First-in-class *Pseudomonas aeruginosa* Rational Vaccine Design and Characterization by *ilmm* Molecular Dynamics

Modestas Filipavicius, Clare Towse, Valerie Daggett
Department of Bioengineering, University of Washington

Abstract

P. aeruginosa is a common bacteria that infects thousands of hospital visitors each year. It attaches to host cell membrane via a Type IV PAK pili, which is an oligomer of the 144 amino acid long long monomer protein. Here we show the development of the first-in-class small peptide vaccine targeting P. aeruginosa pilin epitope. The antigenic epitope spans 15 residues, and it was stabilized by performing a series of full length protein truncations and mutations at key salt-bridge forming residues, followed by exhaustive MD simulations with *il*mm. The final product design (ID3) is only 50 amino acids long, exhibits stable secondary structure, minimized non-native interactions, and sub-angstrom RMSD and RMSF values at the binding epipe.

Introduction

P. aeruginosa is a common hospital bacteria that infects thousands of patients each year. Its resilience comes from the incredible ability of attaching to abiotic surfaces such as stainless steel via PAK pili (Giltner et al., 2006). It breaches the host cell by first attaching PAK pili to cell surface and then penetrating the cell membrane.

Type IV Pili are filaments made up of repeating pilin proteins, C-terminal of which has a receptor-binding site that mediates adherence to host epithelium (Keizer et al., 2001). The C-terminus of type IV pilus, also known as receptor binding domain (CTRBD), could be targeted by antibodies, which upon strong binding will prevent pili adhesion to host cell. Upon injection of stable protein (vaccine immunogen) the body's immune response (B-cell and T-cell maturation, random mutagenesis events) will produce antibodies that target the antigen's epitope.

Designing a small 30-50 residue-long peptide vaccine is advantageous in terms of drug delivery (Li et al., 2014). Also, peptide vaccines can be used for induction of broad-spectrum immunity against multiple serological strains same bacterial family by targeting epitopes conserved between different serovars/strains of a pathogen. Smaller peptides are preferred, since they will diminish the need for a carrier molecule, add chemical stability and reduce non-specific immune response.

To do that, we need to identify the epitope that is targeted by these broadly neutralizing antibodies (across different species) and design more stable pilin protein.

BD Sykes group has created a monomeric pilin that retains oligomer's antigenicity profile despite removing the first 28 N-terminal residues (Keizer et al., 2001). Tertiary structure was determined by NMR. Antigenic epitope of PAK pilin is well conserved within residues 128-144 (Lee et al., 1996). Pilin interacts with target cell receptors via S131, Q136, I138, P139, G141 and

K144, therefore an antigen design should focus on presenting these residues intact to the immune system.

Here we used Foldit, Modeller and ilmm to rationally redesign the pilin protein and subject it to molecular dynamics simulations (Beck et al., 2008). Special emphasis was on designing a robust epitope stabilization by adjacent alpha-helix and beta-sheet, through evaluating different mutants in these secondary structures. We have successfully determined a first-in-class 51-residue peptide that will be used as a basis for future.

Methods

Designs of Pilin fragments

The first insights into 1DZO native dynamics were acquired in Rosetta Foldit, by freezing the epitope (K104-R120) and applying course minimization of side chains, automatic mutagenesis, and repacking protocols. On native 1DZO, we tried several mutations: P42A successfully stabilized the alpha-helix, T101R had to enhance solvation of the beta-sheet.

Then I explored truncations. Fragment was first made by deleting the V40-P109 region, which resulted in two chains: D27-S39 and K110-R144 (Figure S1.A).

Why this fragment: space is key, epitope interacting partners are located in the C-terminal half of alpha helix and two out of four sheets, so we cut the rest. There were no hydrogen bonds or salt bridges between the epitope and the other two beta-sheets.

To connect the two chains, we inversed the numbering on alpha helix to S27-D39, mutated S27D (to cap the helix) and connected D39 with K110 by either a G40G41G42 or MGF linker. Linker was made with PyMOL Editing mode, one residue at a time from K110 towards D39 (Figure S1.B). The linker loop was refined with Model/Refine tool in Chimera, that uses Modeller

algorithm, and model with the lowest DOPE score was chosen (DOPE = -0.59) (Figure S1.C). The numbering was reset at D1, A2, L3, … , C49, S50, R51.

51 Residue Fragment Mutations:

1. No mutations were made in the hydrophobic core, since increased stability following core repacking rarely happens, especially without causing backbone movements (Mooers et al., 2003). Important here, since avoiding the epitope perturbation is the main goal.
2. S39D (S1D on new numbering) on N-terminus of alpha-helix. Asp interacts with water and other charged residues on helix, potentially forms helix-capping bridges.
3. W34D creates a salt bridge with R27, and therefore links the two beta-sheets together, stabilizing the epitope.
4. I22S is on the solvent surface and creates polar interactions with water.
5. GGG and MGF flexible linkers. Both are commonly seen in nature, the former is more flexible and unstable, while MGF interacts well with hydrophobic environment through buried Met and Phe residues (Towse, personal communication).
6. G7A stabilizes the helix by replacing a highly flexible and therefore helix-breaking Gly with high alpha-helix propensity Ala.
7. S5R creates a salt bridge with D1 following the i+4 rule, thus capping alpha-helix.
8. E13P is located right after the helix, and it breaks it. No conformation exploring.

All fragment constructs are shown in Figure 1.

Molecular Dynamics with illm

Simulations were done with *in lucem* molecular mechanics (*ilmm*) (Beck et al., 2008). Design pdb files were cleaned up with clean_ilmm_pdb.pl, minimized with 1000 steps of SD minimization protocol, hydrogens were added and water box that surrounds protein by at least 10 Å was added (Levitt et al., 1995). Temperature was set to 310 (physiological T), water density was adjusted to 0.985506 g/ml and water was added at most within 1.8A from the protein surface. Next, the solvated protein went through a 1000-step SD minimization at a given T, plus 1 ps of raising T by 40K.

Checkpoint.mdr file was created with velocity, position and acceleration terms for all simulation atoms. Simulation length here is 5 ns for ID, ID1 and ID2; 6 ns for ID3 and 4; and 10 ns for 1DZO. We used the standard 2 fs timestep, structures were saved every 1 ps for later analysis.

MD simulation analysis

Overall and individual residue RMSD and RMSF were calculated every 2 fs, with the MD simulation starting structure taken as a reference. 3D structure snapshots were recorded to individual .pdb files every 100 ps with Filet tool.

Secondary structure interactions and motifs were classified as loops, 3/10, alpha and pi helices, and alpha, beta, sheets or salt bridges, and visualized with DSSP package as outlined in McCully et al., 2013). VCONT analysis revealed the contacts between residue side chains, which were classified as non-native, if they did not appear in the original structure. Asymmetric contacts plot has non-native contacts on the top left and native on bottom right, colored according to fraction of time in contact. Total contacts plot reveals native contacts before the simulation.

Results

RMSD backbone movements

The initial full length pilin designs had RMSD values in the range of 2.0 – 3.0 Å (Fig. S1). This is not surprising in the absence of backbone movement. However, the truncated fragments with two chains held together by disulfide bond, FWT and FMD, had RMSD in the range of 4.0 – 5.0 Å.

For our second fragment design iteration we tested five constructs. These designs (ID – ID4) produced a variety of RMSD values (Figure 2). ID3 on average was 2.8 Å for the last 2 ns of simulation, followed by ID (3.3 A), ID2 (4.7 A), ID1 (5.4 A) and ID4 (6.2 A). ID3 matched the RMSD of IDZO crystal structure.

The total RMSD of the whole structure is a rough indicator of overall stability. To gain better insights into local fold movements we performed RMSD by residue analysis (Figure 3). As expected the N-terminal alpha-helix D1-12F exhibited huge movements as indicated by RMSDs in the 4-12 Å range. Nevertheless, ID3 and ID had relatively low values, showing that the helix was stabilized. Meanwhile ID4, which differs from ID3 only by a single helix ending E13P mutation, had the most unstable helical region. Interestingly, I22S and W34D mutations in ID1 on the beta sheet, right before the epitope, significantly destabilized the protein compared to ID1.

Most importantly for our vaccine design, ID3 had least movement in the key epitope region defined by K35-R51 (K128 – R144 in 1DZO), with residues 38 to 51 forming a loop, which are notoriously hard to stabilize. ID3 showed a low value of 1.5 Å for R51 (compared to 3.0 Å in other 4 designs), indicative of Arg stabilization by another residue. Only two residues, F44 and I45, were more stable in ID2 and ID designs.

ID3, with a MGF loop, and stabilized alpha helix with Asp1-Arg5 salt bridge and G5A mutation, shows least overall fluctuations in 3D space and especially the epitope region (Figure 4). Amazingly, a single helix breaking residue change in ID3 (E13P), which correlates to ID4, produces the most variable structure, especially in the alpha-helix region, with RMSF values above 6 A. Not surprisingly, the ID2 (MGF linker) was less flexible than ID1 (GGG linker) due to hydrophobic burial of M14 and F16 into A11, F11 and I22.

DSSP analysis

DSSP algorithm revealed that the secondary structure of initial full protein designs was largely intact, however the fragments lost two beta-sheets and beta-bridges in the epitope (Fig. S2). Our fragment designs sought to prevent the destruction of epitope's secondary structure. As seen in Figure 5, only ID3 retained the native (1DZO) beta sheets at residues 25-28 and 34-37, and beta bridges at K47 and S50. The native three residue, E42, Q43, F44 3/10 helix was apparent only in 30% of simulation time. Despite absence of alpha-helix, this ID3 region was stabilized by S5R mutation. Movies generated from by ilmm Filet algorithm collaborate the above findings (*see* Movies 1-5).

VCONT analysis

Initial full protein designs did not introduce non-native contacts, while very unstable fragments made new transient interactions (Figure S4). When we focused on improving the 51 aa-long fragment designs, there was a special emphasis on not introducing new bonding partners between truncation destabilized alpha helix and the epitope.

Unlike 1DZO and ID3, the ID, ID4, ID2 and especially ID1 displayed a multitude of new non-native contacts within the alpha-helix and with epitope (Figure 6). ID3, however, had the least new non-native interactions, as most of them matched 1DZO epitope-alpha-helix contacts

between residues A4 and P46-C49; R10 and Q40, I45-K47; A11 and Q40. The only exception being R5 and P46, K47 salt bridges, which are due to S5R mutation. This salt bridge probably helped to stabilize the helix to the epitope.

Alpha helix was completely disordered and randomly coiled in ID1, ID2 and ID4 for majority of the simulation. It made aberrant and non-native hydrogen bonds within itself forming Pi helices (Figure 5, Figure 6, Movies 2, 3 and 5).

Discussion

From this study we have determined that ID3, a redesign of PAK pilin from P. aeruginosa strain K, is the most stable antigen and therefore best candidate for further vaccine design. Main goal here was to create a small peptide antigen, as opposed to using the full protein, even if that meant sacrificing native folds. For example, our best design, ID3 completely lacks alpha helical structure. However, through clever helix capping we have intra-connected the disordered helix with D1-R5 salt bridge.

Antigen ID3 exhibited high relative stability at physiological temperatures, compared to other designs, which was partially due to enhanced alpha-helical and epitope interactions. S5R mutation created a salt bridge to P46 and K47, which stabilizes alpha-helix and epitope interaction. This would not be possible without the incorporation of a long charged residue like Arg. The insight to put Arg at that position came from exploring rotomer space with with Dynameomics Rotomer Library (Towse et al., 2016). Opponents might claim that it interferes with the epitope presentation. We must determine this experimentally.

Some promising design did not deliver. For example, the mutations I22S and W34D on the surface of beta-sheets of ID1 and ID2 exchange a hydrophobic residue for a polar or charged residue, respectively. However, we did not consider that protein folding is also driven not only by

burial of hydrophobics into the protein core, but also by release of water from exposed nonpolar groups. Our initial expectations were to reduce antigen peptide aggregation, since numerous hydrophobic residues on the protein surface during the dynamic folding/unfolding thermodynamic equilibrium post a risk of replacing intramolecular hydrophobic packing with intermolecular hydrophobic packing, or aggregation.

Meanwhile, the inclusion of MGF flexible linker instead of GGG helped to stabilize the linker loop. In ID2 (with MGF linker) the region 14-16 was less flexible than ID1 (GGG linker) due to hydrophobic burial of M14 and F16 into A11, F11 and I22 (Figure 3).

One recurring problem was Asp1-Ala11 alpha helix collapse during the ID, ID1, ID2 and ID4 simulations (Figure 5). Hydrogen bond formation between amide hydrogen (donor) and carboxy group of amino acid 4 residues earlier (acceptor) is driven by 1 kcal/mol enthalpic contribution, and is the primary force behind alpha-helix formation (Marshall et al., 2002). Meanwhile, random coil is favored by entropy, it costs around 1.5 – 2 kcal/mol per residue to adjust dihedrals from random conformations to those suitable for helix.

We first tried to stabilize it by adding an N-terminal Asp or Glu. According to Pace and Scholtz (1998), uncharged Glu is the second best helix former, right after Ala, and much better than Asp (by 0.27 kcal/mol). However, both Glu and Asp have lower (by around 0.25 kcal/mol) helical propensities in their charged states, which are the most physiologically relevant states, higher than pH=4. Difference between Asp and Glu could be explained by the fact that Glu has a shorter side chain, which means smaller entropic penalties when new interactions are formed when random coil turns into a helix (Pace and Scholtz, 1998).

In an alpha-helix, the first four >N-H groups and last four >C=O groups necessarily lack intrahelical hydrogen bonds (Anil et al., 2004). Instead, such groups are often capped by

alternative hydrogen bond partners. Charged side-chains may interact with the helix dipole, which consists of partial positive and negative charges located at the N and C termini, respectively, of each a-helix (Marshall et al., 2002). This dipole arises from unsatisfied hydrogen bonds at each of the helix termini (3 of them). Interactions between side-chains and the helix dipole have been demonstrated to impact the stability of both model peptides and proteins (Anil et al., 2004). The design rules stipulate no positively charged residues at N-terminus and no negatively charged residues on C-terminus. To fix helix capping issue, we introduced a mutation S5R, so that positive Arg5 forms a salt bridge with Glu1 following the i+4 rule (ID3).

Next, we observed that G7 is moving and kinking a lot in ID-ID3, but after mutation to A7, these movements stopped and helix was stabilized. Gly is a known helix breaker, while Ala has the best propensity for alpha helix (Pace and Scholtz, 1998). Finally, we mutated E13P, which is right after the original alpha helix ends. Pro is a great helix breaker, and since it's upstream of the alpha helix, it will prevent anything above it from acquiring helical conformation, it will cap helix's C-terminus.

In our future work we will try to address further alpha helical stabilization, explore alternative helix to beta-sheet linker sequences by an extensive Modeller scan and test if the ID3 epitope could elicit a broadly neutralizing antibody response in infected patient blood titers. All in all, here we used Foldit, Modeller and *il*mm to rationally redesign the pilin protein and subject it to molecular dynamics simulations (Beck et al., 2008), which helped us to stabilize the 16 residue epitope within a 51-residue peptide that will be used as a basis for future vaccine.

References

Beck DA, Daggett V. Methods for molecular dynamics simulations of protein folding / unfolding in solution. Methods, 34: 112-120, 2004.

Beck, D., McCully, M., Alonso, D. & Daggett, V., 2000-2016. *in lucem* molecular mechanics software, University of Washington

Giltner, Carmen L., et al. "The Pseudomonas aeruginosa type IV pilin receptor binding domain functions as an adhesin for both biotic and abiotic surfaces." *Molecular microbiology* 59.4 (2006): 1083-1096.

Lee, Kok K., et al. "Anti-adhesin antibodies that recognize a receptor-binding motif (adhesintope) inhibit pilus/fimbrial-mediated adherence of Pseudomonas aeruginosa and Candida albicans to asialo-GM1 receptors and human buccal epithelial cell surface receptors." Canadian journal of microbiology 42.5 (1996): 479-486

Levitt M, Hirshberg M, Sharon R, Daggett V. Potential-Energy Function and Parameters for Simulations of the Molecular-Dynamics of Proteins and Nucleic-Acids in Solution. Computer Physics Communications, 91: 215-231, 1995.

Li, Weidang, et al. "Peptide vaccine: Progress and challenges." *Vaccines* 2.3 (2014): 515-536.

Pace, C. Nick; Scholtz, J. Martin (1998). "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins". Biophysical Journal 75. pp. 422–427.

PyRosetta Lecture series with Jeff Gray (JHU)

Scouras A.D. and Daggett V. The dynameomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. Protein Science 20:341-352, 2011

Towse, Clare-Louise, et al. "Insights into Unfolded Proteins from the Intrinsic ϕ/ψ Propensities of the AAXAA Host-Guest Series." *Biophysical journal* 110.2 (2016): 348-361.

Figure legends

Figure 1. Mutations featured in five fragment designs (ID-ID4). Mutations showed in pink, epitope in orange. Sticks represent six residues necessary for pilin-host membrane interaction. Picture created with PyMOL, sequences aligned with Chimera.

Figure 2. RMSD of five new fragment redesigns. ID – has GGG connecting loop; ID1 – has GGG connecting loop and I22S and W34D mutation; ID2 – same as ID1, but GGG is replaced with MGF at G14M and G16F; ID3 – has MGF loop and new mutations at alpha-helix G7A and S5R; ID4 – same as ID3, but with E13P mutation which breaks the alpha helix. ID, ID1 and ID2 MD simulations ran for 5 ns, ID3 and ID4 for 6 ns, while 1DZO for 1 ns.

Figure 3. RMSD values by residue. ID, ID1 and ID2 MD simulations ran for 5 ns, ID3 and ID4 for 6 ns, while 1DZO for 1 ns.

Figure 4. RMSF data for all five mutations reveals that ID3, with a MGF loop, and stabilized alpha helix with Asp1-Arg5 salt bridge and G5A mutation, shows least overall fluctuations in 3D space and especially the epitope region. ID – has GGG connecting loop; ID1 – has GGG connecting loop and I22S and W34D mutation; ID2 – same as ID1,but GGG is replaced with MGF at G14M and G16F; ID3 – has MGF loop and new mutations at alpha-helix G7A and S5R; ID4 – same as ID3, but with E13P mutation which breaks the alpha helix.

Figure 5. DSSP graph reveals a stabilized epitope in ID3 structure. ID4 has a stronger 3/10 helix character and less beta-bridge.at positions K47 and S50. Despite stable epitope region, ID3 had no helix at residues 2-11, but instead this region was stabilized by a salt bridge at R5. In ID4, an alpha helix formed but that destabilized the two beta sheets and possibly our epitope.

Figure 6. VCONT secondary structure. tot_contacts were generated by superimposing asymmetrical and ref_totatal plots.

Figure 7. VCONT asymmetric contacts. In ID3 strong interactions between residue 4 and 46-49, 5 and 46, 47; 10 and 40, 45-47; 11 and 40. Cap worked because 5 is in contact with 1, 2, 3 and 10 all the time.

Figure S1. Three stage (A-C) fragment design starting from a full length pilin 1DZO sequence.