**The modeling of Pilus Glycan**

**Abstract**

**Acknowledgements**

**TABLE OF CONTENTS**

**Abstract**

**Acknowledgements**

**List of Tables**

**Glycan type**

**Sasa**

**List of Figures**

**Chapter I: Introduction**

**Chapter II: Methods**

**Chapter III: Results**

**Chapter IV: Discussion**

**Chapter IV: Conclusion**

**References**

**List of Tables**

**List of Figures**

**Chapter I: Introduction (8 pages)**

1. **Type IV Pilus,Pilin, PilA, ACICU, M2 (2 pages)**

**Pilus**

**infection**

**Major vs minor**

**Pilus function**

**classification，n.m,n.g,a.b…**

**How to evade immune system? Variable come from: Seq diversity and glycosylation**

**Class I and ii pilus (N.m specific), 算是补充论证以上两点。**

**Glycosylation process**

**ACICU**

**M2**

**Pilus structure**

Type IV pili are extracellular filamentous adhesive appendages expressed by both Gram-negative (3,4), and Gram-positive (5-7) bacteria, as well as archaea (8). They are primarily assembled from protein subunits, called pilin (1). Type IV pilins are small (~7-20 kDa) structural proteins with a conserved hydrophobic alpha-helical N terminus that is both a transmembrane domain and a protein-protein interaction domain.

Major and minor pilin

In contrast, within a given species, the minor pilins are typically well-conserved. Only the major pilin is highly variable (13-15) and then only in those regions left exposed in the assembled pilus (16).

Type IV pili provide several properties to the bacteria including twitching motility (9), horizontal gene-transfer (10), host-cell adhesion (11) and microcolony/biofilm formation (12). Besides these properties, type IV pili are also the main virulent factor of the bacteria. The efficacy of the bacteria to proliferate in the blood during productive infection depends on its ability to evade type IV pili specific antibodies. There are two sources of pilin virulence. First of all, the antigenicity of pilins can be diversified by transferring DNA from the silent cassettes to the expression locus to generate multiple different antigens. Secondly, another source of antigenic cariation is post translational modification and in particular glycosylation. It was recongnized that the class II pilins in *Neisseria meningitidis* [11, 12] are lack of gene conversion, but they can successfully evade the immune system, which proved that the main source of virulence of these pilins might be glycosylation.

Pilus

Pilin

PilA

from *A. baumannii* ACICU and *A. nosocomialis* M2, ACICU, M2

*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 (42)

A.n=A.b

*A. baumannii* M2 (32,43,44) was isolated in 1996 from a hip infection of a patient at Cleveland MetroHealth Systems (Cleveland, OH).

*N. gonorrhoeae* Type IV pilus filament (Protein Data Bank ID ‘2HIL’)

*P. aeruginosa* PAK pilin (donate seq, different disulfide bond positions comparing to ACICU)

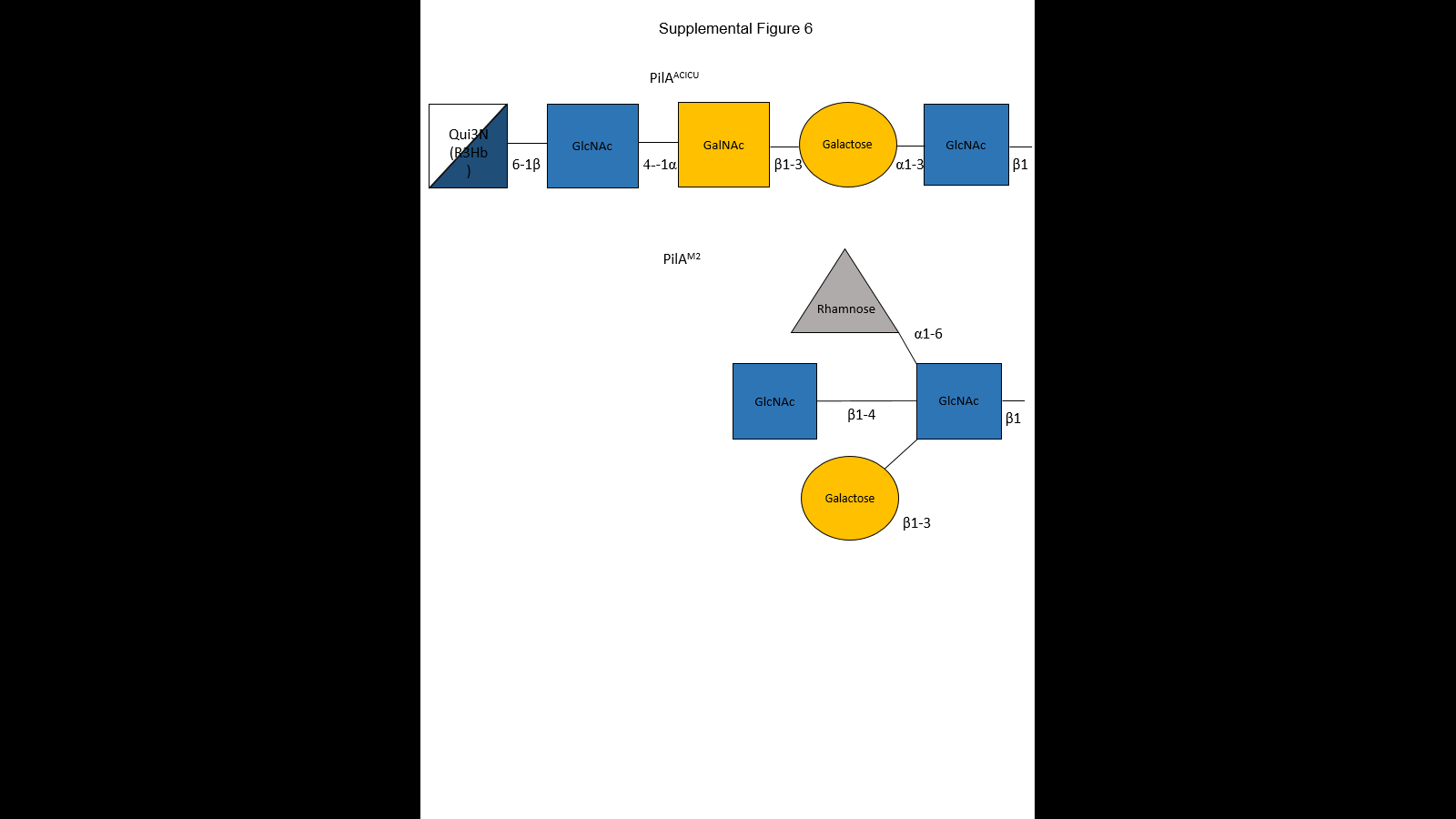
PilAACICU contains two disulfide bonds; one, between residues 123 and 136, is the C-terminal disulfide bond also found in PilAPAK, which is nearly universal in type IV pili from Gram-negative bacteria and the other, between residues 74 and 91, spans the first two strands of the central β-sheet (Figure 2B). However we note that this additional disulfide bond in PilAACICU (relative to PilAPAK) does not result in any substantial rearrangement of the protein backbone.

1. Virulence (2 pages)

because pili are prominent surface structures, they also are occasionally targets of host defenses, and alternating induction and shutoff of pilus expression allows for successful evasion of the immune system.

Glycan Modification will influence the …

1. Main goal of this thesis
2. Glycan Modeling
   1. As a major type of the post-translational modifications of proteins, glycosylation plays important on 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 (Benz and Schmidt, 2002; Szymanski and Wren, 2005). In particular, most glycoproteins of bacterial pathogens are either surface localized or trafficked for secretion and appear to influence interactions with the host. Prime examples of pili among Gram-negative species include pilin subunits of P. aeruginosa (Castric et al., 2001) and neisserial type IV pili (Tfp) (Stimson et al., 1995). In many instances, glycosylation-defective mutants have been shown to be attenuated in virulence- associated properties and colonization (Szymanski et al., 2002; Grass et al., 2003; Schirm et al., 2003; Hendrix- son and DiRita, 2004; Arora et al., 2005).
   2. Importance of glycosylation for proteins and for evading immune systems
      1. Glycosylation facilitates solubilization of pilin monomers and pilus fibres. The removal of the terminal glycan residues of the modification had no significant effect on the adhesive abilities of pili (Stimson et al. 1995, Marceau et al. 1998)
      2. The role of pili in mediating bacteriacal interaction has been well characterized, especially in capsulated bacteria.
   3. Glycan



The terminal glycan residues have been linked to the terminal serine (Ser 139) of the ACICU pilin. This glycan chain was comprised of an N-Acetylglucosamine (GlcNAc), a galactose, an N-Acetylgalactosamine (GalNAc), an N-Acetylglucosamine (GlcNAc) and a 6-deoxy glucose called quinovose with its R3Hb side chain.

For the M2 pilin, the terminal glycan residues have been linked to the terminal serine (Ser 136) of the M2 pilin. Unlike the string-like glycan chain of pilin ACICU, the glycan chain of M2 had a compact globular structure started with an N-Acetylglucosamine (GlcNAc). An galactose, an N-Acetylglucosamine (GlcNAc), and a L-Rhamnose were linked to the GlcNAc.

* + 1. In both cases…FloppyTail Algorithm
  1. PyRosetta modeling introduction
     1. PyRosetta is a Python-based interactive platform for the protein structure prediction and design. It can manipulate the conformations of proteins and ligands, calculate energies of structures, and
  2. Stuff specific for glycan modeling, non-low-res, different energy weight and movemap?

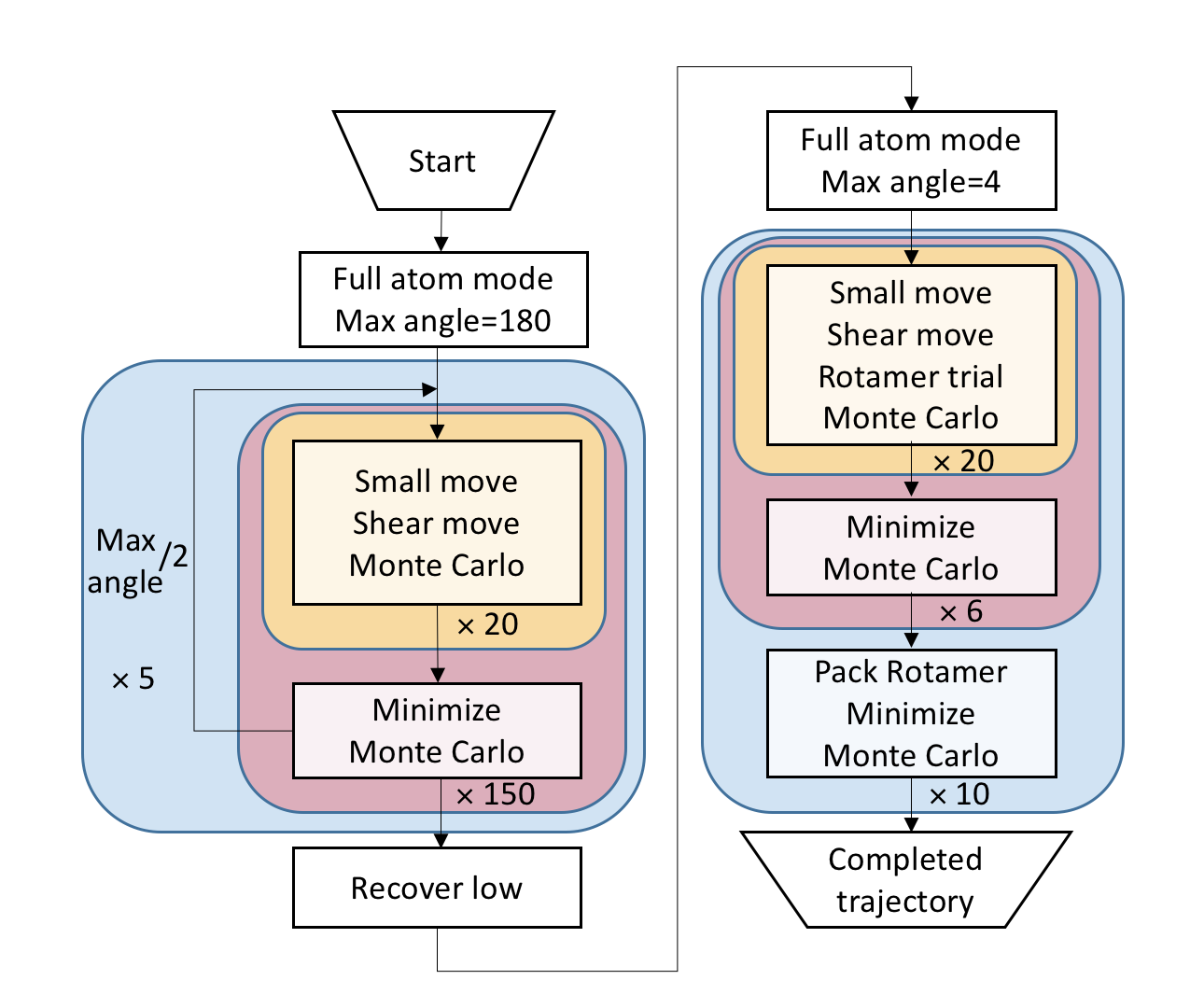
**Chapter II: Methods**

To measure the extent to which C-terminal glycosylation of PilAACICU and PilAM2 would mask the pilin protein from binding, we modelled the full length pilins and pilus fibers and measured the effect of glycosylation on the accessible surface area of each protein in its native context. We modelled an ensemble of each glycan based on the repeating unit of the major polysaccharide glycan and minimized each structure using PyRosetta. The ten best-scoring glycan conformations were then combined to approximate the native conformational ensemble. We then measured the differences in accessible surface area using a 10Å particle probe to approximate the surface area needed for protein binding. The resulting models are shown in Figure 6 and the change in accessible surface area for each protein in Figure 7. 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.

1. *Initial* Structure Generating.

The initial structures of PilA ACICU and M2 were modeled using the full-length *P. aeruginosa* PAK pilin (45). The initial models of the pili were created by superimposition onto a model of the *N. gonorrhoeae* Type IV pilus filament (Protein Data Bank ID ‘2HIL’) (47), and adjustment of the N-terminal helix position to eliminate clashes between subunits. The resulting models then underwent rigid-body minimization by UCSF Chimera (66). And adjusted in the Discovery Studio Visualizer (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: Dassault Systèmes, 2015.).

1. Glycan Modelling Algorithm



* 1. Low-resolution Gradient-based modeling

Monte Carlo

[Chapter 4 (Monte Carlo methods) of M. P. Allen & D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, 1989.

Z. Li & H. A. Scheraga, “Monte Carlo-minimization approach to the multiple-minima problem in protein folding,” *Proc. Natl. Acad. Sci. USA* **84**, 6611-6615 (1987).]

Torsion angle

Rotamer and repacking

Minimization

Score: reu

* 1. High-Resolution Rotamer Repacking
     1. Rotamer is …
     2. The low-resolution modeling stage was followed by a high-resolution refinement stage, where the lowest energy structure from the low-resolution stage was refined by side chain packing, swapping rotamers of a random residue, torsion angle perturbation and test each move using the Metropolis criterion.

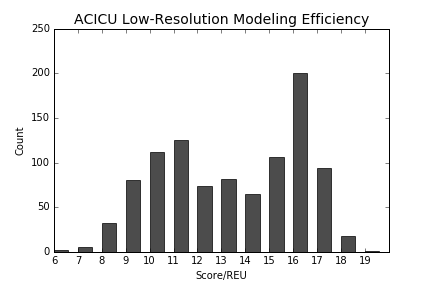
[figure algorithm]

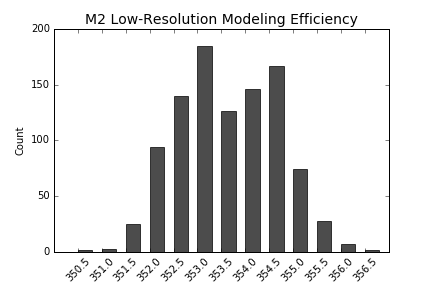
The glycans attached to PilAACICU and M2 were modeled using PyRosetta protein structural modeling suite (67). And then the glycans were modeled using the *FloppyTail* Algorithm (68). The protocol consists two parts. In the low-resolution part, a random perturbation of the torsion angles was applied. The structures were then refined in the high-resolution part by applying a more precise perturbation of the torsion angles, the side-chain packing and the minimization. With this protocol, 6000 structures of PilAACICU and 6000 structures of PilAM2 were generated. And 1000 structures without glycans were generated.

**Results**

1. Low-resolution modeling efficiency (1 figure)

To examine the searching efficiency of the low-resolution modeling step, there were 1000 decoys for ACICU and M2 generated after the low-resolution sampling ended(Fig.).





After the gradient-based low-resolution stage, the score of ACICU ranged from 6 to 19.

The score of M2 was ranging from 350.5 to 357, which is a smaller range comparing with ACICU. The solid hydrogen bonds and short, globular glycan chain is responsible for the score ranging difference.

For both ACICU and M2, the low-resolution modeling step was efficient to find a low energy structure among the energy landscape.

1. Score with & without glycans (Figure 2(ACICU, M2) abc\*WITH,WITHOUT)
   1. Lowest
   2. Top 10
   3. Random 20

Interestingly, there were several good structures which were apparently different with others. Unlike other good structures, their glycan side chains changed to an almost opposite direction at GLY 137 of the protein tail. The GLY 137 of the tail region was flexible due to its highly variable torsion angle, thus resulted in totally different structures. However, there was a score compensation for direction change, which was approximately 3 to 4 REU. Although the scores were slightly higher, these structures might also be favored conformations.

1. Rmsd+low20decoy rmsd (Figure 2)
2. Surface area with & without glycans (Figure 2, table 1)
   1. Top 10 surface
3. Hbond
4. Resi by resi energy
   1. M2 and ACICU comparison
   2. Energy well for sugar\_bb term (is resonable)
   3. Dunbrack energy

**Conclusion**

**Reference**