Computational Modeling of the C-Terminal Glycan of Type IV Pilin

By

Xiaotong Zuo

A thesis submitted to the Johns Hopkins University in conformity with the requirements of the degree of Master of Science in Engineering

Baltimore, Maryland

May 2016

ABSTRACT

Type IV pili are extracellular, filamentous, adhesive appendages that are assembled from a protein monomer called pilin. Type IV pili provide several properties to the bacterium, including twitching motility [1], DNA uptake [2], host-cell adhesion [3], and biofilm formation [4]. Besides these properties, type IV pili are also considered as the main virulence factor of bacterial pathogens. It has been shown that the C-terminal glycosylation of pilin, the pilus monomer, is an important source of bacteria virulence [5].

To discover how glycans influence the conformation and virulence of the Type IV pilin, I used the Rosetta software suite to model the C-terminal glycans along with the terminal protein residues of two types of Type IV major pilins, PilA ACICU and PilA M2, and analyzed the obtained structures from the perspectives of structures, root-mean-square deviations, hydrogen bonds, energies and surface areas. The results show that PilA ACICU has a tendency to be more flexible while PilA M2 is more constrained.

Advisor: Dr. Jeffrey J. Gray

ACKNOWLEDGMENTS

I wish to express my sincere gratitude for my advisor, Dr. Jeffrey J. Gray, who has always supported me throughout this research project. His steady guidance, patience, and encouragements were extremely important in bringing this research project to completion.

I also wish to express my deepest appreciation to my excellent collaborators, Dr. Eric J. Sundberg and Dr. Kurt H. Piepenbrink from the School of Medicine at the University of Maryland. This project could not be realized without their valuable guidance and advice.

A special thanks goes to Dr. Jason W. Labonte, who taught me from the very beginning of coding to every detail of my modeling protocol. Meanwhile, I wish to thank all the lab members for their support during my master's study. I am very proud to work with the most inspired and creative peers in our lab.

Last but not least, I wish to thank Dr. Michael J. Betenbaugh for serving as a reader of my master's thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

I. Pilus

Type IV pili are long, extracellular, filamentous adhesive appendages frequently expressed by Gram-negative [8–10] and Gram-positive [11,12] bacterial pathogens, as well as by archaea [13]. They are primarily composed of major pilin subunits, also known as PilA [9], which are repeatedly assembled and disassembled to mediate pilus function. There is also a small group of pilin-like proteins called minor pilins, which function in priming pilus assembly.

A Type IV major pilin (PilA) is a small (~7–20 kDa) structural protein consisting of a conserved-fold N-terminal α-helix of ~50 amino acid residues and a C-terminal, soluble, globular domain of ~100 residues formed by a four-stranded antiparallel β-sheet, referred to as the pilin headgroup (**Figure 1.1**). The N-terminal α-helix and the β-sheet are connected by the αβ-loop. The N-terminal α-helix consists of two parts: a ~30-residue, hydrophobic, α-helical, N-terminal domain, referred to as the α1-N domain, and a ~20-residue, hydrophobic, α-helical, C-terminal domain near the headgroup, referred to the α1-C domain. The α1-N domain retains the monomers in the inner membrane prior to assembly. When pilins assemble into a pilus, the conserved hydrophobic N-terminal α-helices are buried in the center of the pilus fiber, forming an inner core of the pilus, while the C-terminal soluble headgroups are exposed and form the surface of the pilus fiber [1].

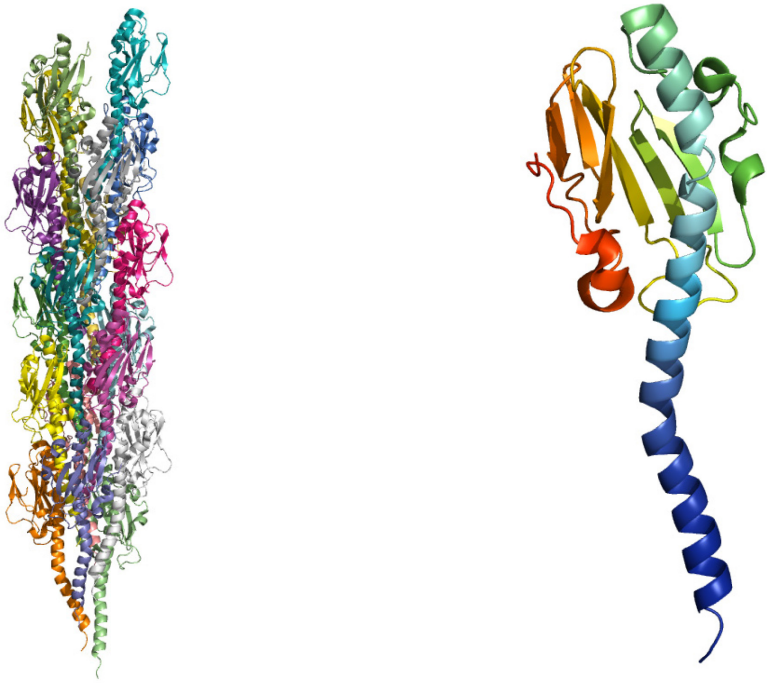


Figure 1.1 Cartoon models of a pilus fiber (left) and a pilin (right). Structures from the *Neisseria gonorrhoeae* Type IV pilus, created from PDB entry 2HIL [6] using the visualization tool, PyMol [7]. For the pilin, the α-helical N-terminal domain is colored blue, and the headgroup is colored orange and green. The C-terminal region is colored red. Coordinates courtesy of Dr. Kurt Piepenbrink (U. Maryland).

Type IV pili are involved in several functional properties of bacteria including twitching motility [1], DNA uptake in natural genetic transformation [2], host-cell adhesion [3], and microcolony or biofilm formation [4]. Besides these properties, as the prominent surface structures of bacteria, type IV pili are the main virulence factor of bacterial pathogens. They are also one of the main targets of the host immune system. The efficacy of the bacterial pathogens to proliferate in the blood during productive infection depends on a bacterium's ability to evade type IV pili-specific antibodies, and alternating induction and shutoff of pilus expression allows for successful evasion of the immune system. There are two main sources for the virulence of pilins. One is transferring DNA from the silent cassettes to the expression locus to generate multiple different antigens to

diversify the virulence of pilins. Another source of antigenic variation is post-translational modification, in particular, glycosylation. O-linked glycosylation has been found in multiple strains of both *Pseudomonas aeruginosa* [14] and *Neisseria* [15,16]. Additional glycosylation sites have been found in class II strains of *Neisseria meningitidis*, where they are hypothesized to play a role in immune evasion [17]. It has been recognized that the class II pilins of *Neisseria meningitidis* were lacking gene conversion while they could successfully evade the immune system, which suggested that glycosylation might also be an essential source of the virulence of bacterial pathogens [17].

In different bacterial species, the glycosylation of pilins may vary both in the glycosylation sites selected and in the identity of the glycan side chains. Genetic study of *Neisseria meningitides* showed that the sites of glycosylation were determined by the primary structure of the pilin, while which glycan was added was determined by the genetic background of the pilin [17]. Thus, as the structures of pilins vary, it is possible for a pilin to own a single glycan side chain or multiple glycan side chains at different sites with various carbohydrate compositions.

Within given species, the minor pilins are typically well-conserved. Only the major pilins are highly variable [18–20] and then only in those regions left exposed in the assembled pilus [21]. To be specific, the N-terminal α-helix is well-conserved. From the 30^{th} amino acid residue to the last residue of the PilA, the sequence exhibits considerable variations.

This project focuses on the modeling of glycans of two distinct PilAs, PilA ACICU and PilA M2, which differ in sequence and structures, and represent two major groups of PilA in two *Acinetobacter baumannii* strains, *A. baumannii* ACICU and *A. baumannii* M2, respectively. *Acinetobacter baumannii* is a Gram-negative, opportunistic pathogen. *A. baumannii* ACICU (also known as H34) is an epidemic, multi-drug-resistant strain belonging to the European clone II group,

which was isolated in an outbreak in Rome in 2005 [22]. *A. baumannii* M2 (referred to as *A. nosocomialis* M2 in some publications) [23–25] was isolated in 1996 from a hip infection of a patient at Cleveland MetroHealth Systems (Cleveland, OH).

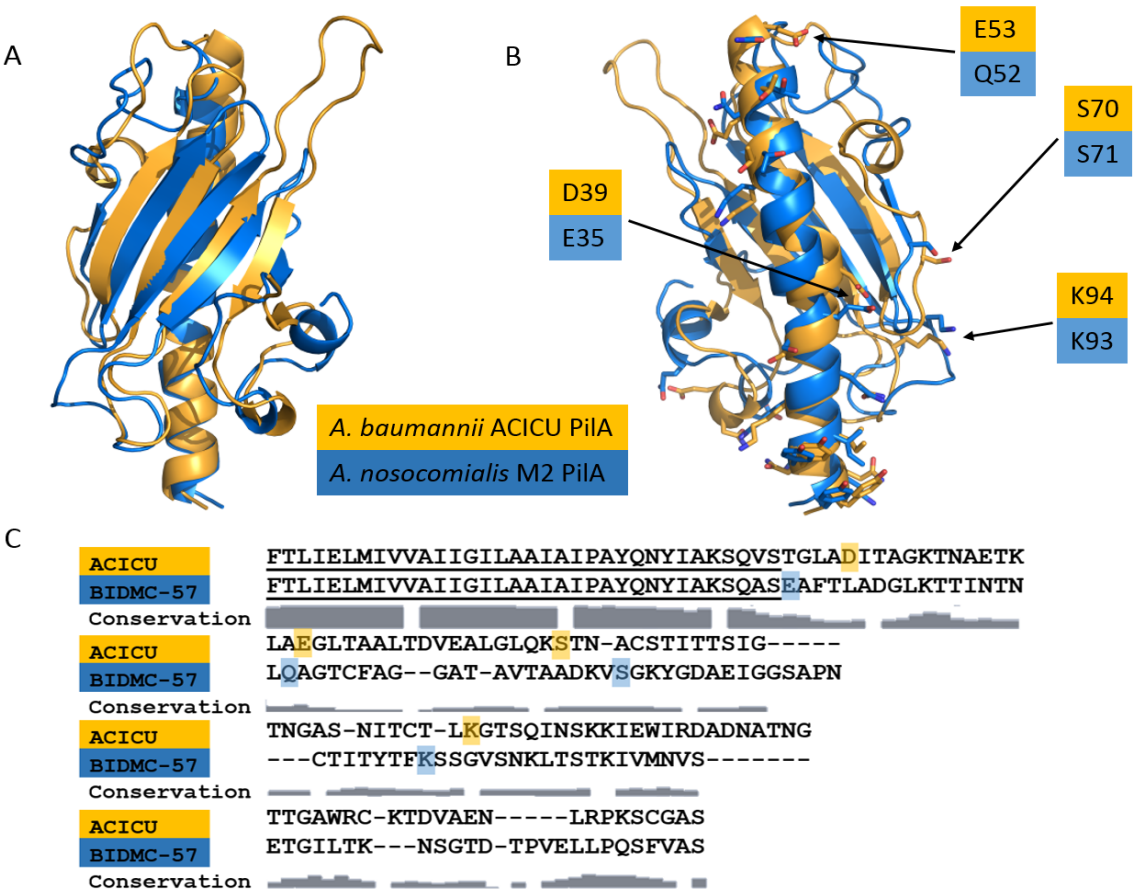


Figure 1.2 Structures and sequences comparisons of PilA ACICU and PilA M2 [5]. (A) The superimposed structures of PilA ACICU and PilA M2. (B) The main chemical moieties of PilA ACICU and PilA M2. (C) The Sequence alignment of PilA ACICU and PilA M2. Figure from K. Piepenbrink, E. Lillehoj, C.M. Harding, J.W. Labonte, X. Zuo, C.A. Rapp, *et al.*, "Diverse Type IV Pili in Multidrug-resistant Acinetobacter Mask Potential Antigens through C-terminal O-Glycosylation," *J. Biol. Chem.* (under review) (2016).

PilA ACICU and PilA M2 share functional similarities while their structures and sequences vary a lot. From the perspective of sequence alignment (**Figure 1.2C**), beginning with Ala 23, there is a sequence identity of only 33% between the PilA ACICU and PilA M2 headgroups. By superimposing the two structures (**Figure 1.2A**), Many chemical moieties (functional groups) can

be found in similar positions (**Figure 1.2B**) that are not obvious from the sequence alignment, which indicates that pilins may be assembled through similar networks of non-covalent interactions.

In addition to the sequence differences, PilA ACICU and PilA M2 also differ in the number of disulfide bonds they contain. PilA ACICU has two disulfide bonds. One, between residues Cys 123 and Cys 136, is the C-terminal disulfide bond, which is nearly universal in type IV pili from Gram-negative bacteria, and the other, between residues Cys 74 and Cys 91, spans the first two strands of the central β-sheet (**Figure 1.3a**). However, this additional disulfide bond in PilA ACICU does not result in any substantial rearrangement of the protein backbone.

Unlike PilA ACICU, PilA M2 contains only a single disulfide bond, between residues Cys 56 and Cys 86, in the αβ-loop and the first strand of the β-sheet, respectively, rather than a disulfide bond at the C-terminus of the pilin headgroup (**Figure 1.3b**). The addition of covalent disulfide bonds is typically understood to be a mechanism of stabilization in polypeptides, and hence the C-terminal disulfide bond that is nearly ubiquitous in type IV pilins is thought to be conserved as a mechanism to stabilize the pilin fold [26]. I chose to model these two groups of PilAs because they are two representative groups of Type IV PilA with identical structures and sequences.

As a major type of post-translational modification of proteins, glycosylation plays an important role in protein properties and functions. Recently, heightened attention has been drawn towards protein glycosylation in bacteria primarily because of the increasing frequency with which it is seen in pathogenic species [27,28]. In particular, most glycosylated proteins of bacterial pathogens are either surface localized or trafficked for secretion and appear to influence interactions with the host. Typical examples of pili among Gram-negative species include pilin subunits of *P. aeruginosa* [29] and *Neisserial* type IV pili (Tfp) [30]. Glycosylation facilitates solubilization of

pilin monomers and pilus fibers [24]. In many instances, glycosylation-defective mutants have been shown to be attenuated in virulence-associated properties and colonization [31–35]. As the glycosylation of Type IV pili is a possible virulence source of bacterial pathogens, I modeled the glycan side chains of PilA ACICU and PilA M2.

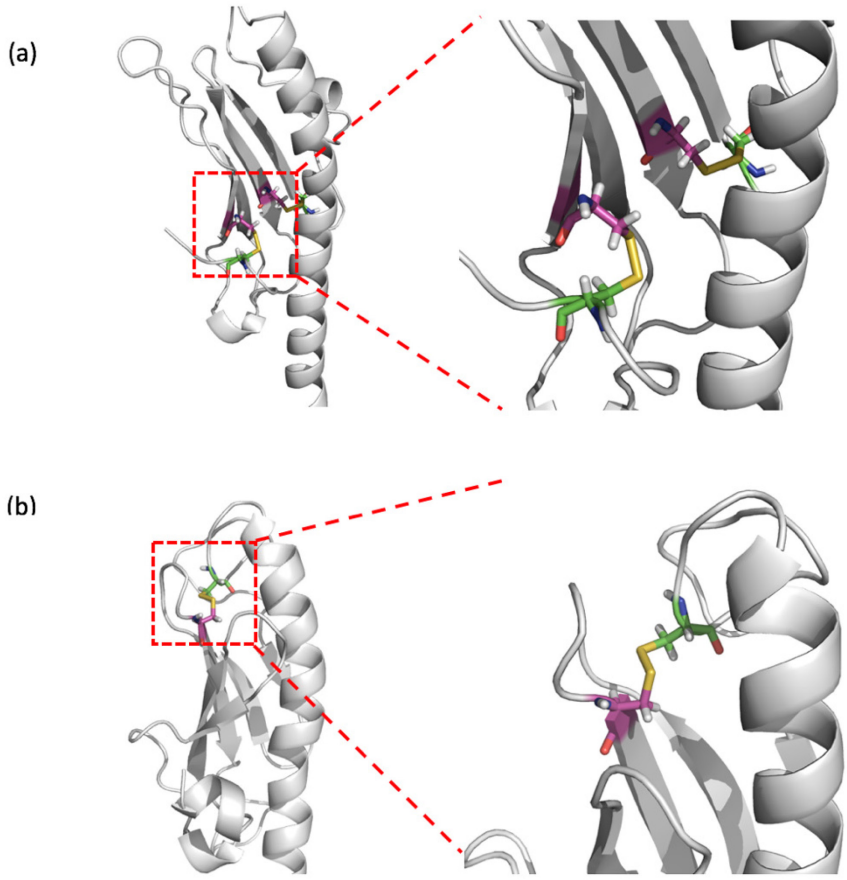


Figure 1.3 Disulfide bonds within PilA ACICU and PilA M2. (a) The two disulfide bonds within PilA ACICU. One is between Cys 74 and Cys 91, and the other is between Cys 123 and Cys 136. (b) The single disulfide bond in PilA M2 is between Cys 56 and Cys 8.

This thesis focuses on the glycans attached to pilins, for which an ensemble of each glycan was modeled based on the major polysaccharide glycan and each model minimized using Rosetta. The compositions of the polysaccharide glycan of PilA ACICU and PilA M2 are shown in **Figure 1.4**, and the information for each carbohydrate residue is listed in **Table 1.1**.

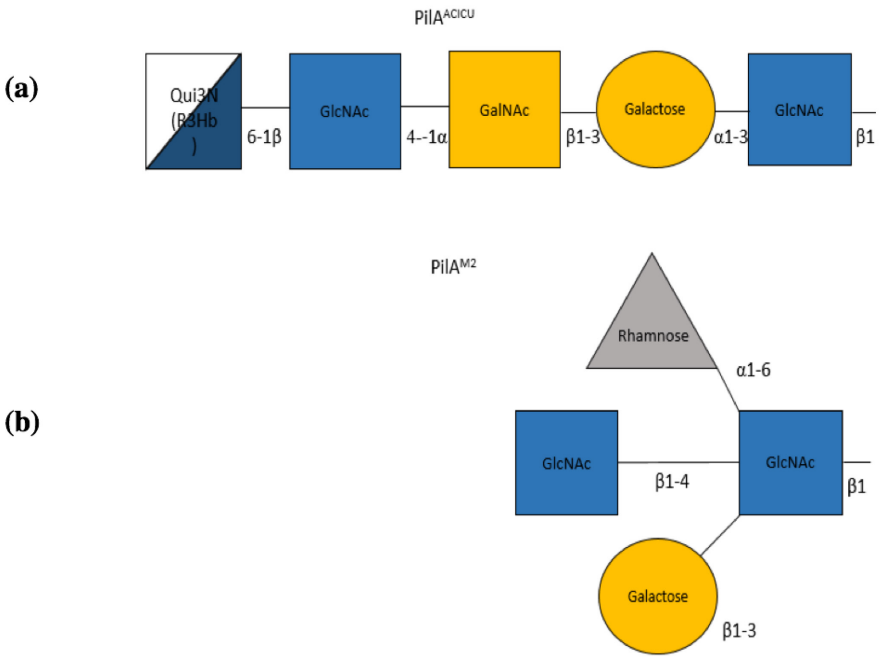


Figure 1.4 The glycans of PilA ACICU (a) and PilA M2 (b) [5].

For PilA ACICU, the glycan residues link to the C-terminal residue Ser 139 of the pilin monomer. This long glycan side chain is comprised of an *N*-acetylglucosamine (Glc*N*Ac), a galactose (Gal), an *N*-acetylgalactosamine (Gal*N*Ac), another Glc*N*Ac and a 6-deoxy glucose called quinovose with its R3Hb ((*R*)-3-hydroxybutyrate) side chain.

For PilA M2, the glycan residues link to the terminal Ser 136 of the M2 pilin monomer. Unlike the linear glycan chain of PilA ACICU, that of M2 has a compact globular conformation starting with a Glc*N*Ac residue. A galactose, another Glc*N*Ac, and a rhamnose are linked to the Glc*N*Ac at its 3rd carbon, the 4th carbon, and the 6th carbon respectively.

Table 1.1 Glycan information for PilA ACICU and PilA M2.

Residue Number	Glycan residue name for PilA ACICU	IUPAC Designation	Class
1	GlcNAc	→3)-2-(Acetylamino)-2-deoxy-β-D-glucose	HexNAc
2	Gal	→3)-α-D-Galactose	Hex
3	GalNAc	→4)-2-(Acetylamino)-2-deoxy-β-D-galactose	HexNAc
4	GlcNAc	→6)-2-(Acetylamino)-2-deoxy-α-D-glucose	HexNAc
5	Qui (Quinovose)	6-Deoxy-β-D-Glucose	dHex

Residue Number	Glycan residue name for PilA M2	IUPAC Designation	Class
1	GlcNAc	[→6)-]-[→4)-]-→3)-2-(Acetylamino)-2-deoxy-β-D-glucose	HexNAc
2	Gal	β-D-Galactose	Hex
2’	GlcNAc	2-(Acetylamino)-2-deoxy-β-D-glucose	HexNAc
2’’	Rha (Rhamnose)	6-Deoxy-α-L-Mannose	dHex

II. Rosetta modeling

Rosetta [36] is a comprehensive software suite for modeling macromolecular structures. As a flexible, multi-purpose application, it includes tools for structure prediction, design, and remodeling of proteins and nucleic acids.

Rosetta is a structure-prediction tool which offers a wide variety of effective sampling algorithms to explore backbone, side-chain, and sequence-space for macromolecules and its excellence has generalized to more community-wide exercises including RNA-puzzles[37] and Critical

Assessment of PRotein Interactions (CAPRI) [38]. Rosetta boasts broadly tested scoring (energy) functions and contains an unparalleled breadth of applications from folding to docking to design [36].

The scoring function of Rosetta calculates the energy of a structure based on a combination of physics-based and statistics-based potentials. The energy unit I used here, the Rosetta Energy Unit (REU), does not match up with actual physical energy units (*e.g.*, kcal/mol or kJ/mol). Instead, Rosetta energies are on an arbitrary scale. The value of 1.00 REU could change based on the particular score function used, so an "REU" for one protocol might not be comparable to an "REU" for another protocol.

The basic idea of modeling a molecule is to perturb different torsion angles of residues and then compare the energy changes to decide whether accept the perturbation or not. Each perturbation applied to the molecule is called a 'move', which is realized by a 'mover'. In Rosetta, several components are typically used to make a complex algorithm. One such component is a Monte Carlo object. By performing a Monte Carlo object, all energy changes for each move are kept, and whether to accept a move or revert it back is based on the energy change, which is decided using the Metropolis criterion. If the new move results in a lower score, the move is accepted. If the new score is higher than the old one, the acceptance rate for this movement is $P = e^{-\Delta E/kT}$, where the default value of kT is 1.00 REU. Monte Carlo searches are often paired with minimization, which is to minimize the energy of current structure to a local energy minimum before energy comparison. This Monte Carlo plus minimization method [39] makes the Monte Carlo searching more efficient.

III. Main goal of this thesis

The goal of my thesis project is to model the C-terminal glycans of PilA ACICU and PilA M2, to find the best conformations of the C-terminal glycans and the most probable conformations of the structures, as well as to enhance the capacity of current ligand predictive models.

As former studies have shown that multisite glycosylation of the Type IV pilin resulted in coverage of the pilus surface, so that the virulence of the bacterial pathogens might be impacted by glycosylation, it is also worth comparing changes in surface area to measure the extent to which C-terminal glycosylation of PilA ACICU and PilA M2 would mask the pilin protein from binding.

CHAPTER II

METHODS

I. Generation of initial structures

Initial structures of PilA ACICU and PilA M2 monomers were built by Dr. Kurt Piepenbrink (U. Maryland) [40] using the structure of the full-length *P. aeruginosa* PAK pilin. Initial assembled models of the pili were created by superimposition onto a model of the *N. gonorrhoeae* Type IV pilus filament (Protein Data Bank ID 2HIL, 12.5 Å resolution) [6], and adjustment of the N-terminal helix position to eliminate clashes between subunits. The resulting models then underwent rigid-body minimization by UCSF Chimera [41]. These were further adjusted by Dr. Jason Labonte (Johns Hopkins University) using the Discovery Studio Visualizer software [42].

My test runs indicated that the glycan of a single PilA ACICU protein monomer could not interact with any neighbor monomers. However, for PilA M2, though the glycan of PilA M2 was short and would not interact with the main body of the pilin protein to which it linked, the glycan would interact with the headgroup of its neighbor pilin. Thus, I used the monomer for PilA ACICU modeling, while I used the dimer for PilA M2 modeling.

Because the initial structures, the monomer of PilA ACICU and the dimer of PilA M2, were chosen from the already assembled pilus fiber complex models, there were steric clashes found at the joints of two neighbor pilins, which resulted in extremely poor Rosetta scores. So before modeling, I deleted several residues of the N-terminal α-helix. To be specific, the first 17 residues of PilA ACICU and the first 25 residues of PilA M2 were deleted. These deleted regions are far from the glycan side chain; thus, the deletion does not influence the modeling results. However, there were

still clashes in the αβ-loop of PilA M2, which increased the total score of PilA M2. I kept the clashes since they would not influence the modeling of the carbohydrates.

Before modeling, the structures were pre-packed and minimized to avoid any unexpected atom placing error or clash.

II. Computational model

a. MoveMap

Within the code for a Rosetta protocol, a MoveMap object specifies which degrees of freedom are fixed and which are free to change. In this project, the backbone and the side chains of the saccharide residues and the amino acid "tail" of the pilin are free to move. Other degrees of freedom are fixed. The tail region refers to the amino acid residues after the last C-terminal α-helix of the pilins. The glycans of PilA ACICU and PilA M2 are both linked to the terminal serines of their pilin protein monomers. So for both PilA ACICU and PilA M2, the glycan side chain was modeled along with the tail region, since together they acted as a long tail and were able to swing relatively freely.

Before the tail region, it was assumed that the protein monomer would not make large movements due to hydrogen bonds and other interactions between secondary structures, and the protein body of the pilin has little influence on the glycan side chain. Thus, for pilin ACICU, the amino acid tail region is 15 residues, from Thr 125 to Ser 139. For pilin M2, the tail region is 19 residues, from Lys 118 to Ser 136.

b. Score function

A score function was created based on the default Rosetta full-atom score function, a well-trained scoring function that includes comprehensive physics-based and statistics-based potentials. **Table 2.1** lists the weight of each score term of the original score function and the score function I used in this project. The weight of each term of my score function remains the same as the original, except the fa_intra_rep score term, which stands for the full-atom intra-residue repulsive Van der Waals energy. The weight of the fa_intra_rep score term was tuned up from 0.004 to 0.440 due to the Dunbrack rotamer energy scoring difference between amino acids and carbohydrates. The fa_dun term, which refers to the Dunbrack rotamer energy, has already

Table 2.1 Score terms and weights for the full-atom score functions used.

Scores	Original Weight	Modified Weight
fa_atr	0.800	0.800
fa_rep	0.440	0.440
fa_sol	0.750	0.750
fa_intra_rep	**0.004**	**0.440**
fa_elec	0.700	0.700
pro_close	1.000	1.000
hbond_sr_bb	1.170	1.170
hbond_lr_bb	1.170	1.170
hbond_bb_sc	1.170	1.170
hbond_sc	1.100	1.100
dslf_fa13	1.000	1.000
rama	0.200	0.200
omega	0.500	0.500
fa_dun	0.560	0.560
p_aa_pp	0.320	0.320
ref	1.000	1.000
sugar_bb	1.000	1.000

included the fa_intra_rep term for amino acids, thus the weight of the fa_intra_rep is set to a small value. However, the fa_dun term does not include the fa_intra_rep for glycans. As the goal is to model glycans, so I tuned up the weight of fa_intra_rep to 0.44, the same as the one of fa_rep, the full-atom repulsive Van der Waals energy.

III. Glycan modeling algorithm—*FloppyTail* algorithm

To model the terminal glycan side chains of both PilA ACICU and PilA M2, a refined *FloppyTail* algorithm [43,44] was applied due to its outstanding performance in the modeling of terminal residues of molecules which may have an ensemble of native-like structures as the swinging of terminal tail makes the energy of each structure indistinctive.

The two basic moves, small/shear moves, are applied to change the torsion angle. A small move is the simplest move, which perturbs ϕ/ψ of a random residue by a random small angle. A shear move perturbs ϕ of a random residue by a small angle and ψ of the same residue by the same small angle of opposite sign [45].

The fundamental idea of the *FloppyTail* algorithm (**Figure 2.1**) is to apply a set of torsion angle perturbations of the backbone to collapse the tail into a folded conformation from an initially straight-out-into-space extended conformation, which is to say the algorithm roughly searches for energy minima along the energy landscape of the structure. And then recovers the lowest energy structure and uses more precise small/shear moves with side chain repacking to refine its position. This is conceptually similar to how *ab initio* folding works in Rosetta, although it was not designed for that purpose (and does not contain temperature-scheduling, *etc.*) Because the tail of the starting structure was manually made, the starting structure was far away from the native ones, and the energy of the starting pose is considerably higher than the top-scoring structure.

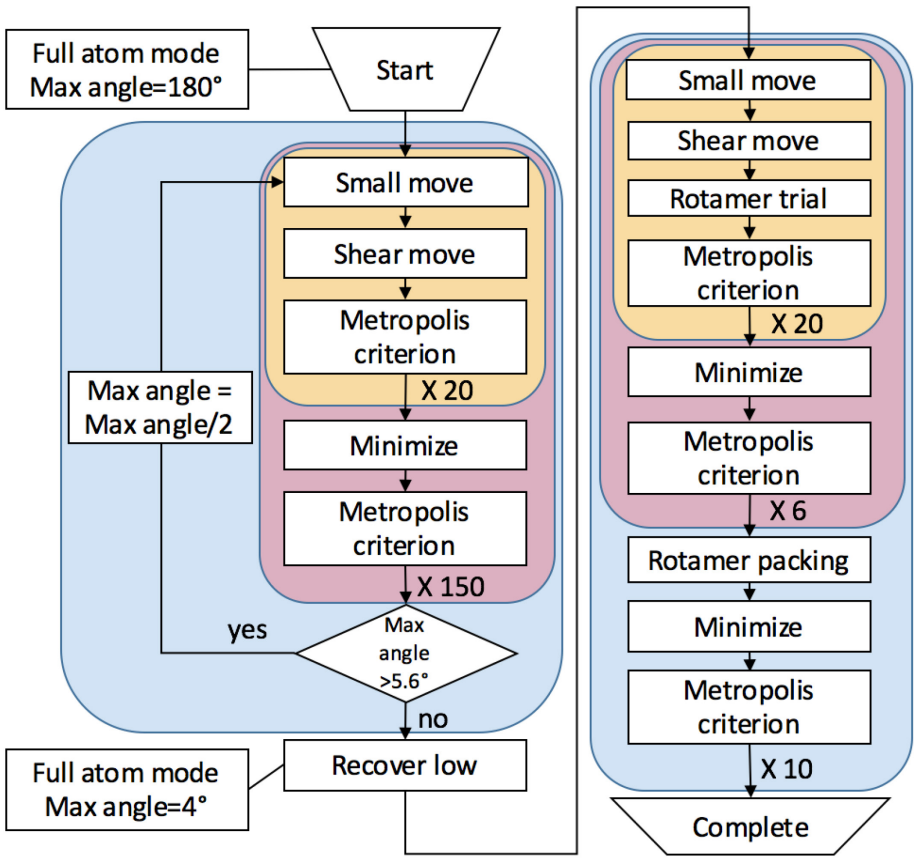


Figure 2.1 A flowchart of the *FloppyTail* algorithm. The left half is the broad search stage, and the right half is the refined search stage.

A refined two-stage *FloppyTail* algorithm was made for the glycan side chain modeling of PilA ACICU and PilA M2.

a. The broad search: gradient based modeling stage

In the broad search stage, a gradient-based perturbation of the torsion angles is applied. At the beginning, the maxima of the phi (ϕ) and psi (ψ) torsion angles are set to 180°. After a set of small moves and shear moves, the structure is minimized, and its local energy minimum is located by applying the Metropolis criterion. Then, the maxima of the ϕ/ψ torsion angles are divided into half. The process is repeated five times until the maxima of the ϕ/ψ torsion angles become 5.6°.

b. The refined search: rotamer packing stage

Rotamers are a library of the most likely, low-energy side-chain conformations, which are used to pack side chains for both amino acid and carbohydrate residues. In the refined search stage, there are two tasks: a precise torsion angle perturbation and rotamer packing, and packing rotamers is the main factor of increasing the total computational complexity for the protocol.

In the refined search refinement stage, the lowest energy structure from the broad search stage is recovered and then refined in the refined search step by applying a more precise perturbation of the torsion angles, with the maximum ϕ/ψ torsion angles set as 4°. Importantly, in the refined search stage, side-chain packing and rotamer packing are involved. The resulting structure is then minimized to locate its local energy minimum with acceptance or rejection according to the Metropolis criterion.

With this protocol, 8,000 structures were generated each for PilA ACICU and PilA M2. 1,000 structures each for PilA ACICU and PilA M2 without the glycan were also generated for comparison.

CHAPTER III

RESULTS

I. Analysis of modeling effectiveness

The premise of an effective broad search modeling stage is to make sure that there are enough trajectories of torsion angle perturbations at the broad search stage so that the structure can go through all possible conformations and then use Monte Carlo to accept a move or not. Enough generated structures are needed to avoid ineffective sampling. In this project, 8,000 structures were generated for each target. At the broad search stage, there are 14,250 perturbations of torsion angles, including phi (ϕ), psi (ψ) and (ω). At the refined search stage, there are 6,000 times precise torsion angle perturbations and rotamer packings.

The broad search stage plays a fundamental role in the modeling process. From a structural perspective, the conformation is dramatically changed by gradient-based torsion angle perturbations as the maximum magnitudes of perturbations decrease from 180° to 5.6°. From an energy perspective, the structure can jump out of a local minimum and search along the energy landscape to find another, lower minimum at the broad search stage. Thus, the effectiveness of the broad search stage should be validated first. To assess the searching effectiveness of the broad search modeling stage, there were 1000 structures for PilA ACICU and PilA M2 generated from broad search sampling. The score distributions of these structures are shown in **Figure 3.1**.

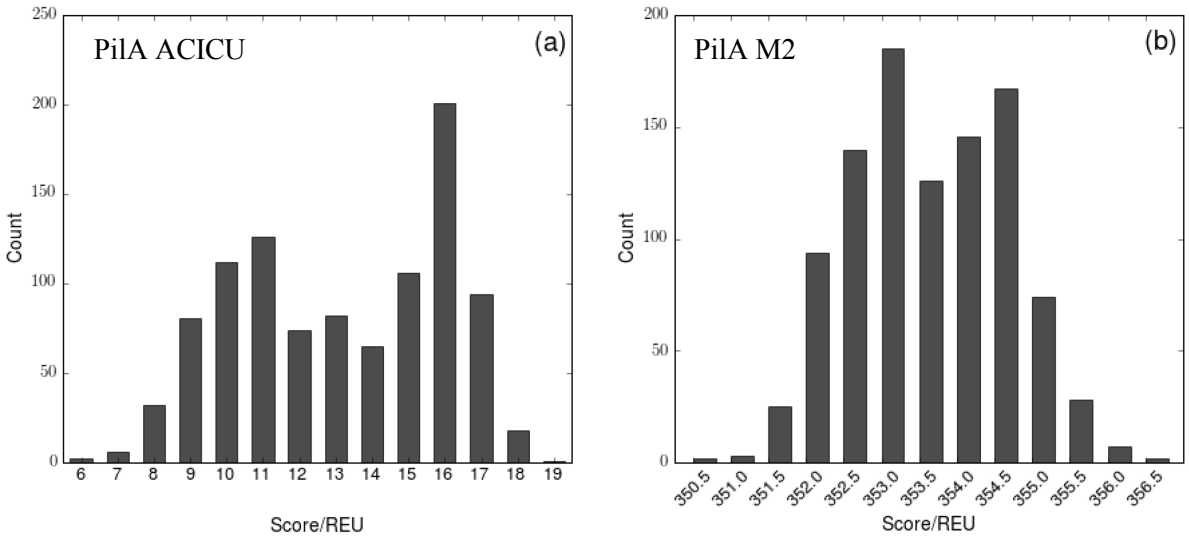


Figure 3.1 Score distributions of PilA ACICU and PilA M2 after the broad search stage. (a) The score distribution of PilA ACICU after the broad search stage. (b) The score distribution of PilA M2 after the broad search stage.

For PilA ACICU, after the gradient-based broad search stage, the scores of the whole pilin with glycan of the 1000 structures ranged from 6 to 19 Rosetta Energy Units (REU). The scores of 4% of the structures are below 8 REU, which is the top 18.9% of the score range.

The scores of the 1000 PilA M2 structures with glycans range from 350.5 to 357 REU. The score range of PilA M2 is nearly the half of PilA ACICU's. The scores of 3% of structures are below 351.5 REU, which is the top 23.4% of the score range.

In this project, PilA M2 always has a smaller score range compared with PilA ACICU. Although the tail region is very long and able to move in a large space, compared with PilA ACICU, the protein part of PilA M2 is more compact, and there are more hydrogen bonds found in the tail region. Also, unlike the long glycan of PilA ACICU, the short, globular glycan side chain of PilA M2 has a smaller space to move around, and this glycan frequently interacts with its neighbor pilin monomer, which prevents it from moving around. Thus, the conformations of PilA M2 varies little, so the score range of PilA M2 is smaller than PilA ACICU.

Figure 3.2 shows the score distributions of those two structures after the whole protocol. Interestingly, though the conformations of PilA M2 remain similar to each other, the score range of PilA M2 becomes larger after the refined search stage compared with its 1000 broad search results, which implies that the refined search stage might have the major impact on score changes. That is to say the side-chain packing might be the major source of score differences for PilA M2.

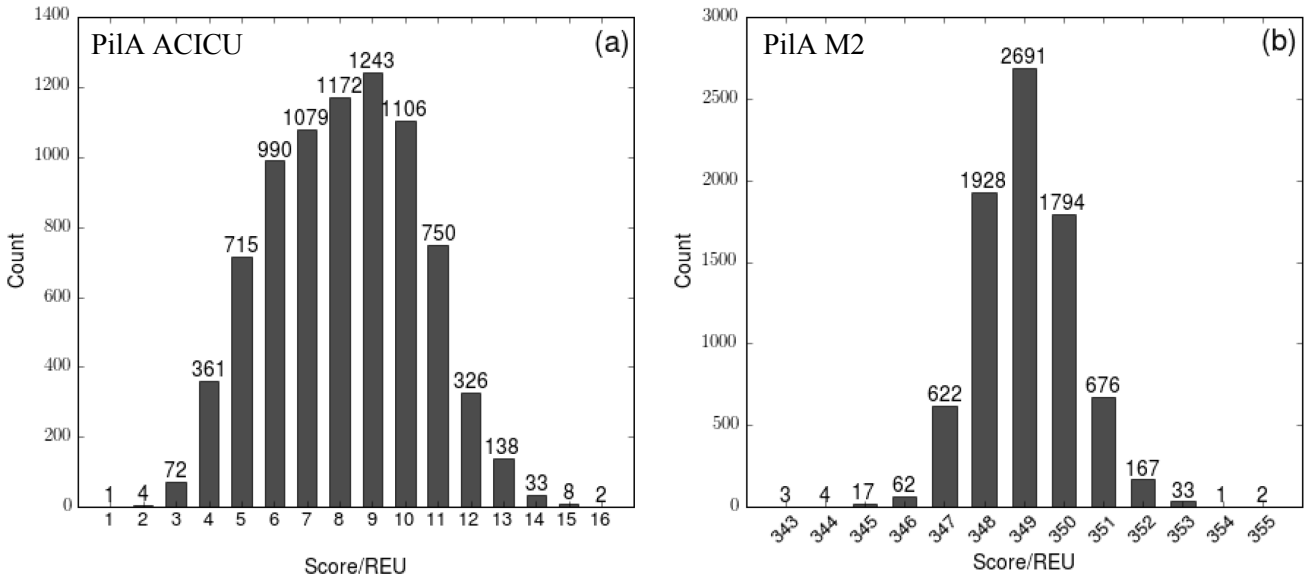


Figure 3.2 Score distributions of PilA ACICU and PilA M2 of the 8000 structures after the entire protocol (a) Score distribution of PilA ACICU (b) Score distribution of PilA M2.

II. Structures analysis

There were 8000 structures each generated for PilA ACICU and PilA M2 with the protocol. The top-scoring structure, ten top-scoring structures, and 20 randomly chosen structures are shown in **Figure 3.3**.

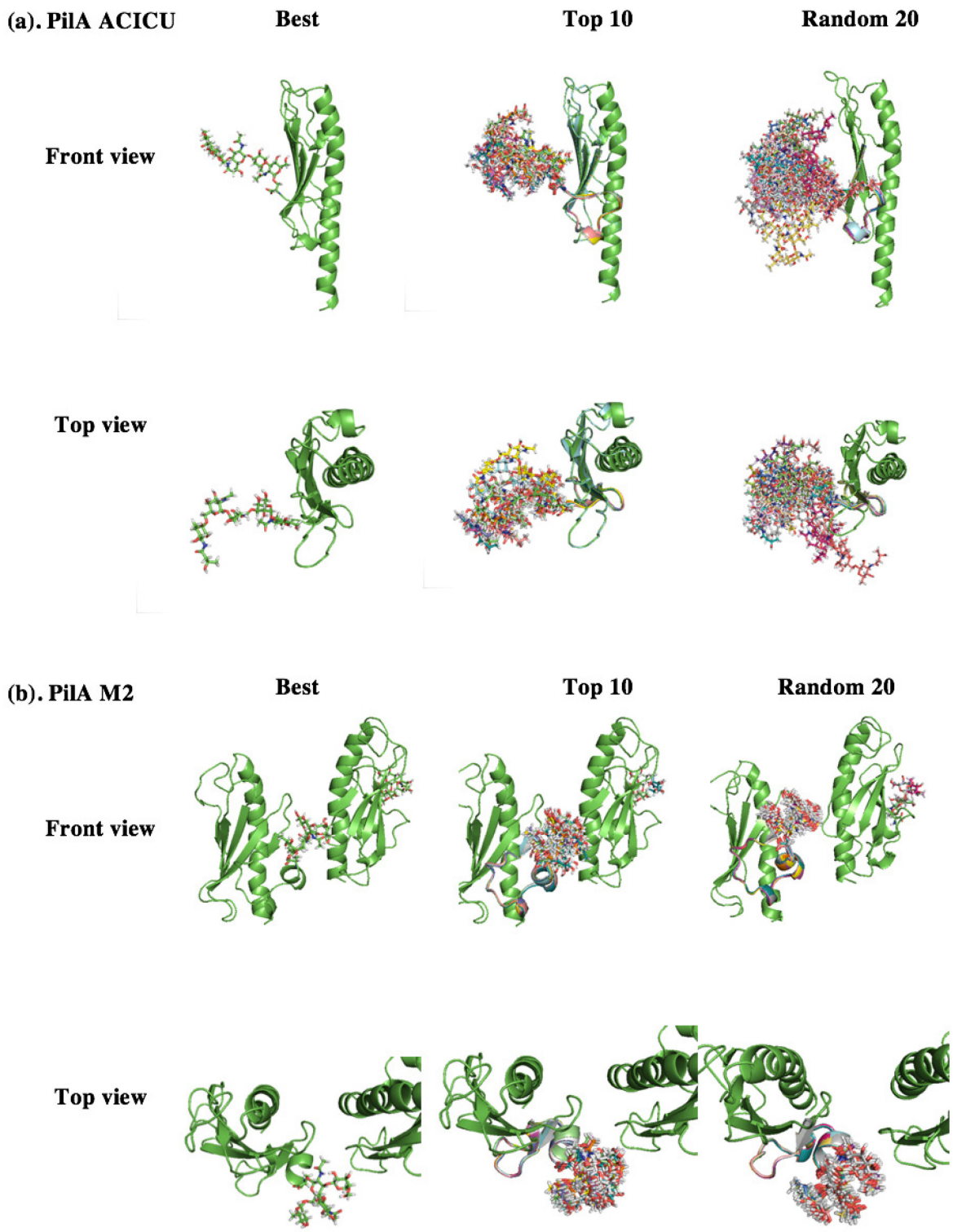


Figure 3.3 Structures obtained with the modeling protocol. (a) The front views and top views of the best, ten top-scoring, and randomly chosen 20 structures of PilA ACICU monomer. (b) The front views and top views of the best, ten top-scoring, and randomly chosen 20 structures of PilA M2 dimer.

For PilA ACICU, the tail regions of all the structures stay at similar positions, but the long carbohydrate portion curls, beginning at the 4th glycan Glc*N*Ac. The glycans of the top 10 structures form a 70° sector near the headgroups of the pilin proteins. Theoretically, as the glycan side chain does not have permanent contact with its protein body, it is able to swing around the pilin protein within a large angle range without drastic energy changes. However, the tail region does not move much, because the bond angles and torsion angles of the first glycan is important to the entire conformation of the glycan side chain. Any unfavorable angles would contribute to high scores. So the glycan side chains of the top-scored structures stay in the small 70° fan-shaped sector.

For PilA M2, the initial model contains two pilin monomers, chain A and chain B. I only modeled the protein tail region and the glycan side chain of chain B (the left monomer in the figure). Both the tail region and the glycan side chain retain a stable and conserved conformation. Looking at the ten top-scoring structures and random 20 structures of PilA M2, there are no obvious differences among those structures. The root-mean-square deviation (rmsd) analysis also stresses the structural similarity of the ensemble of PilA M2.

III. Rmsd analysis

Root-mean-square deviation (rmsd) represents the extent of position deviation between two structures, derived from the distances between their corresponding Cα atoms of amino acid residues and C1 atoms of carbohydrates. For all the rmsd *vs.* score plots in this paper, the reference positions (at rmsd=0 Å) are set as the positions of the lowest scored structures, which are assumed to be the top-scoring structures. A large value of rmsd means the structure is far from the lowest scored structure.

For PilA ACICU, as mentioned above, the glycan side chains of ten top-scoring structures form a 70° sector near the headgroups of the pilin proteins. Within this sector, the conformations are considered stable and have low energies. In **Figure 3.4a**, the lowest-scoring structure is set as the reference structure (rmsd = 0 Å). The rmsds of the other top 10 structures, for which the glycan side chains are within the 70° sector, are around 1.0 Å. Also, the center of the darkest area above the top 10 structures, which stands for the majority of the 8000 structures, is around 1.0 Å as well. This phenomenon indicates that conformations with glycan side chain within the 70° sector near the headgroups of the pilin proteins are favored. The score differences of these structures may mainly come from rotamer packing.

Interestingly, three "good" structures of PilA ACICU are found with low energies but show a wide variation in conformations compared with the top-scoring structures. These structures are far away from the lowest-scored structure, and their rmsds are larger than 4.0 Å. So here I call them "far" structures. Unlike the best-scoring structures, the glycan side chains of the far structures bend to an almost opposite direction at Gly 137 of their protein tail region (**Figure 3.5a**). As shown in **Table 3.1** and **Figure 3.5b**, Gly 137 of the tail region is flexible due to its highly variable backbone torsion angles (especially the ϕ angle in this case [46]). This results in a very different glycan side chain position. However, this direction change of the glycan side chain is realized at the expense of a score increment, which is approximately 2 to 4 REU (blue points of **Figure 3.4a**). Although the scores are slightly higher, these structures may still be considered as favored conformations.

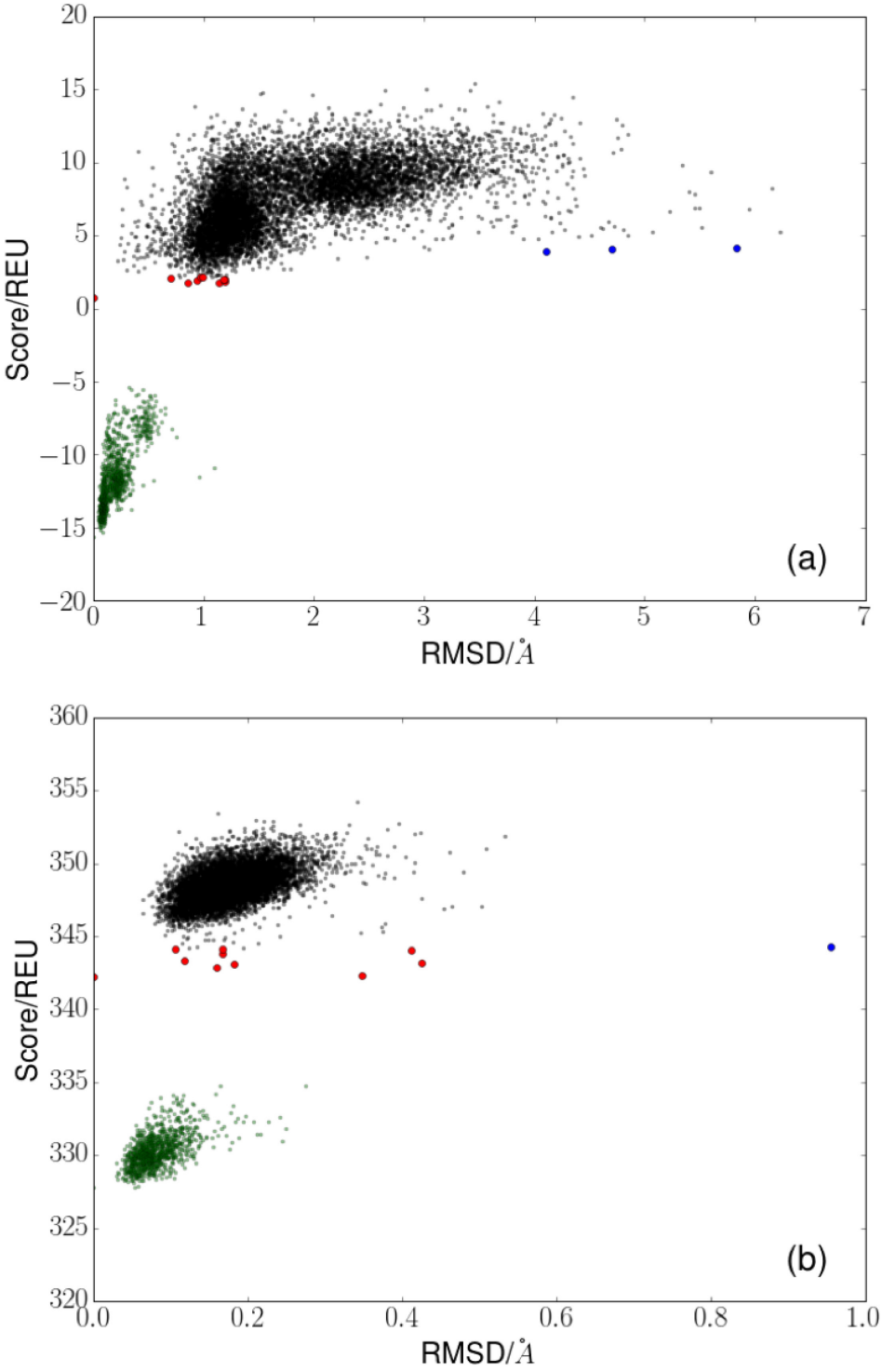


Figure 3.4 Funnel plots of PilA ACICU (a) and PilA M2 (b). Each of 8000 glycostructures for each protein is plotted. The ten top-scoring structures are colored red. The blue points are the structures farthest from the top scoring pose, but still having good scores. The green points are the 1000 structures without glycan side chains, generated with the same protocol.

For PilA M2, the rmsds of all the structures are within 1.0 Å. The rmsds of the majority are within 0.4 Å. Though PilA M2 has a long tail region to move, 19 residues from Lys 118 to Ser 136, the tails are held to the pilin protein bodies by the strong forces between pilins and tails. Also, the short and globular glycan side chain is compact, so it cannot move to a far position.

Table 3.1 ϕ/ψ torsion angles comparison of the top-scoring structures and far structures, from residue 135 to 138.

structures	135		136		137		138	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
Far_1	68.40	11.31	-97.34	147.35	**126.12**	-157.09	-149.23	155.95
Far_2	92.57	-6.63	-96.84	148.60	**68.27**	-122.93	-68.98	144.97
Far_3	74.44	15.66	-117.27	144.52	**102.80**	152.02	-61.32	148.38
Best	75.38	11.66	-109.03	135.14	**-124.62**	-142.63	-153.12	154.65

By employing the same modeling protocol, and with the setting that only the tail region was free to move, there were 1000 structures generated for each structure without glycans. These structures are shown in **Figure 3.4** as the green points. For PilA ACICU, the scores range from -15.63 to -5.39 REU. The rmsds are less than 0.50 Å for 96.2% of all structures. For PilA M2, the scores range from 327.80 to 334.77 REU and most structures have a rmsd less than 0.20 Å. For both PilA ACICU and PilA M2 without carbohydrates, the lowest scores are much smaller than the ones with the glycan side chains, suggest the score contributions of glycan side chains are around 15 REU. The ranges of their rmsds are smaller compared with the former results, suggesting that the total rmsd of a structure is contributed to by the glycan side chain.

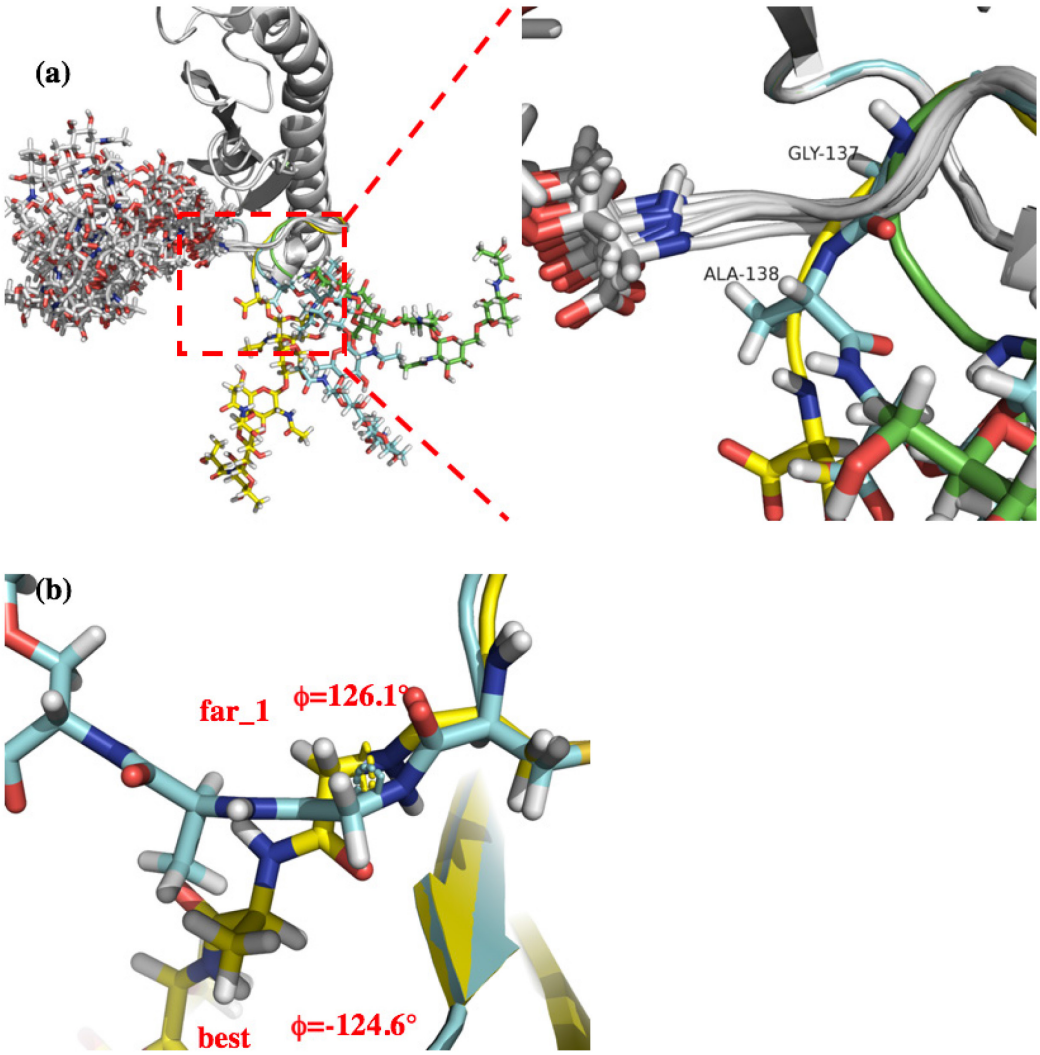
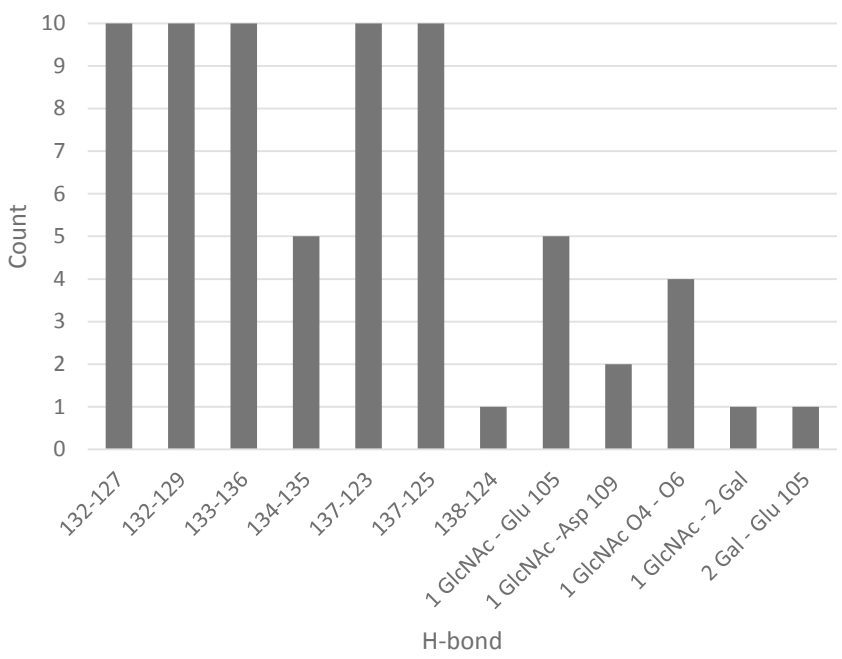


Figure 3.5 The structures with good scores but high rmsds of PilA ACICU. (a) The structures change direction at Gly 137. (b) A ϕ of Gly 137 comparison between the best and the far_1 structures.

IV. Hydrogen bonds

For PilA ACICU, from residue Arg 132 to the end of the glycan side chain of pilin ACICU, there are 12 hydrogen bonds observed (**Figure 3.6 a**), five of which exist in all the ten top-scoring structures (**Figure 3.7**). These five hydrogen bonds are between the tail and the main body of the pilin protein. They should play an important role in holding the protein tail back to the protein body, stabilizing the conformation and minimizing the energy of the structure.

(a)



(b)

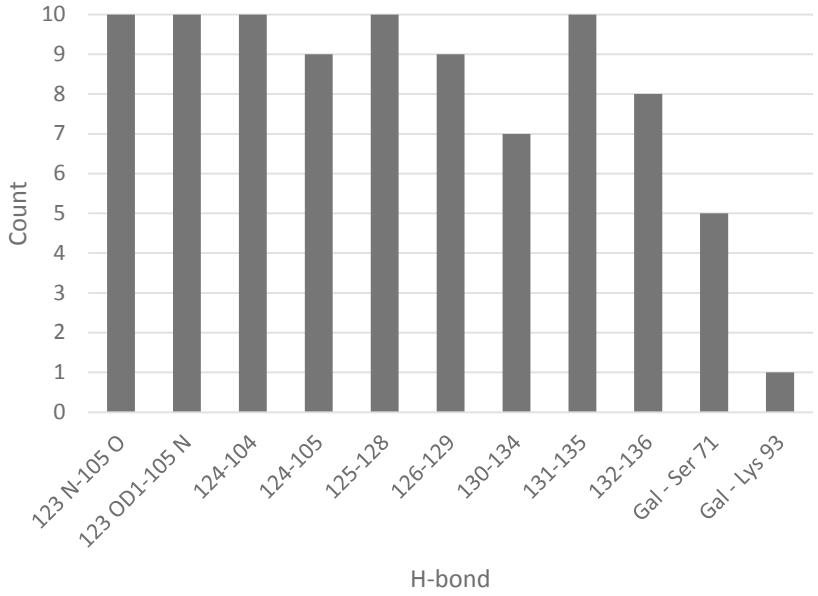


Figure 3.6 Hydrogen bond counts of the ten top-scoring structures of PilA ACICU (a) and PilA M2 (b).

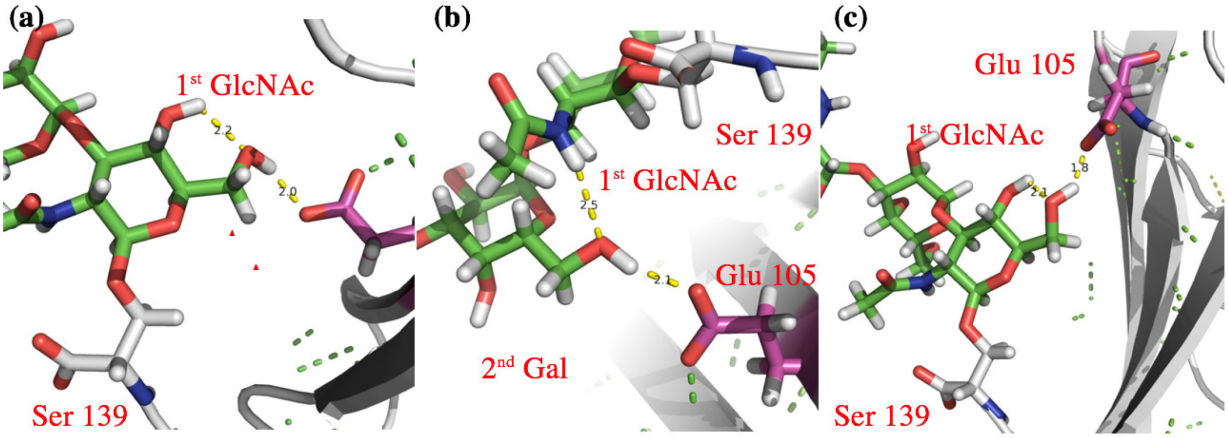


Figure 3.7 Hydrogen bonds of PilA ACICU between the glycan side chain and Glu 105, Asp 109. (a) Hydrogen bonds 1st GlcNAc O4-1st GlcNAc O6, and 1st GlcNAc O6-Asp 109, from the lowest score structure. (b) Hydrogen bonds 1st GlcNAc-2nd Gal and 2nd Gal-Glu 105. (c) Hydrogen bonds 1st GlcNAc O4-1st GlcNAc O6 and 1st GlcNAc-Glu 105.

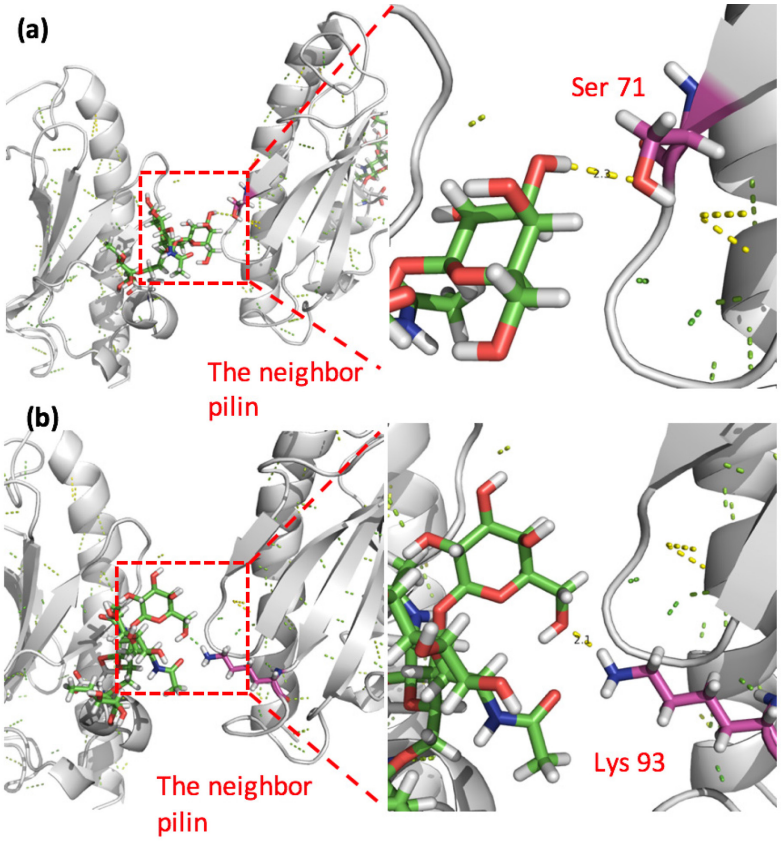


Figure 3.8 Hydrogen bonds of PilA M2 between the glycan side chain and the neighbor pilin. (a) Hydrogen bond between galactose and Ser 71 of the neighbor pilin. (b) Hydrogen bond between galactose and Lys 93 of the neighbor pilin.

For PilA M2, from residue Asp 123 to the ends of glycan of pilin M2, there are 11 hydrogen bonds found in the best 10 structures (**Figure 3.6 b**). Interestingly, for the glycan of pilin M2, the four carbohydrate residues have fewer hydrogen bonds compared with pilin ACICU. The hydrogen bonds with the glycan side chain involved are found between the galactose, which links to the center glucose at the 6^{th} carbon, and the amino acids of the neighbor pilin monomer (Ser 71 and Lys 93, respectively) (**Figure 3.8**). The energies of these two hydrogen bonds are -0.356 and -1.14 REU respectively. The interaction with Ser 71 is more common. The interactions between monomers may restrain the short glycan side chain from moving around and slightly contribute to the pili assembling as well. More importantly, the glycan side chain might influence the virulence of pilins and pili by covering the neighbor pilin protein surface.

V. Energy comparison

To explore which factors contribute to the total energy differences, I compared the energy of each residue within the tail region and glycan of the top 10 conformations for both PilA ACICU and PilA M2 (**Figure 3.9**). For both PilA ACICU and PilA M2, within the tail region, the energy of each residue does not vary much. So the energy lines are overlapped at the tail regions. Specifically, for PilA ACICU, hydrogen bonds are found among residue 133, 136, and 137 in all conformations; thus, the energy of these residues are lower than those of other peptide residues. For PilA M2, the figure shows a gentle upwards-sloping line within the tail region, indicates that the tail region has a stable structure due to the interactions between tail region and the main body of the pilin.

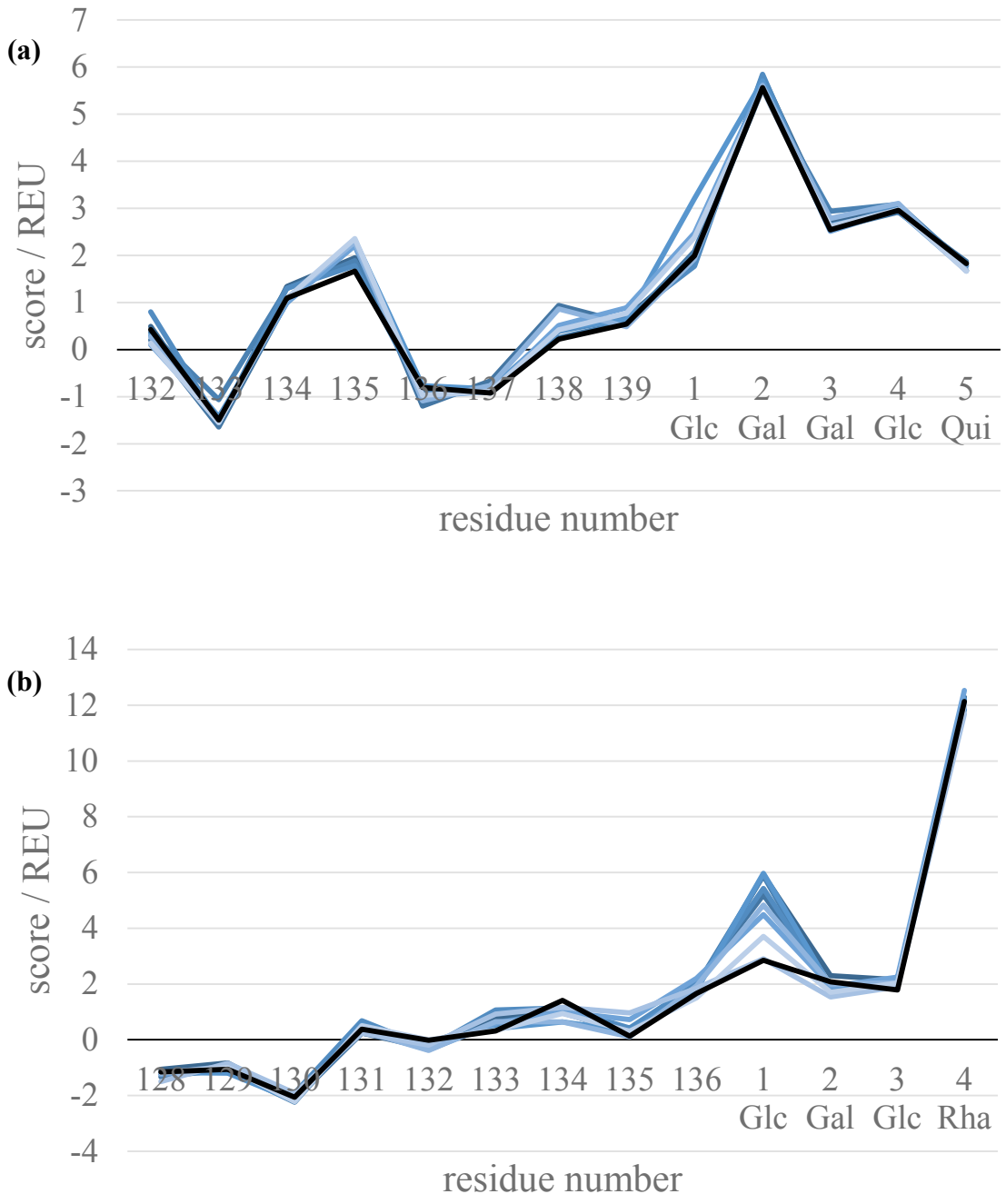


Figure 3.9 Residue-by-residue energy comparison for top 10 structures of PilA ACICU and M2. (a) The energy comparison of PilA ACICU. (b) The energy comparison of PilA M2. The black lines are the lowest score structures. The blue lines are other 9 top-scoring structures.

When it comes to glycans, there are significant energy gains for several glycan residues. For example, the 2nd glycan Gal of PilA ACICU has an energy of 6 REU. What is more, the energy of the 1st glycan Glc of PilA M2 is 6 REU, and the 4th glycan Rha of PilA M2 has an energy above 12 REU. These high scores result from the "sugar backbone" energy term, which can only be found in glycans and is always positive. The default weight of the "sugar backbone" is 1 in the score function. So it is meaningless to compare energy of glycans with the energy of peptide tail regions. In PilA M2, the energy of the 1st glycan Glc seems different from other glycan residues. Its energies vary from 3 to 6 REU though the conformations are very similar to each other. This Glc locates at an essential position since other three glycans are connected to it, so any slight changes of these glycan conformations will influence the torsion angles of the 1st Glc, thus the energies of the 1st Glc vary largely.

VI. Surface area comparison

The solvent-accessible surface area (SASA) is the surface area of a molecular that is accessible to a solvent. In Rosetta, there is a series of functions for SASA calculation for molecules. Here, I used Rosetta to calculate the SASA of PilA ACICU and PilA M2 to find the surface area changes of different conformations.

I analyzed the SASA of the ten top-scoring structures of PilA ACICU and PilA M2, which differ in the surface area in total (A_{tot}) or just in the protein (A_p) or glycan (A_g) portions or in the contact area (interface) between the protein and the glycan (A_{pg}). These area terms can be described in a simple formula:

$$A_{pg}=(A_p+A_g-A_{tot})/2$$

For PilA ACICU (**Table 3.2** and **Figure 3.10**), the 5th, 7th, and 8th structures are the most representative structures to discuss the area changes. The 5th structure has the lowest score, the largest total surface area, and the smallest protein-glycan contact area (A_{pg}) among the ten top-scoring structures. The 7th and 8th structure have the smallest total surface area and the largest protein–glycan contact area among the ten top-scoring structures. Additionally, the 8th structure has the largest glycan surface area among the ten top-scoring structures.

Table 3.2 SASA comparison of PilA ACICU.

score rank	score (REU)	total surface area A_{tot} (Å^2)	protein surface area A_p (Å^2)	glycan surface area A_g (Å^2)	protein–glycan contact area A_{pg} (Å^2)
1	0.78	8209.14	7214.28	1311.39	158.26
2	1.75	8279.12	7210.14	1305.23	118.12
3	1.76	8262.20	7214.43	1294.61	123.42
4	1.85	8319.18	7208.85	1348.11	118.89
5	**1.96**	**8333.02**	**7227.03**	**1342.72**	**118.37**
6	2.04	8283.86	7211.96	1313.63	120.86
7	**2.06**	**8121.25**	**7192.87**	**1296.75**	**184.19**
8	**2.09**	**8220.96**	**7238.16**	**1374.54**	**195.87**
9	2.15	8296.69	7194.96	1354.04	126.15
10	2.16	8298.87	7214.12	1332.53	123.89
Min	0.78	8121.25	7192.87	1294.61	118.12
Max	2.16	8333.02	7238.16	1374.54	195.87
Avg	1.86	8262.43	7212.67	1327.36	138.80
σ (stddev)	0.41	63.19	13.30	27.03	29.59

For PilA M2 (**Table 3.3** and **Figure 3.10**), because I modeled a single glycan side chain of a dimer, and the glycan side chain is contact with the neighbor monomer, the area of PilA M2 calculated here has different definitions compared with PilA ACICU, but the protein–glycan contact area A_{pg} is comparable to that of PilA ACICU. There is no significant difference found in the surface area of PilA M2 due to the similarity of the top 10 conformations. However, the protein–glycan contact area A_{pg} of PilA M2 is larger (with an average of 189.9 Å^2) than that of PilA ACICU (with an

average of 138.8 Å^2). This phenomenon is led by the contact between glycan and the neighbor pilin protein; thus, the glycan covers some part of the neighbor pilin

protein, which reduces the total surface area and increases the protein–glycan contact area.

Table 3.3 SASA comparison of PilA M2 dimer. A_{pg} here refers to the protein–glycan contact area between a single glycan side chain and its neighbor pilin protein. Other area terms refer to the area of the dimer. terms are referred to area of a dimer.

score rank	score (REU)	total surface area A_{tot_dimer} (Å^2)	protein surface area A_{p_dimer} (Å^2)	glycan surface area A_{g_dimer} (Å^2)	protein–glycan contact area A_{pg_dimer} (Å^2)	protein–glycan contact area A_{pg} (Å^2)
1	342.26	13069.00	12074.82	1717.19	361.50	176.75
2	342.35	13044.83	12031.22	1755.70	371.05	185.70
3	342.89	13068.62	12039.15	1743.62	357.08	172.22
4	343.13	12995.86	12037.90	1740.33	391.19	206.53
5	343.15	12964.77	12042.27	1703.63	390.56	205.95
6	343.36	13040.09	12065.22	1748.46	386.80	201.99
7	343.81	13041.82	12047.80	1742.90	374.44	189.78
8	344.05	13036.38	12071.31	1744.74	389.83	205.71
9	344.09	13014.75	12041.04	1757.08	391.69	207.16
10	344.15	13071.66	12037.76	1696.50	331.30	146.72
Min	342.26	12964.77	12031.22	1696.50	331.30	146.72
Max	344.15	13071.66	12074.82	1757.08	391.69	207.16
Avg	343.32	13034.78	12048.85	1735.02	374.54	189.85
σ(stddev)	0.70	34.44	15.63	21.45	19.91	19.99

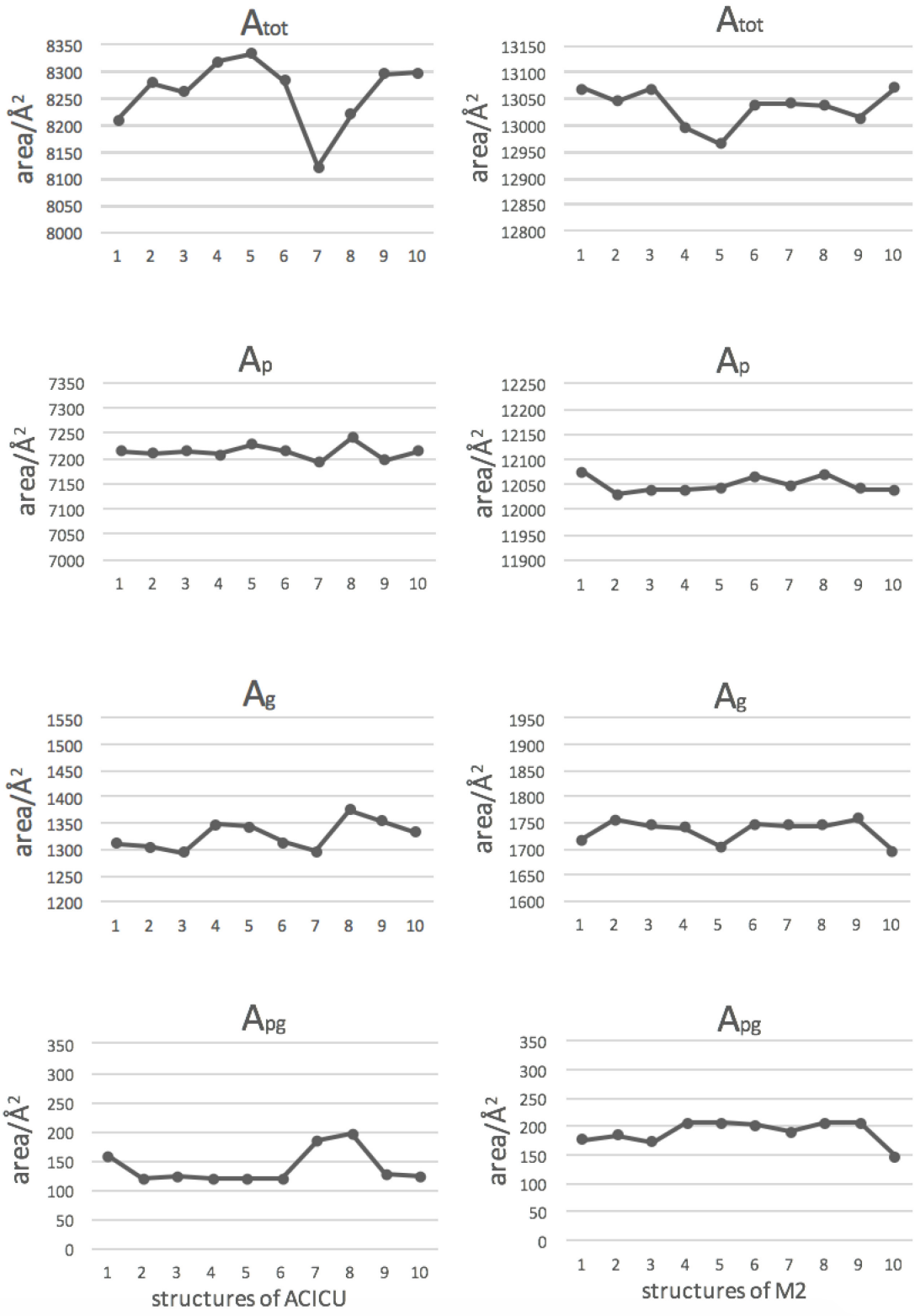


Figure 3.10 SASA comparison of PilA ACICU (a) and PilA M2 (b).

As shown in **Table 3.2** and **Figure 3.11**, there are two small contact regions found in the 5th structure, and the contact area of the 5th structure is extremely small, since its glycan is vertical to the β-strands of the headgroup. Thus, the protein–glycan contact area between the glycan and the pilin protein is less than in other cases (**Figure 3.11c**). On the contrary, the 7th and the 8th structures have similar conformations. The glycans are tangent to the β-strands, so that some part of the β-strands are buried (**Figure 3.11a,b**). The total surface area of the 7th and the 8th structures is reduced by enlarging the protein–glycan contact area (**Figure 3.11d,e**).

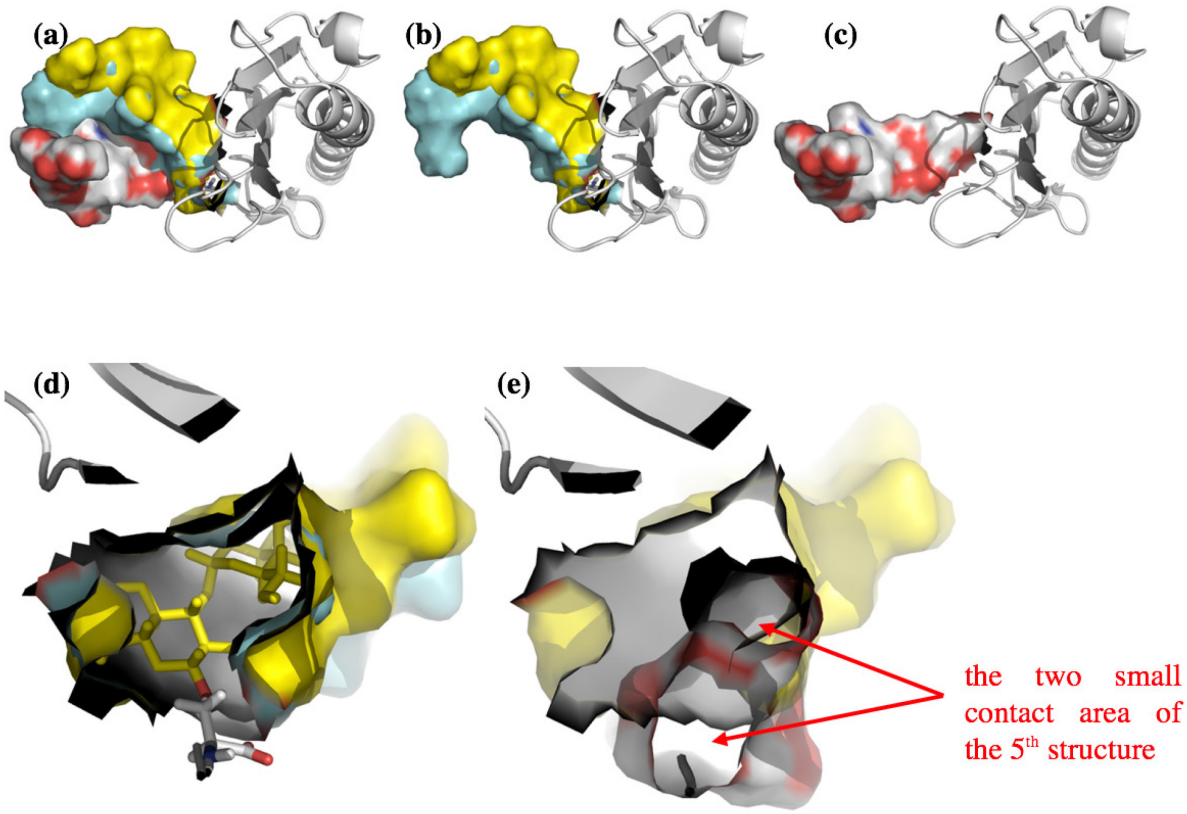


Figure 3.11 SASA comparison of PilA ACICU. (a)–(c) The top views of the glycan surfaces. The 5th structure with a lowest protein-glycan contact area is colored gray, and the 7th structure is colored yellow. The 8th structure is colored blue. (d) The protein–glycan contact area (A_{pg}) of the 7th and 8th structures. (e) The A_{pg} comparison of the 5th structure and the 7th structure.

From an experimental view, the accessible surface area was then measured by Dr. Kurt Piepenbrink (U. Maryland) using a 10 Å particle probe to approximate the surface area needed for protein binding. The resulting models are shown in **Figure 3.12** and the change in accessible surface area for each protein in **Figure 3.13**. In both cases, C-terminal glycosylation significantly reduces the surface area available for antibody binding. While the total area masked by the glycan is similar for the two structures, it is distributed differently between the two; all of the buried surface area for the ACICU glycan is contained within a single subunit while in the case of the M2 glycan, it is split between two neighboring subunits [5].

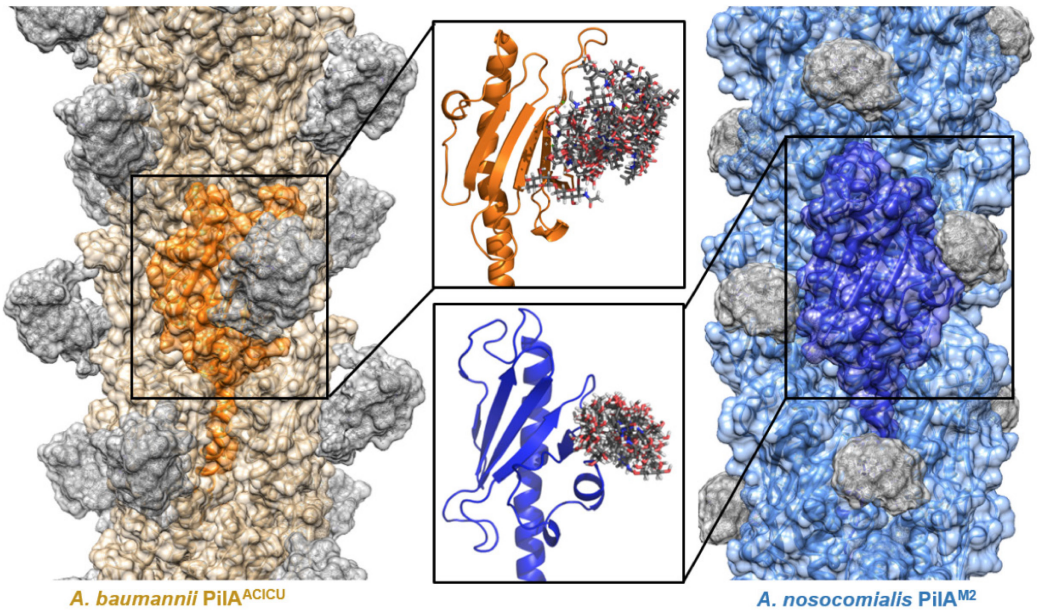


Figure 3.12 Models of glycosylated Acinetobacter Type IV Pili. models of assembled type IV pili from *A. baumannii* ACICU (orange) and *A. nosocomialis* M2 (blue) are depicted with semi-transparent surfaces; glycan residues are shown in grey. Inset panels show detail of the computed glycan conformations [5]. Figure from K. Piepenbrink, E. Lillehoj, C.M. Harding, J.W. Labonte, X. Zuo, C.A. Rapp, *et al.*, "Diverse Type IV Pili in Multidrug-resistant Acinetobacter Mask Potential Antigens through C-terminal O-Glycosylation," *J. Biol. Chem.* (under review) (2016).

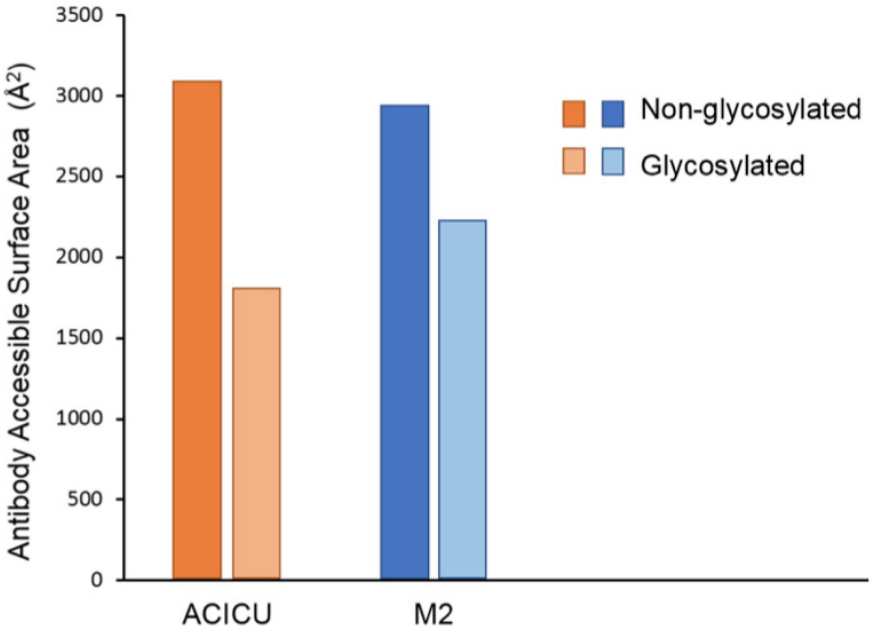


Figure 3.13 Accessible surface area calculations. The surface area of a single pilin monomer in an assembled pilus exposed to a 10 Å probe is shown for both *A. baumannii* ACICU (orange) and *A. nosocomialis* M2 (blue), with and the without the C-terminal glycan [5]. Figure from K. Piepenbrink, E. Lillehoj, C.M. Harding, J.W. Labonte, X. Zuo, C.A. Rapp, *et al.*, "Diverse Type IV Pili in Multidrug-resistant Acinetobacter Mask Potential Antigens through C-terminal O-Glycosylation," *J. Biol. Chem.* (under review) (2016).

Chapter IV

CONCLUSION

This project modeled the repeat unit of the C-terminal glycan side chains of the Type IV major pilins, PilA ACICU and PilA M2, and low energy conformations and other potential conformations were obtained. However, an unexpected phenomenon in this work happens to the modeling of PilA M2. Although the modeling protocol and initial setting of movemaps for those two pilins are the same, all PilA M2 structures obtained from the process do not have apparent different conformations, which leads to totally different conformation and energy analyses for PilA ACICU and PilA M2, and suggests that the glycan might have a rigid structure. However, though the conformations of PilA M2 are highly similar to each other, the scores of PilA M2 structures vary largely, which implies that PilA M2 may have a favored backbone conformation, so that its energy changes mainly come from side-chain packing, not from the backbone conformation. On the contrary, the score changes of PilA ACICU are mainly from the backbone movements compared with PilA M2.

Glycosylation reduces the accessible surface area of the Type IV pilin, which to some extend prevent pilins from binding with antibodies. The linear glycan of PilA ACICU is more flexible while the globular glycan of PilA M2 retains constrained. Specifically, as linear glycan is more flexible than globular glycan, it can cover more surface area.

This project is one of the first applications which modeling peptide residues and carbohydrates simultaneously. In this project, protein and glycan residues were modeled simultaneously using the same algorithm, but these two components differed a lot, especially in the energy score calculation and rmsd calculation. As we know, carbohydrates and other types of ligands are

structural different from amino acids. Unlike amino acid residues, carbohydrates have different structural properties (different backbone torsion angles and energy scoring terms, *etc.*) which complicated the overall modeling process. To meet the requirements of glycan modeling, I modified the score function. Moreover, I calculated the rmsd of structures based on the combination of Cα atoms of amino acids and C1 atoms of carbohydrates. Moreover, when calculating energies of protein and glycan residues, there are different energy terms used, thus the comparison of energy between protein and glycan is not proper. We can only compare energy changes within the same residue or component. This project has enhanced the capacity of current ligand predictive models.

Based on the Type IV PilA ACICU monomer and PilA M2 dimer modeling results, future work on the glycan modeling of the assembled Type IV pili can be realized. Also, the glycans I modeled in this project are just for specific Type IV pilin species, so it is promising to model other types of glycans and predict the conformations involving various types of glycans based on the current results.

REFERENCES

[1] D. Wall, D. Kaiser, Type IV pili and cell motility., Mol. Microbiol. 32 (1999) 1–10. doi:10.1111/j.1365-2958.2003.03977.x.

[2] H.S. Seifert, R.S. Ajioka, C. Marchal, P.F. Sparling, M. So, DNA transformation leads to pilin antigenic variation in Neisseria gonorrhoeae., Nature. 336 (1988) 392–5. doi:10.1038/336392a0.

[3] H. Takahashi, T. Yanagisawa, K.S. Kim, S. Yokoyama, M. Ohnishi, Meningococcal pilV potentiates Neisseria meningitidis type IV pilus-mediated internalization into human endothelial and epithelial cells, Infect. Immun. 80 (2012) 4154–4166. doi:10.1128/IAI.00423-12.

[4] J.B. Chem, S. Park, J.H. Exton, K.O. Neill, M. Wigler, M.C. Biol, et al., Type IV Pili , Transient Bacterial Aggregates , and Virulence of Enteropathogenic Escherichia coli, 280 (1998).

[5] K. Piepenbrink, E. Lillehoj, C.M. Harding, J.W. Labonte, X. Zuo, C.A. Rapp, et al., Diverse Type IV Pili in Multidrug-resistant Acinetobacter Mask Potential Antigens through C-terminal O-Glycosylation, J. Biol. Chem. under review (2016).

[6] L. Craig, N. Volkmann, A.S. Arvai, M.E. Pique, M. Yeager, E. Egelman, et al., Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions, Mol. Cell. 23 (2006) 651–662. doi:10.1016/j.molcel.2006.07.004.

[7] Schrödinger, LLC, The PyMOL Molecular Graphics System, Version-1.8, 2015.

[8] B.J. Stone, Y. Abu Kwaik, Expression of multiple pili by Legionella pneumophila: identification and characterization of a type IV pilin gene and its role in adherence to mammalian and protozoan cells., Infect. Immun. 66 (1998) 1768–1775.

[9] M.S. Strom, S. Lory, Structure-function and biogenesis of the type IV pili, Annu. Rev. Microbiol. 47 (1993) 565–596. doi:10.1146/annurev.micro.47.1.565.

[10] T. Taniguchi, Y. Fujino, K. Yamamoto, T. Miwatani, Sequencing of the gene encoding the major pilin of pilus colonization factor antigen III (CFA / III) of human enterotoxigenic Escherichia coli and evidence that CFA / III is related to type IV pili . Sequencing of the Gene Encoding the Major Pilin of Pi, 63 (1995) 724–728.

[11] K. Piepenbrink, G.A. Maldarelli, C.F.M. de la Pena, T. Dingle, G. Mulvey, A. Lee, et al., Structural and Evolutionary Analyses Show Unique Stabilization Strategies in the Type IV Pili of Clostridium difficile, Structure. (2014) 385–396. doi:10.1016/j.str.2014.11.018.

[12] K.H. Piepenbrink, G.A. Maldarelli, C.F.M. de la Pena, G.L. Mulvey, G.A. Snyder, L. De Masi, et al., Structure of clostridium difficile pilj exhibits unprecedented divergence from known type IV pilins, J. Biol. Chem. 289 (2014) 4334–4345. doi:10.1074/jbc.M113.534404.

[13] K. Lassak, A. Ghosh, S.V. Albers, Diversity, assembly and regulation of archaeal type IV pili-like and non-type-IV pili-like surface structures, Res. Microbiol. 163 (2012) 630–644. doi:10.1016/j.resmic.2012.10.024.

[14] S. Voisin, J. V. Kus, S. Houliston, F. St-Michael, D. Watson, D.G. Cvitkovitch, et al., Glycosylation of Pseudomonas aeruginosa strain Pa5196 type IV pilins with mycobacterium-like α-1,5-linked D-Araf oligosaccharides, J. Bacteriol. 189 (2007) 151–159. doi:10.1128/JB.01224-06.

[15] F.E. Aas, Å. Vik, J. Vedde, M. Koomey, W. Egge-Jacobsen, Neisseria gonorrhoeae O-linked pilin glycosylation: Functional analyses define both the biosynthetic pathway and glycan structure, Mol. Microbiol. 65 (2007) 607–624. doi:10.1111/j.1365-2958.2007.05806.x.

[16] P.M. Power, K.L. Seib, M.P. Jennings, Pilin glycosylation in Neisseria meningitidis occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in Escherichia coli, Biochem. Biophys. Res. Commun. 347 (2006) 904–908. doi:10.1016/j.bbrc.2006.06.182.

[17] J. Gault, M. Ferber, S. Machata, A. Imhaus, C. Malosse, A. Charles-orszag, et al., Neisseria meningitidis Type IV Pili Composed of Sequence Invariable Pilins Are Masked by Multisite

Glycosylation, (2015) 1–24. doi:10.1371/journal.ppat.1005162.
[18] A. Cehovin, M. Winterbotham, J. Lucidarme, R. Borrow, C.M. Tang, R.M. Exley, et al., Sequence conservation of pilus subunits in Neisseria meningitidis, Vaccine. 28 (2010) 4817–4826. doi:10.1016/j.vaccine.2010.04.065.
[19] A.K. Criss, K.A. Kline, H.S. Seifert, The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*, Mol. Microbiol. 58 (2005) 510–519. doi:10.1111/j.1365-2958.2005.04838.x.
[20] C. Toma, H. Kuroki, N. Nakasone, M. Ehara, M. Iwanaga, Minor pilin subunits are conserved in Vibrio cholerae type IV pili, FEMS Immunol. Med. Microbiol. 33 (2002) 35–40. doi:10.1016/S0928-8244(02)00273-0.
[21] T.E. Blank, H. Zhong, A.L. Bell, S. Thomas, M.S. Donnenberg, T.S. Whittam, Molecular Variation among Type IV Pilin (bfpA) Genes from Diverse Enteropathogenic Escherichia coli Strains Molecular Variation among Type IV Pilin (bfpA) Genes from Diverse Enteropathogenic Escherichia coli Strains, 68 (2000) 7028–7038. doi:10.1128/IAI.68.12.7028-7038.2000.Updated.
[22] M. Iacono, L. Villa, D. Fortini, R. Bordoni, F. Imperi, R.J.P. Bonnal, et al., Whole-genome pyrosequencing of an epidemic multidrug-resistant Acinetobacter baumannii strain belonging to the European clone II group, Antimicrob. Agents Chemother. 52 (2008) 2616–2625. doi:10.1128/AAC.01643-07.
[23] M. Strain, M.D. Carruthers, C.M. Harding, B.D. Baker, R. a Bonomo, K.M. Hujer, et al., Draft Genome Sequence of the Clinical Isolate Acinetobacter, Genome Announc. 1 (2013) 1–2. doi:10.1128/mBio.00360-13.7.
[24] K.M. Clemmer, R.A. Bonomo, P.N. Rather, Genetic analysis of surface motility in Acinetobacter baumannii, Microbiology. 157 (2011) 2534–2544. doi:10.1099/mic.0.049791-0.
[25] C.M. Harding, E.N. Tracy, M.D. Carruthers, P.N. Rather, L.A. Actis, R.S. Munson, Acinetobacter baumannii strain M2 produces type IV Pili which play a role in natural transformation and twitching motility but not surface-associated motility, MBio. 4 (2013) 1–10. doi:10.1128/mBio.00360-13.
[26] L. Craig, M.E. Pique, J.A. Tainer, Type IV pilus structure and bacterial pathogenicity, Nat. Rev. Microbiol. 2 (2004) 363–378. doi:10.1038/nrmicro885.
[27] I. Benz, M.A. Schmidt, MicroReview Never say never again : protein glycosylation in pathogenic bacteria, Mol. Microbiol. 45 (2002) 267–276.
[28] C.M. Szymanski, B.W. Wren, Protein glycosylation in bacterial mucosal pathogens., Nat. Rev. Microbiol. 3 (2005) 225–237. doi:10.1038/nrmicro1100.
[29] P. Castric, F.J. Cassels, R.W. Carlson, Structural Characterization of the Pseudomonas aeruginosa 1244 Pilin Glycan, J. Biol. Chem. 276 (2001) 26479–26485. doi:10.1074/jbc.M102685200.
[30] E. Stimson, M. Virji, K. Makepeace, A. Dell, H.R. Morris, G. Payne, et al., Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose, Mol Microbiol. 17 (1995) 1201–1214. doi:10.1111/j.1365-2958.1995.mmi_17061201.x.
[31] C.M. Szymanski, D.H. Burr, P. Guerry, Campylobacter Protein Glycosylation Affects Host Cell Interactions Campylobacter Protein Glycosylation Affects Host Cell Interactions, Infect. Immun. 70 (2002) 2242–2244. doi:10.1128/IAI.70.4.2242.
[32] S. Grass, A.Z. Buscher, W.E. Swords, M.A. Apicella, S.J. Barenkamp, N. Ozchlewski, et al., The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis, Mol. Microbiol. 48 (2003) 737–751. doi:10.1046/j.1365-2958.2003.03450.x.
[33] M. Schirm, E.C. Soo, A.J. Aubry, J. Austin, P. Thibault, S.M. Logan, Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori, Mol. Microbiol. 48 (2003) 1579–1592. doi:10.1046/j.1365-2958.2003.03527.x.
[34] D.R. Hendrixson, V.J. DiRita, Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract, Mol. Microbiol. 52 (2004) 471–484. doi:10.1111/j.1365-2958.2004.03988.x.

[35] S.K. Arora, A.N. Neely, B. Blair, S. Lory, R. Ramphal, Role of motility and flagellin glycosylation in the pathogenesis of Pseudomonas aeruginosa burn wound infections, Infect. Immun. 73 (2005) 4395–4398. doi:10.1128/IAI.73.7.4395-4398.2005.
[36] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, R. Jacak, K. Kaufman, et al., Rosetta 3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules, (2014). doi:10.1016/B978-0-12-381270-4.00019-6.R.
[37] Z. Miao, R.W. Adamiak, M. Blanchet, M. Boniecki, M. Bujnicki, S. Chen, et al., RNA-Puzzles Round II : assessment of RNA structure prediction programs applied to three large RNA structures, (2015) 1066–1084. doi:10.1261/rna.049502.114.
[38] J.J. Gray, S.E. Moughon, T. Kortemme, O. Schueler-Furman, K. Misura, A. V. Morozov, et al., Protein-protein docking predictions for the CAPRI experiment, Proteins Struct. Funct. Genet. 52 (2003) 118–122. doi:10.1002/prot.10384.
[39] Z. Li, H. a Scheraga, Monte Carlo-minimization approach to the multiple-minima problem in protein folding., Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 6611–6615. doi:10.1073/pnas.84.19.6611.
[40] L. Craig, R.K. Taylor, M.E. Pique, B.D. Adair, A.S. Arvai, M. Singh, et al., Type IV pilin structure and assembly: X-ray and EM analyses of Vibrio cholerae toxin-coregulated pilus and Pseudomonas aeruginosa PAK pilin, Mol. Cell. 11 (2003) 1139–1150. doi:10.1016/S1097-2765(03)00170-9.
[41] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., UCSF Chimera - A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612. doi:10.1002/jcc.20084.
[42] Ref. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: Dassault Systèmes, 2015.
[43] J. Zhang, S.M. Lewis, B. Kuhlman, A.L. Lee, Article Supertertiary Structure of the MAGUK Core from PSD-95, Struct. Des. 21 (2013) 402–413. doi:10.1016/j.str.2012.12.014.
[44] G. Kleiger, A. Saha, S. Lewis, B. Kuhlman, J. Raymond, NIH Public Access, 139 (2010) 957–968. doi:10.1016/j.cell.2009.10.030.Rapid.
[45] S. Chaudhury, S. Lyskov, J.J. Gray, PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta, Bioinformatics. 26 (2010) 689–691. doi:10.1093/bioinformatics/btq007.
[46] R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement, Acta Crystallogr. Sect. A. 47 (1991) 392–400. doi:10.1107/S0108767391001071.

Xiaotong Zuo

+1 (667) 212-1771 • xiaotongzuo@gmail.com

EDUCATION

- **The Johns Hopkins University** — **Baltimore, MD**
  Master of Science in Engineering, Chemical and Biomolecular Engineering — *Sep. 2014 – Aug. 2016*
  Coursework: Computational Biology and Bioinformatics, Application of Molecular Evolution to Biotechnology, Interfacial Science, Statistic Mechanics
- **Fudan University** — **Shanghai, China**
  Bachelor of Science, Biological Science — *Sep. 2009 – Jun. 2013*
  Non-degree Certificate, Economics — *Sep. 2011 – Jun. 2013*
  Awards: 2 x Overall Academic Excellence Award
  Coursework: Statistics, Biochemistry, Genetics, Microbiology, Cell Biology

PROFESSIONAL EXPERIENCE

- **Simulation of Protein Structure** — **Baltimore, MD**
  Researcher, The Johns Hopkins University — *Dec. 2014 – May 2016*
  ○ Simulated the Type IV pilin protein with glycans. Obtained near-native protein structures by the Monte Carlo Search. Analyzed the virulence deduction caused by the glycosylation of the Type IV pilin protein.
  ○ Improved the performance of *FloppyTail* Protein Modeling Algorithm, so that the goal of simultaneously modeling of both proteins and glycans was achieved, 42% computing time was saved as well.
- **Bioinformatic Annotation of SNPs** — **Baltimore, MD**
  Course Student, The Johns Hopkins University — *Mar. 2015 – May 2015*
  ○ Built a functional genomic decision tree that could systematically distinguish mechanisms for SNPs from GWAS studies done on specific diseases.
- **Protein Structural Analysis** — **Beijing, China**
  Researcher, Tsinghua University — *Oct. 2013 – May 2014*
  ○ Purified R233H crystallin mutant protein by applying AKTA purifier, size-exclusion, ion and Ni-NTA chromatography. Applied biophysical methods (UV, Fluorescence, Circular Dichroism, *etc.*) to analyze its misfolded protein structure and biophysical properties (side-chain packing and refolding mechanism).
  ○ Tested a new drug *L*anosterol which successfully reversed the misfolded crystallin protein aggregation in cataract disease. This work was published in Nature (doi:10.1038/nature14650).
- **Genetic Mapping of Enpitope Motif** — **Shanghai, China**
  Researcher, Fudan University — *Mar. 2015 – May 2015*
  ○ Successfully obtained the smallest epitope motif within the length of 8 amino acids of Foot-and-Mouth Disease Virus by using plasmid recombination, transformation, protein induced expression of segmented epitope motifs and western-blot.

SKILLS

- **Bio-experimental Skills:** protein expression and purification, chromatography, spectroscopy, genetic and protein engineering, plasmid recombination, PCR, western-blot
- **Programming Languages:** Python, R, MATLAB, Linux Shell, SQL, HTML, CSS, JavaScript
- **Softwares:** Microsoft Office, Github, PyMol, PyRosetta, LaTeX