Computational Modeling of *Francisella novicida* Target Protein OppA with LL-37 Antimicrobial Peptide

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Sarah Parron
Master of Arts
Virginia Polytechnic and State University, 2011
Bachelor of Science
Virginia Polytechnic and State University, 2009

Director: Monique van Hoek, Professor
Department of Bioinformatics and Computational Biology

Fall Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my husband, Jon, ever supporting of my academic adventures, and my beautiful children, Jedidiah and Miriam, whose unwavering love helps me to conquer all.

ACKNOWLEDGEMENTS

I would like to thank Dr. Amy Smith for her support in completing my thesis work, and her inspiration as a talented woman wearing many hats in a challenging career. I would like to thank Dr. Monique van Hoek for her amazing abilities to mentor with endless encouragement and support and her inspiring skills in organization, writing, multitasking, and lab management. I would like to thank Dr. Iosif Vaisman for seeing my potential in bioinformatics, supporting me throughout my program and enabling my degree to be completed remotely.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Antimicrobial peptide .....AMP
Angstroms .....Å
Protein Database .....PDB
Visual Molecular Dynamics ..... VMD

ABSTRACT

COMPUTATIONAL MODELING OF *FRANCISELLA NOVICIDA* TARGET PROTEIN OPPA WITH LL-37 ANTIMICROBIAL PEPTIDE

Sarah Parron, M.S.

George Mason University, 2021

Thesis Director: Dr. Monique van Hoek

As a potential biological threat agent, the virulent bacterium *Francisella tularensis* remains an important topic of research. Understanding the mechanisms of action of antimicrobial peptides on *F. tularensis* is crucial to developing alternative treatments in the event of engineered or natural antibiotic resistance. The antimicrobial peptide LL-37 produces bactericidal effects in *Francisella* species and has demonstrated interaction with the membrane as well as periplasmic and intracellular proteins. One *Francisella* protein identified as an LL-37-binding protein is the periplasmic oligopeptide permease (Opp) complex protein, OppA.

To perform computational structural modeling of the *Francisella*'s oligopeptide substrate-binding protein, crystal structure templates of homologs from several species were used to create models of *F. novicida*'s OppA sequence. To characterize the interaction of OppA and LL-37, computational docking was performed on the LL-37

fragment, KR-12 and then on the longer fragment LL-20. These studies revealed the favorable binding of KR-12 and LL-20 within the cavity of the OppA protein. This computational modeling supports the experimental data. These studies also revealed significant insight on the effect of open or closed protein conformations on the computational docking process. This approach illustrates the power of computational docking to reveal information about potential bacterial protein targets of antimicrobial peptides.

INTRODUCTION

In the last century, the discovery and developments within microbiology have led to new capabilities of genetic manipulation and engineering. While these techniques pave the way for potential opportunities for stopping antibiotic resistant bacteria, curing diseases, and understanding biological mechanisms, there is also the possibility for weaponizing bacteria. Although *Francisella tularensis* currently has low infection rates and is treatable with antibiotics, it is an example of a potential biothreat if genetically engineered. To combat potential problems like bioweapons and antibiotic resistant bacteria, research has extended to natural or synthetic molecules that induce bactericidal effects through cellular membrane disruption or compromising intracellular targets. The use of antimicrobial peptides (AMPs) as a potential treatment against engineered strains is promising because many AMPs are natural components of host immune systems and are often imported into cells via peptide transporters, like ATP-binding cassette (ABC) transporters. Understanding the mechanisms of action of these AMPs including cell entry and interaction with selected intracellular targets is crucial to developing AMPs into a viable treatment option.

Antimicrobial Peptides

Antimicrobial peptides are a diverse class of naturally occurring molecules that are produced as a defense mechanism in all multicellular organisms and many

prokaryotes. Most AMPs utilize several mechanisms that can kill bacteria, yeasts, fungi, viruses, and cancer cells. Some AMPs in higher eukaryotic organisms include immunomodulatory activities and are referred to as “host defense peptides” (HDPs) [1]. The human host defense peptide LL-37 is part of the cathelicidin family of HDPs and is highly researched due to its diverse antimicrobial capabilities on both gram-positive and gram-negative bacteria. Cathelicidins are defined by a conserved pre-pro region that gets cleaved by a neutrophil protease and is not part of the antimicrobial actions. The cathelicidins antimicrobial functions include direct antibiotic activities, orchestrating immune response to infection, modulating inflammation, and promotion of wound healing [2]. LL-37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES) is a strong cationic peptide with an alpha-helical structure that starts with two Leu residues and is 37 residues long. Unlike other helical AMPs that adopt helical conformation only near biological membranes, LL-37 maintains its helical structure in aqueous solutions at physiological salt concentrations, which is believed to account for its overall capabilities [3]. Due to its strong positive surface charge, LL-37 is highly attracted to negatively charged bacterial membranes, making it a candidate in developing new treatments for infections caused by *Francisella tularensis*. As with many AMPs, LL-37 exhibits two different mechanisms of action: cell lysis through membrane pore formation at higher concentrations, or cell-penetration and binding to intracellular targets at lower concentrations [4]. Further exploration of these intracellular mechanisms will provide support for developing LL-37 into a viable antimicrobial treatment.

Francisella

Francisella tularensis is a gram-negative bacterium that is the causative agent of tularemia, commonly referenced as “rabbit fever”. Tularemia is found in animal populations of rabbits, squirrels, muskrats, rodents, reptiles, and less commonly among water-associated mammals. Tularemia has a history of zoonotic transmission to humans by direct contact with infected animals and mosquito or tick-transmission most often to hunters, farmers, or forest workers [4]. The symptoms associated with tularemia are chills, fever, headache, body ache, enlarged lymph nodes, diarrhea, or pneumonia depending on how the infection occurs. During active years, there are thousands of human cases each year mainly occurring during the summer months throughout Europe, Russia, and the United States [5]. In the early 1900s, the United States saw about 2000 cases per year but now the current U.S. rates are approximately 200 cases per year [5]. Despite the case reduction, the virulent form of *F. tularensis* is categorized as a Tier 1 threat agent by the Center for Disease Control (CDC) due to its high infectivity if inhaled by human lungs and remains an important topic of research [5].

The species *F. tularensis* has four subspecies: *F. tularensis* (Type A), *F. tularensis holarctica* (Type B), *F. tularensis mediasiatica*, and *F. tularensis novicida*. While *F. tularensis ssp. tularensis* is the virulent strain in humans, *F. tularensis ssp. novicida* shares 97% homology with *F. tularensis* and is not infectious in healthy humans. This facilitates easier laboratory experimentation with *F. tularensis spp. novicida* U112 experiments often reflecting the same results for the virulent *strain F. tularensis spp. tularensis* SchuS4. Alternative treatments like antimicrobial peptides are

being explored in case fully antibiotic resistant strains emerge naturally or from deliberate bioengineering.

Francisella **Target Proteins for LL-37**

From experimental research from the van Hoek lab, it is known that LL-37 interacts with two specific *Francisella* proteins: AcpP and OppA. These experiments involved immobilizing tagged LL-37 and binding *Francisella* cell lysate to the immobilized LL-37. Increasing salt washes were done to remove lightly bound proteins and to leave behind tightly bound proteins. The resulting tightly bound proteins were run on an SDS-PAGE gel and protein bands identified by silver staining. Highly abundant bands were cut out of the gel and subjected to mass-spectrometry for identification.

Acyl-carrier proteins (Acp) carry metabolites like fatty-acids for important biosynthetic pathways such as cell growth. Previous experimental research demonstrated binding of AcpP and LL-37, which led to increased antimicrobial effect of LL-37 from within the cytoplasm [3]. These results demonstrate an unstudied mechanism of action of LL-37 for this AMP in *Francisella*, indicating that LL-37 is using intracellular protein targets for cell death, not just membrane disruptions such as pore formation. This is evidence that a transmembrane transportation of LL-37 must occur for intracellular targets like AcpP to be reached.

OppA and ABC Transporter Roles

One major protein importer in microorganisms is the oligopeptide permease (Opp) complex. OppABCDF transporters are formed from five oligopeptide permeases. First described in *Salmonella typhimurium* as a periplasmic transport system able to

translocate peptides, the OppABCDF has experimentally exhibited itself as a transmembrane transporter of AMPs [6, 7].

The OppABCDF transporter consists of five subunits: OppA, the substrate-binding protein (SBP), two homologous integral membrane proteins, OppB and OppC that form a translocation pore, and two homologous nucleotide-binding domains, OppD and OppF that perform ATP-binding and hydrolysis. OppA is a periplasmic protein which is classified within a large superfamily of substrate-binding proteins associated with ATP-binding cassette (ABC) transporters that share similar tertiary structures [8]. In gram-negative bacteria like *Francisella* the SBP binds the peptide substrate in the periplasm and shuttles it to the transporter, in opposition to being membrane-anchored as found in gram-positive species [9].

While the sequence identities of this superfamily are not highly conserved between species (20-35%), data shows that these OppA (or similar proteins) are highly structurally conserved [9]. The SBPs have a common structure of two rigid, conserved lobes with α/β -folds connected by a flexible hinge region that rotates to create open or closed conformations. During the open conformation ligand-binding occurs inside the large cavity and once bound, the closed conformation transports the ligand to the membrane-embedded transporter [1]. Most importantly, this process demonstrates how the large binding cavity accounts for various sized ligands or peptides and requires less specificity for ligand-binding.

Within *F. novicida* the novel interaction of LL-37 and OppA was confirmed when *F. novicida* mutants with knocked out Opp genes were created and tested for LL-37

susceptibility. The results indicated that oppA mutants were significantly more resistant to the AMP than the wildtype, while the other mutants demonstrated similar susceptibility to wildtype [unpublished]. To confirm OppA as a vital mechanism for transmembrane transport, fluorescence microscopy and flow cytometry analysis experiments were performed. Fluorescence microscopy with labelled LL-37 peptides showed an increase of peptide fluorescence within wildtype *F. novicida* and decreased fluorescence of LL-37 within oppA mutants. Flow cytometry analysis supported these results, showing the decrease of FAM-LL-37 within the oppA mutants compared to wildtype. These results confirm that OppA is necessary for LL-37 cellular transport [unpublished].

Since the structure and function of this ABC transporter, and specifically the OppA subunit, is conserved among bacteria, we hypothesized that a structure for *Francisella*'s OppA protein can be computationally modeled using homologous OppA templates. Crystal structures are available for many OppA proteins including DppA from *Escherichia coli*, OppA from *Salmonella typhimurium*, OppA from *Yersinia pestis*, OppA from *Hameophilus influenzae*, and OppA from *Lactococcus lactis* [9]. OppA and peptide binding studies that have been performed in *L. lactis* and *H. influenzae* indicate a wide binding cavity that accounts for the promiscuous binding of small and large peptides seen from these proteins [9, 10]. This is an important distinction from many transport proteins that only bind small peptides. The computational modeling of *Francisella*'s OppA will determine structure and binding capabilities with LL-37 in *F. novicida* as well as in the virulent species *F. tularensis* SchuS4 which shares 98% homology of the OppABCDF

transporter genes. The understanding of the antimicrobial mechanism of action within *Francisella* will provide key insights into developing potential antimicrobial treatments.

Computational Modeling of Protein Structure

The structure of a protein holds a wealth of information that may infer its function, transport and storage abilities, role in chemical reactions, and other regulatory functions [11]. The traditional experimental methods used to determine protein structure, including crystallography, electron microscopy, or NMR, are expensive and time consuming, which has led to the development of computational options for predicting protein structures. To create a protein structure for *Francisella*'s OppA protein, computational modeling programs that predict 3D structure from homologous protein structures were used. This process, known as homology modeling, uses structural information such as residue-residue contact patterns, secondary structure, and solvent accessibility, to align the target sequence against the structural models available in the protein library [11]. During the alignment process, an optimized potential energy function (an objective or scored function) is created to assess the fitting quality of a sequence to a 3D configuration. Finally, the objective scores are statistically analyzed to assess the possibility of the target amino acid sequence adopting one of the available structural folds [11]. Once a predicted model for OppA is selected from the various computational programs and models, the structure(s) will be used in the computational docking of OppA protein and peptide LL-37 fragments.

Computational Modeling of Docking OppA and LL-37

With the experimental evidence indicating that the antimicrobial peptide LL-37 has intracellular targets, it is possible that LL-37 is entering the cell though the peptide transporter OppA. To explore this possibility, it is valuable to computationally model the binding of *Francisella*'s OppA protein with peptide LL-37. There have been several studies completed on the mechanism of action of substrate-binding proteins of the SBP_bacterial_5 superfamily, and on peptide binding capacity of *L. lactis*'s OppA protein [9, 12]. Crystal structures of OppA-peptide complexes from *Salmonella typhimurium, Yersinia pestis,* and *Escherichia coli* show the main interactions to be with the peptide backbones, independent of amino acid specificity, and peptides 4–35 amino acids long, as demonstrated in *L. lactis* [9, 13]. Additionally, studies on *S. typhimurium*'s OppA indicate that the voluminous hydrated cavity may provide flexible binding requirements through hydrogen bonding, or shielding peptide charges, which may account for larger peptide transport [14]. The information from these studies and the known structure and function from other OppA subunits is helpful in understanding AMP docking to *F. novicida* OppA. Based on the Comprehensive Evaluation of Fourteen Docking Programs on Protein-Peptide Complexes article published in 2020, different software programs perform more accurately based on global or local binding predictions, length of the peptide, the length of the binding site on the protein, and the amount of flexibility given for the initial binding conformations [15]. To best determine the binding action of LL-37, several of the top-ranked software programs will be used to model the docking of peptide KR-12, the smallest active fragment of LL-37, and larger fragments of

LL-37 based on program input restrictions. Analysis for the *F. novicida* OppA-LL-37 binding predictions will include analyzing the amino acids contacts, the binding affinities of the protein-peptide complex and how this predicted binding action compares to other known OppA protein-peptide bindings.

OppA Conformations

Previous crystallization of OppA subunits of ABC-transporter proteins and their docking with small peptides has demonstrated that the OppA proteins maintain two main conformations during the docking process. OppA proteins reside in an open conformation when the peptide is unbound and, after docking, the protein-peptide complex forms a closed conformation [14]. For small peptides, 3-6 amino acids in length, the closed conformation completely encloses the peptide within the large cavity. In peptides greater than 6 residues there is evidence that the docking mechanism may involve the peptide only partially binding within the cavity interior while the remaining residues bind to the exterior protein surface or free-float [14]. Although computational docking alone does not account for OppA conformation changes in real time, a comparison of open and closed models will be used to analyze the influence of the conformation state on docking.

Specific Aims

Aim 1: Computational Modeling of *Francisella* OppA protein onto 3D structures

Currently there are no available crystal structures for *Francisella novicida* OppA protein, however there are many crystallized models for OppA in other species of bacteria. Since this superfamily maintains high structural homology, the available crystal structures in PDB served as templates for OppA homology modeling. Computational

modeling of *F. novicida* OppA was completed using several structure prediction programs. The confidence scores from the models were used in analyzing the quality of the models. The OppA *F. novicida* U112 sequence (A0Q888, FTN_1593) was used for the modeling process. The best models were selected for use in AIM 2 and AIM 3.

Aim 2: AMP Binding Prediction using Computational Tools

In order to determine potential binding locations, a predictive binding program was used to acquire potential binding locations. With the many protein-ligand complexed homologous models available for OppA subunits, these predictive algorithms provided insightful binding information. SPOT-peptide was used to complete peptide binding prediction on the selected *Francisella* OppA models created in AIM 1.

Aim 3: Computational modeling of *F. novicida* OppA's docking of AMP LL-37 fragments

Computational modeling of *F. novicida*'s OppA with LL-37 fragments was completed using HPEPDOCK, MDockPeP, and GalaxyPepDock modeling programs. Due to its length of 37 amino acids, the LL-37 peptide is considered large for computational modeling. To simplify modeling, the smallest active fragment of LL-37, called peptide KR-12, was modeled first. KR-12 represents the helical central part of the peptide and has displayed selective toxic effects on bacteria [2]. Additionally, based on accepted input lengths, a larger fragment, LL-20, was docked in certain programs.

The docking inputs included the model structure(s) from AIM 1 and the peptide fragment KR-12 crystal structure (PDB: 2NA3) and additional fragment sequence from LL-37 crystal structure (PDB: 2K60). The resulting binding affinities, specific amino

acids and binding distances (in Angstroms) were recorded and analyzed for the most favorable docking complexes. Computational analysis and visualization of the docking was completed with Visual Molecular Dynamics (VMD) software [45].

METHODS

Computational Modeling of *Francisella* OppA protein onto 3D structures

The goal of AIM 1 was to create a model of protein OppA of the *F. novicida* U112 sequence (A0Q888, FTN_1593) using multiple web-based homology servers. Five different modeling programs were used in the creation of *Francisella*'s OppA protein models. The protocol for each program is described below.

RaptorX Structure Prediction

RaptorX Structure Prediction Server is a distance-based protein folding powered by deep learning [16-21]. The A0Q888 fasta file was uploaded to the server using default settings. Out of the 100 models produced from RaptorX Structure Prediction, the top 2 models were selected for further analysis.

RaptorX Old Structure Prediction (Homology Based)

RaptorX Old Structure Prediction Server is a template-based protein structure modeling server and the recommended program if the input sequences have close homologs in PDB [16-19]. The A0Q888 fasta file was uploaded to the server using default settings. Out of the 100 models produced by RaptorX Old Structure, which is homology-based, the top model was selected for further analysis.

Swiss Model Homology Prediction

Swiss Model Homology Prediction is a web-based server that does protein structure homology modeling using their ProMod3 pipeline [22-26]. The A0Q888 fasta file was uploaded to the server using default settings. Out of 50 uniquely templated models returned by Swiss-Model, the top 6 with the most negative Q-MEAN from different species were selected for further analysis.

Additionally, the user template mode of Swiss-Model was used to create a *Francisella* OppA model specifically templated off PDB:3TCH from *E. coli*. This template was selected due to its open conformation crystallized structure instead of the closed conformation maintained by all other models.

I-TASSER

I-TASSER (Iterative Threading ASSEmbly Refinement) identifies structural templates from the PDB by the multiple threading approach LOMETS and then threads the 3D models through the protein function database BioLiP [27-29]. The A0Q888 fasta file was uploaded to the server using default settings. The top ranked I-TASSER model was selected for further analysis.

Mod-Base

Mod-Base is a database of comparative protein structure models, calculated by the modeling pipeline ModPipe [30]. The A0Q888 fasta file was uploaded to the database and the server matched a pre-computed *Francisella* OppA model. The server noted that this model was from 08-13-2011 and is not updated based on new protein structure findings. This model was selected for further analysis.

Analysis of OppA Models

A single model or selection of *Francisella*'s OppA models was selected from each program based on the algorithm's specific criteria scores produced with model output data with a total of 11 models chosen for in-depth analysis (**Table 1**).

Pairwise alignment using the web-based server TM-Align, which produces an RMSD score as well as a TM-score, was performed on all 11 models [31]. A matrix of TM-Align RMSD scores for all models was created for analysis (**Table 2**).

Pairwise alignment using the web-based server FATCAT, which produces an RMSD score, was performed on all 11 models (**Table 3**).

The three models that demonstrated consistently low pairwise alignment scores within the matrix were selected for predictive binding and peptide docking of LL-37 fragments. The *Francisella* OppA models selected were RaptorX Old Structure Prediction Model 1 **(RapX_hlog_1)**, Swiss-Model Homology Prediction from template PDB:1B4Z **(SM_hlog_1b4z),** I-Tasser Model 1 **(itasser_mod1).**

Table 1: Francisella OppA Model Values
Computational values for the 11 *Francisella* OppA models selected for further analysis.

Francisella OppA Models						
Program	**Species**	**PDB**	**Seq Identity**	**GMQE**	**QMEAN**	**Molecule**
SM	Salmonella typhimurium	1rkm	34.98%	0.68	-0.95	Oligo peptide binding protein
SM	Salmonella typhimurium	1b4z	34.98%	0.69	-1.39	Oligo peptide binding protein OppA with KDK
SM	Yersina pestis	2z23	35.83%	0.69	-1.57	Oligo peptide binding protein OppA with tri-lysine ligand
SM	Haemophilus influenzae	6dqq	33.92%	0.69	-1.61	OppA complex with endogenous peptide
SM	Burkholderia pseudomallei	3zs6	31.75%	0.68	-2	Burkholderia pseudomallei OppA
SM	Escherichia coli	3tcg	34.53%	0.67	-2.11	Periplasmic oligopeptide-binding protein
Program	**Species**	**PDB**	**Seq Identity**	**RMSD**	**Tmscore**	**Molecule**
RapX_hlog_1	H. influenzae, Y. pestis	6dqq, 2z23				
RapX_pred1	Haemophilus influenzae	6dqq	33.92%	0.65	0.992	OppA complex with endogenous peptide
RapX_pred2	Yersina pestis	2z23	35.83%	0.62	0.993	Oligo peptide binding protein OppA with tri-lysine ligand
Program	**Species**	**PDB**	**Seq Identity**	**RMSD**		**Molecule**
ModBase	Thermotoga maritima	2o7l				Thermophilic cellobiose binding protein with cellobiose
Program	**Species**	**PDB**	**Seq Identity**	**RMSD**		**Molecule**
iTasser	H. influenzae, B. pseudomallei	6dqq, 3zs6				

Table 2: TM-Align Pairwise Alignment
Pairwise alignment RMSD values from all 11 *Francisella* OppA models using TM-Align

TM-Align Pairwise Alignment (RMSD)											
	SM_hlog_1rkm	SM_hlog_1b4Z	SM_hlog_2z23	SM_hlog_6dqq	SM_hlog_3zs6	SM_hlog_3tcg	RapX_hlog_1	RapX_pred_1	RapX_pred_2	MB_hlog_2o7l	itasser_mod1
SM_hlog_1rkm	0	3.06	3.05	3.29	3.33	3.14	3.09	2.29	2.11	4.77	3.29
SM_hlog_1b4Z	3.06	0	1.02	1.19	1.50	1.05	0.97	3.30	3.38	3.77	1.14
SM_hlog_2z23	3.05	1.02	0	1.22	1.39	1.24	0.97	3.28	3.36	3.85	1.23
SM_hlog_6dqq	3.29	1.19	1.22	0	1.63	1.33	1.04	3.45	3.48	3.90	0.74
SM_hlog_3zs6	3.33	1.50	1.39	1.63	0	1.67	1.74	3.40	3.46	3.89	1.61
SM_hlog_3tcg	3.14	1.05	1.24	1.33	1.67	0	0.97	3.27	3.28	3.79	1.19
RapX_hlog_1	3.09	0.97	0.97	1.04	1.74	0.97	0	3.54	3.66	3.88	1.07
RapX_pred_1	2.29	3.30	3.28	3.45	3.40	3.27	3.54	0	1.45	4.55	3.61
RapX_pred_2	2.11	3.38	3.36	3.48	3.46	3.28	3.66	1.45	0	4.59	3.66
MB_hlog_2o7l	4.77	3.77	3.85	3.90	3.89	3.79	3.88	4.55	4.59	0	3.87
itasser_mod1	3.29	1.14	1.23	0.74	1.61	1.19	1.07	3.61	3.66	3.87	0

Table 3: FATCAT Pairwise Alignment
Pairwise alignment RMSD values from all 11 *Francisella* OppA models using FATCAT

FATCAT Pairwise Alignment (RMSD)											
	SM_hlog_1rkm	SM_hlog_1b4Z	SM_hlog_2z23	SM_hlog_6dqq	SM_hlog_3zs6	SM_hlog_3tcg	RapX_hlog_1	RapX_pred_1	RapX_pred_2	MB_hlog_2o7l	itasser_mod1
SM_hlog_1rkm	0	3.05	3.03	3.04	3.17	3.00	3.07	2.10	2.11	3.08	3.04
SM_hlog_1b4Z	3.05	0	1.02	1.28	1.50	1.05	0.89	3.21	3.15	3.07	1.14
SM_hlog_2z23	3.03	1.02	0	1.22	1.43	1.15	0.84	2.73	2.73	3.58	1.15
SM_hlog_6dqq	3.04	1.28	1.22	0	1.63	1.21	0.89	3.25	3.10	3.02	0.74
SM_hlog_3zs6	3.17	1.50	1.43	1.63	0	1.56	1.53	3.26	3.26	3.02	1.61
SM_hlog_3tcg	3.00	1.05	1.15	1.21	1.56	0	0.97	3.09	3.17	3.10	1.20
RapX_hlog_1	3.07	0.89	0.84	0.89	1.53	0.97	0	3.23	3.28	3.08	1.07
RapX_pred_1	2.10	3.21	2.73	3.25	3.26	3.09	3.23	0	1.26	3.23	3.15
RapX_pred_2	2.11	3.15	2.73	3.10	3.26	3.17	3.28	1.26	0	3.13	3.24
MB_hlog_2o7l	3.08	3.07	3.58	3.02	3.02	3.10	3.08	3.23	3.13	0	3.14
itasser_mod1	3.04	1.14	1.15	0.74	1.61	1.20	1.07	3.15	3.24	3.14	0

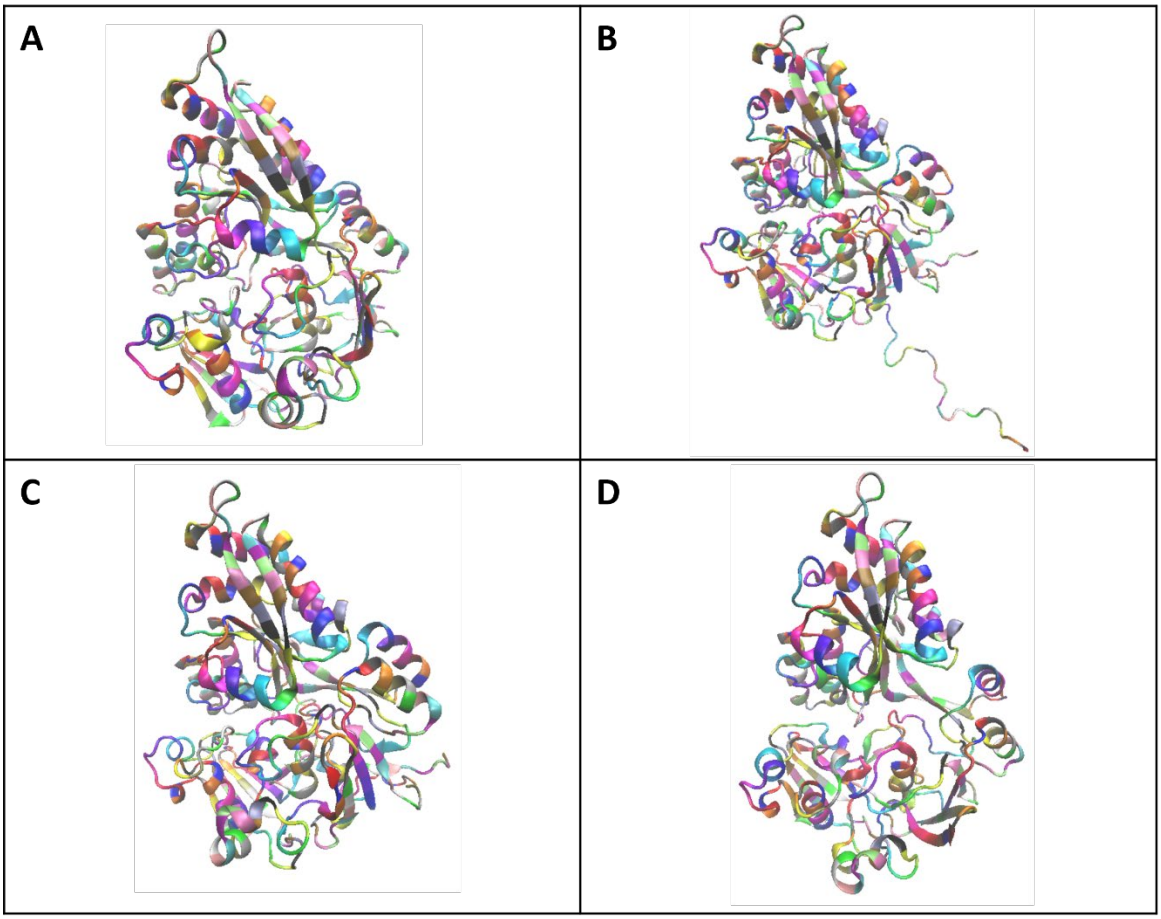


Figure 1: *Francisella* OppA models
***Francisella* OppA models listed left to right: (A) SM_hlog_1b4z created with the Swiss-Model program (B) RapX_hlog_1 created with RaptorX Old Structure Prediction (C) itasser_mod1 created with I-TASSER (D) SM_3tch_open, the open conformation of OppA created with Swiss-Model program.**

AMP Binding Prediction using Computational Tools

The three *Francisella* OppA models used for binding site prediction were **RapX_hlog_1**, **SM_hlog_1b4z,** and **itasser_mod1.** Predictive peptide binding was performed using SPOT-peptide server which predicts binding based on known peptide-binding templates. The model PDB file was uploaded to the server using default settings.

SPOT-peptide predictive binding was also performed on the *Francisella* OppA model **SM_hlog3tch_open** templated from open conformation crystallization of *E. coli'*s OppA (PDB:3TCH) as well as the closed conformation crystalized structure of the same OppA protein (PDB:3TCG), named **SM_hlog3tcg_closed**.

Computational modeling of *F. novicida* OppA's docking of AMP LL-37 fragments

The four *Francisella* OppA models used for peptide docking were **RapX_hlog_1, SM_hlog_1b4z, itasser_mod1** and **SM_hlog_3tch_open**. The AMP sequences used for binding were fragments KR-12 (KRIVQRIKDFLR), the smallest segment of LL-37 that still exhibits independent antimicrobial function, and fragment LL-20 (LLGDFFRKSKEKIGKEFKRI), the first 20 amino acids of the AMP LL-37. The protocol for each program is described below.

HPEPDOCK

HPEPDOCK is a web-based peptide docking program that performs blind protein-peptide docking [32-41]. The receptor input was the PDB file of the *Francisella* OppA model and peptide input was the sequence of the AMP in FASTA format. Default settings were used.

GalaxyPepDock

GalaxyPepDock is a web-based peptide docking program that performs docking based on similar interactions in the structure database and energy-based optimization [42]. The receptor input was the PDB file of the *Francisella* OppA model and peptide input was the FASTA file of the AMP.

MDockPep

MDockPep is a web-based peptide docking program that performs docking of peptides up to 20 AA in length [43, 44]. The receptor input was the PDB file of the *Francisella* OppA model and peptide input was the peptide sequence with default settings.

RESULTS

Computational Modeling of *Francisella* OppA protein onto 3D structures

The 11 homologous models of *Francisella* OppA protein highly resembled one another with the RMSD values from pairwise alignment of most models demonstrating <4.0 Å difference (**Table 1, 2**). From three programs, at least one model demonstrated 1.5Å or less difference to at least 6 other homologous models and was selected for peptide docking.

AMP Binding Prediction using Computational Tools

The SPOT-peptide predictive binding performed on **RapX_hlog_1**, **SM_hlog_1b4z**, and **itasser_mod1** all demonstrated predicted peptide binding in the cavity (**Figure 2**). SPOT-peptide predicted 73% to 82% identical amino acid specificity in the predicted binding sites for all three models (**Table 4**). This is consistent with known evidence that OppA transporter proteins bind peptides within a cavity region during transport [12].

The SPOT-peptide predictive binding performed on the open OppA model **SM_hlog3tch_open** and the closed OppA model **SM_hlog3tcg_closed** revealed very little specificity between the predictions of open and closed conformations of the same protein (**Table 5**). The program suggested two alternative binding predictions for the

open conformation but both predictions showed only 38% and 53% identical amino acid predictions to the closed conformation.

Table 4: Predicted OppA Binding Site Amino Acids in Closed Model Docking
List of predictive binding amino acids on *Francisella* OppA models from SPOT-peptide. Light green highlights identical amino acids in all three models.

Peptide Binding Prediction Tools																												
SPOT-peptide																												
SM_hlog_1b4Z	S59	D60	N61	S64	R65	Y141	Y189	Y273	N274	N275	M298	W435	L439			R451	M452	A453	W454	I455	A456	D457			Y476	Q523	M543	R545
RapX_hlog_1	S59	D60	N61	S64	R136	I188	Y189	Y273	N274	N275	M298	W435	Y438	L442		R451	M452	A453	W454	I455	A456	D457	M467	N474	Y476	Q523		R545
itasser_mod1	S59	D60	N61	S64			Y189	Y273	N274	N275	M298	W435	L439	L442	K443	R451	M452	A453	W454	I455	A456	D457				Q523	M543	R545

Table 5: Predicted OppA Binding Site Amino Acids in Open Model Docking
List of predictive binding amino acids on *Francisella* OppA models from SPOT-peptide. Light green highlights identical amino acids in all three models.

Peptide Binding Prediction Tools																											
SPOT-peptide																											
SM_hlog3tch_open						S140	Y141							W435	L439		A453	W454	I455	A456	D457	N462	T463	Y464	Q523		
SM_hlog3tch_open		D60	N61					I188	Y189	Y273	N274	N275	M298						I455	A456	D457				Q523	V542	
SM_hlog3tcg_closed	S59	D60	N61	S64	Y138				Y189	Y273			M298	W435		R451	A453	W454	I455		D457					V542	

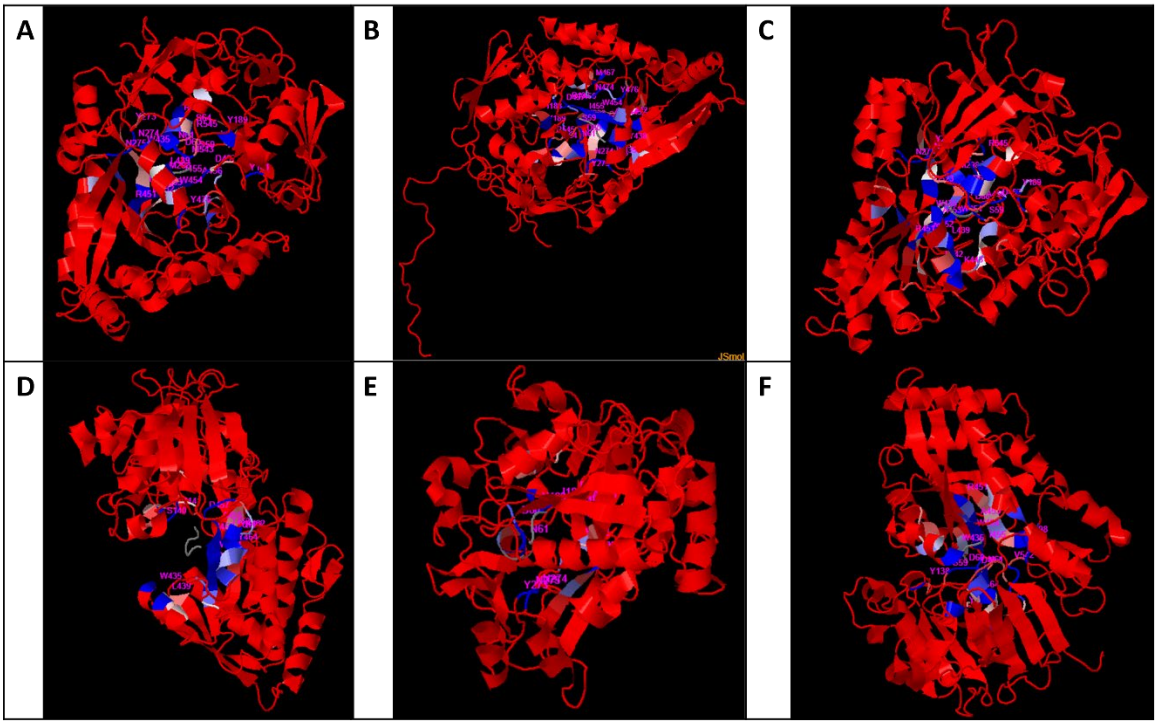


Figure 2: SPOT-peptide Predicted Images
***Francisella* OppA SPOT-peptide predicted peptide binding models listed left to right: (A) SM_hlog_1b4z (B) RapX_hlog_1 (C) itasser_mod1 (D) SM_hlog3tch_open (prediction 1) (E) SM_hlog3tch_open (prediction 2) (F) SM_hlog3tcg_closed. Blue color indicates the predicted binding sites.**

Computational modeling of *F. novicida* OppA's docking of AMP LL-37 fragments

The computational docking images and docking scores were collected for the top ten models of *Francisella* OppA protein models to peptides KR-12 and LL-20, if applicable. Additionally, amino acid contacts were collected for two protein-peptide complex models from each of the peptide-binding programs. From the available output data, the peptide docking scores were collected in a table for in-depth analysis (**Table 6**). While docking scores cannot be compared between programs due to divergent algorithms, patterns were analyzed for each program with the most favorable docking

scores, -200 and above, highlighted in green and the docking values -189 to -199 in yellow.

Table 6: Peptide Docking Scores
Peptide docking scores for KR-12 and LL-20 fragments provided in output data from docking programs. Docking values -189 to -199 in yellow and -200 and below in green.

Peptide Docking Scores											
Protein	Peptide	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
HPEPDOCK											
SM_hlog_1b4Z	KR-12	-195.4	-178.8	-164.2	-163.4	-163.3	-161.6	-159.9	-159.5	-158.9	-156.8
RapX_hlog_1	KR-12	-194.8	-182.7	-176.2	-174.1	-168.5	-166.5	-165.5	-165.3	-164.3	-163.8
itasser_model1	KR-12	-181.3	-168.3	-166.6	-165.7	-163.2	-163.0	-162.0	-160.0	-159.5	-158.1
SM_hlog_3tch_open	KR-12	-251.6	-222.5	-212.3	-195.4	-193.6	-191.2	-185.7	-182.0	-178.4	-178.1
SM_hlog_3tcg_closed	KR-12	-170.8	-169.1	-166.5	-159.2	-158.8	-155.1	-153.3	-152.5	-152.2	-151.9
SM_hlog_3tch_open	LL-20	-194.2	-180.9	-177.0	-172.0	-172.7	-166.0	-160.3	-159.8	-158.9	-158.8
MDockPEP											
SM_hlog_1b4Z	KR-12	-196.4	-181.6	-180.2	-180.1	-178.4	-178.3	-178.2	-177.1	-177.1	-176.6
RapX_hlog_1	KR-12	-185.8	-184.4	-184.1	-182.6	-182.1	-181.3	-180.1	-180.0	-179.7	-179.5
itasser_model1	KR-12	-196.2	-189.7	-189.0	-188.6	-188.4	-188.4	-185.9	-184.2	-182.8	-182.6
SM_hlog_3tch_open	KR-12	-220.3	-217.9	-216.5	-213.0	-208.6	-208.1	-203.1	-203.0	-202.3	-201.1
SM_hlog_3tcg_closed	KR-12	-191.3	-191.3	-186.6	-186.0	-185.1	-184.2	-181.6	-180.5	-180.3	-180.1
SM_hlog_3tchopen	LL-20	-290.6	-289.9	-289.2	-288.7	-287.1	-286.8	-286.6	-286.2	-285.6	-284.7

HPEPDOCK

While HPEPDOCK demonstrated global binding of the KR-12 peptide with all 3 closed conformation models, the bindings were on the exterior of the protein, not in the predicted (known) binding cavity. As seen in **Figure 3**, HPEPDOCK attempted to find favorable binding anywhere possible on the closed conformation protein surface since the cavity accessed remains closed. The modeled images illustrate the significant structural change that occurred from closed to open conformations that allows the peptide access to the binding cavity (**Figure 3**). The binding scores for the closed conformation models

were of limited favorability ranging from the -150's to -190's. Only with the open conformation model, **SM_hlog_3tch_open**, did docking occur within the predicted and known binding cavity and show highly favorable binding (-200 and below) with KR-12 for all ten output models (**Table 6**, **Figure 3**). The docking images for **SM_hlog_3tch_open** and KR-12, the center of the full length peptide, exemplifies how docking in the central zone of the cavity provides docking space for the full length LL-37 peptide (**Figure 3**).

The OppA computational docking with LL-20 performed in HPEPDOCK illustrated LL-20's docking within the cavity in many varied positions, often with three distinctive trends: (1) partially in the cavity with a free-floating tail (2) nose-in cavity binding with forward extending, free-floating tail and (3) partially in the cavity with exterior wrapping and binding to protein surface (**Figure 4**). The **SM_hlog_3tch_open** and LL-20 models had a wider range of binding favorability (-194.2 to -158.8) for all ten output models than KR-12 but overall supports successful computational docking of a larger peptide.

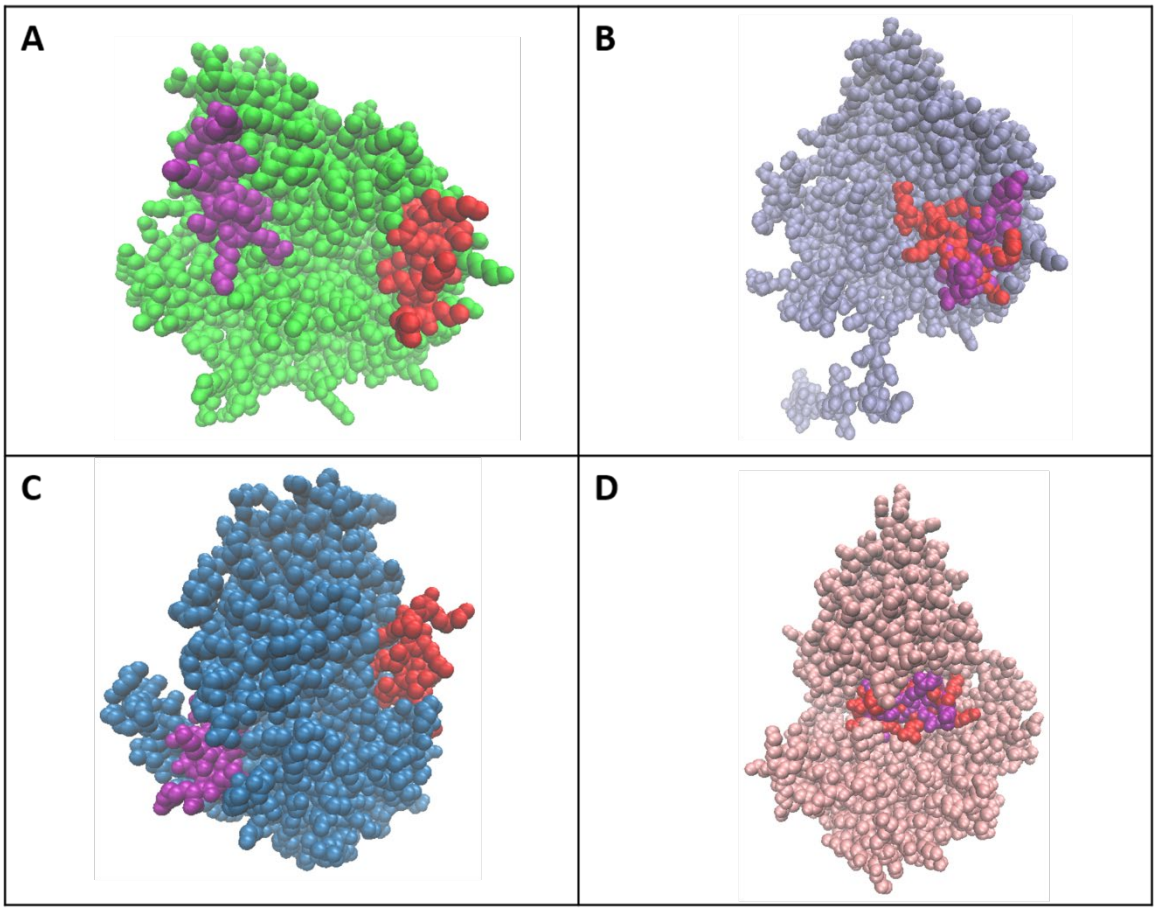


Figure 3: HPEPDOCK Docking of KR-12
HPEPDOCK docking of KR-12 peptide with *Francisella* OppA models (A) SM_hlog_1b4z (B) RapX_hlog_1 (C) itasser_mod1 (D) SM_hlog_3tch_open. For each OppA model, docking model 1 (purple) and model 2 (red) are illustrated on the same protein image.

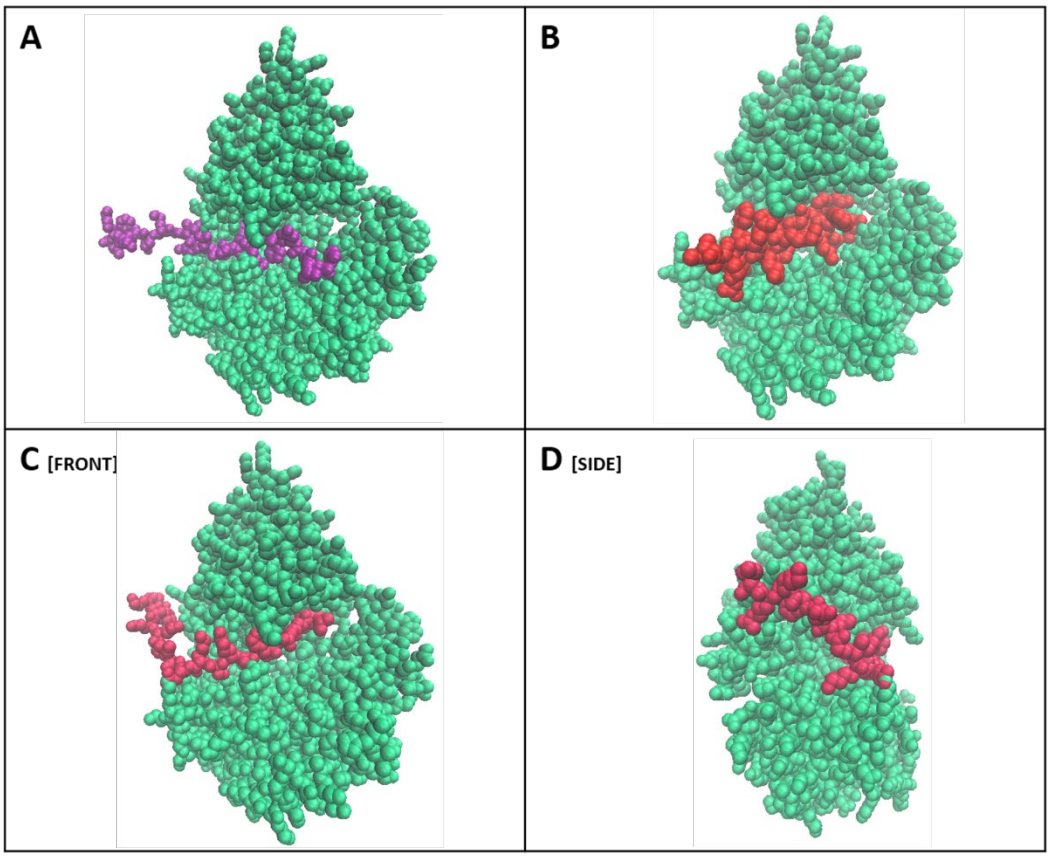


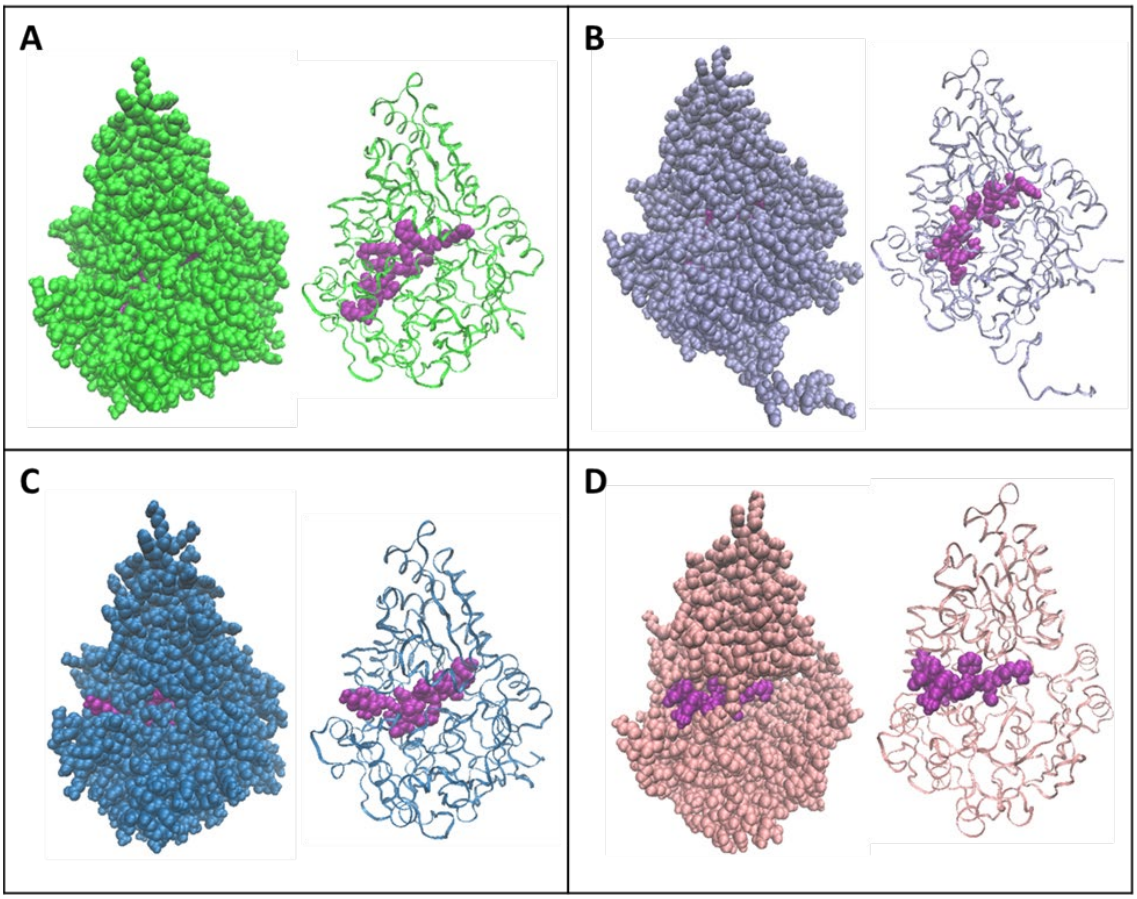
Figure 4: Docking of LL-20 and SM_hlog_3tch_open in HPEPDOCK
SM_hlog_3tch_open docking with LL-20 peptide (A) HPEPDOCK_model_1 [elongated and free floating] (B) HPEPDOCK_model_3 [sticking forward from protein] (C) HPEPDOCK_model_6 Front View (D) HPEPDOCK_model_6 Side View [exterior binding and wrapping]

GalaxyPepDock

In comparison to all other docking programs, GalaxyPepDock demonstrated docking within the cavity for all *Francisella* OppA models, regardless of the closed or open conformations (**Figure 5**). This result is believed to be a product of the program's algorithm, which is based on known templates of peptide-docking from a database, similar to SPOT-peptide prediction. The models in **Figure 5** illustrate the major structural transformation of the closed conformation on the protein's structure, with the KR-12

peptide being completely invisible due to its interior docked position, as well as the large cavity space within this transporter protein.

While no docking scores were produced in the output data, the contact amino acids of significant binding distance, < 6.5 Å, were collected for 2 models from the program for additional analysis.



GalaxyPepDock docking of KR-12 peptide with *Francisella* OppA models (A) SM_hlog_1b4z (B) RapX_hlog_1 (C) itasser_mod1 (D) SM_hlog_3tch_open. For each docking model 1 (purple), the same image is illustrated in ribbon.

MDockPep

Similar to HPEPDOCK results, MDockPep attempts to dock KR-12 to the closed conformation models demonstrated global binding over the exterior protein surface (**Figure 6**). The majority of docking scores demonstrated docking in the -189 to -199 range and were not as favorable as docking scores for the **SM_hlog_3tch_open** protein (**Table 6**). Only with the open conformation model **SM_hlog_3tch_open** did docking occur within the known binding cavity and with highly favorable binding scores of (-220.3 to -201.1) (**Table 6**). Additionally, the **SM_hlog_3tch_open** with KR-12 docking models illustrated the flexibility of the KR-12 docking position in the cavity, which may aid the docking of the larger full-length peptide LL-37 (**Figure 6**).

The docking of peptide LL-20 and **SM_hlog_3tch_open** performed in MDockPep demonstrated the highest favorable binding scores (-290.6 to -284.7) of all simulations (**Table 6**). Similar to HPEPDOCK dockings, these complexed models also demonstrated the variety of favorable binding positions of LL-20 in the cavity region including previous and new distinctive trends: (1) partially in the cavity with exterior wrapping and binding to protein surface and (2) nose-in cavity binding with the forward extension binding to exterior surface and (3) single-side docking within cavity allowing ample space for full protein length (**Figure 7**).

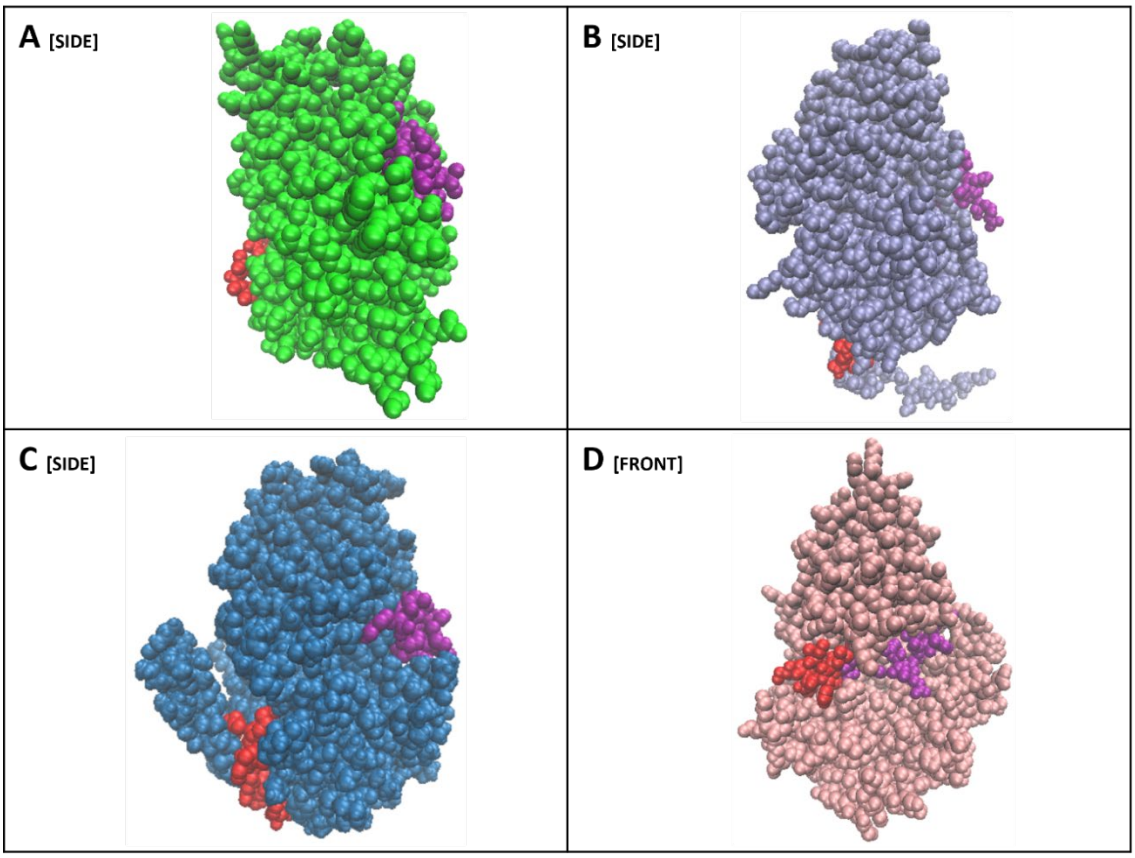


Figure 6: MDockPep Docking of KR-12
MDockPep docking of KR-12 peptide with Francisella OppA models (A) SM_hlog_1b4z (B) RapX_hlog_1 (C) itasser_mod1 (D) SM_hlog_3tch_open. For each OppA model, docking model 1 (purple) and model 2 (red) are illustrated on the same protein image.

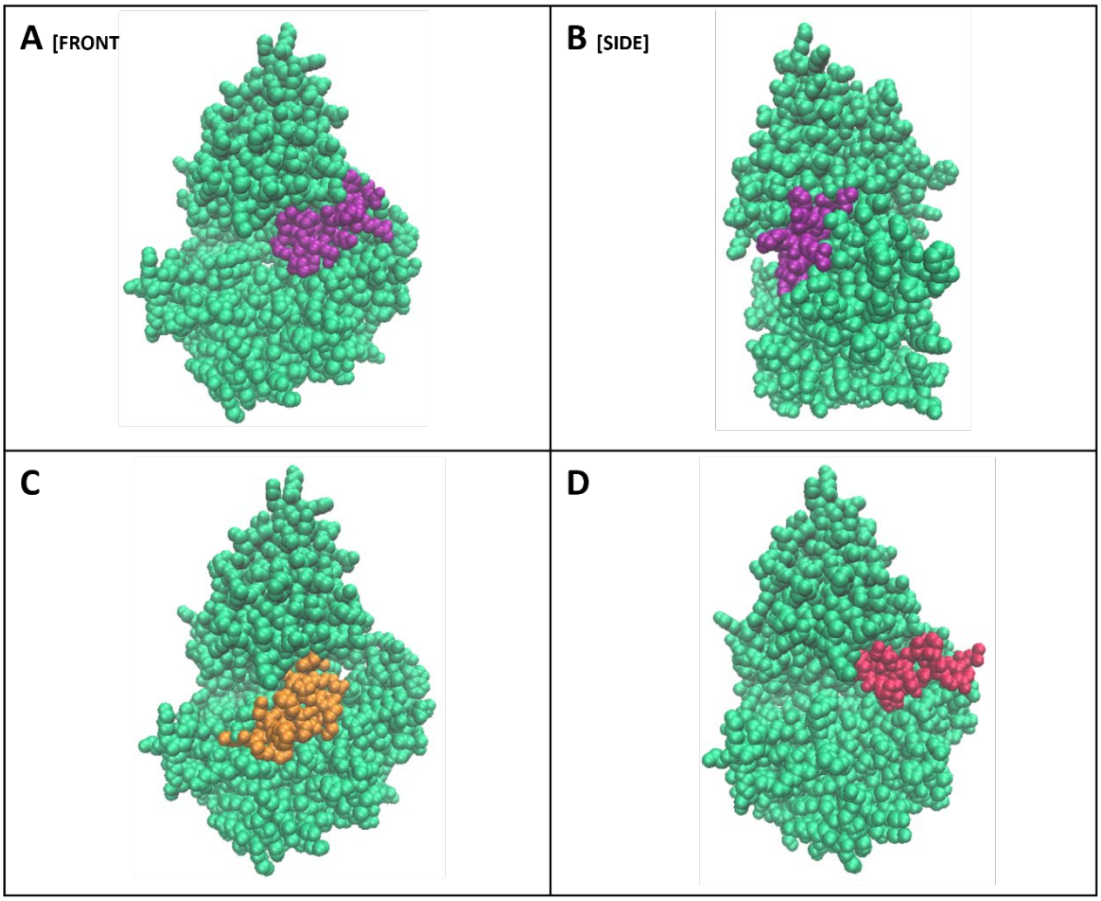


Figure 7: MDockPep Docking of SM_hlog_3tch_open and LL-20
MDockPep Docking of SM_hlog_3tch_open and LL-20 (A) MD_model_1 Front View (B) MD_model_1 Front View [peptide wrapping and binding] (C) MD_model_4 [forward extension and binding] (D) MD_model_5 [single side docking to maximize cavity space]

Cavity Specificity and Size

To confirm the understanding of cavity specificity, the amino acid contacts from significant binding distance, < 6.5 Å, were collected for 2 models from all docking programs and compared (data not shown). The results demonstrated very little amino acid specificity or significant binding patterns. This result supported the well-documented

promiscuous nature of the cavity as seen in other literature as well as the open conformation *Francisella* OppA docking results from this study [9].

Less prominently studied is large peptide docking (>20aa) within the OppA cavity. Although it is not computationally possible to model the full LL-37 AMP at once, the models of individual fragments were superimposed to help visualize the cavity size with a full-length LL-37 peptide. LL-20 and VQ-17, the front and back segments of LL-37 respectively, were modeled as fragments in MDockPEP (data not shown) and visualized in VMD (**Figure 8**). Even with some un-natural overlap occurring due to this methodology, it is clear to see that the OppA cavity is very spacious and could easily accommodate a large peptide of > 20+ amino acids, especially since peptide docking requires low specificity and there is available space to shift horizontally as necessary to fit (**Figure 8**). Additionally, **Figure 8 (D)** demonstrates how the LL-20 fragments ability to wrap along the exterior surface gives space inside the cavity for the second half fragment, VQ-17 to bind. This is positive evidence in support of the full-length LL-37's membrane-transport via the OppA protein of the OppABCDF complex.

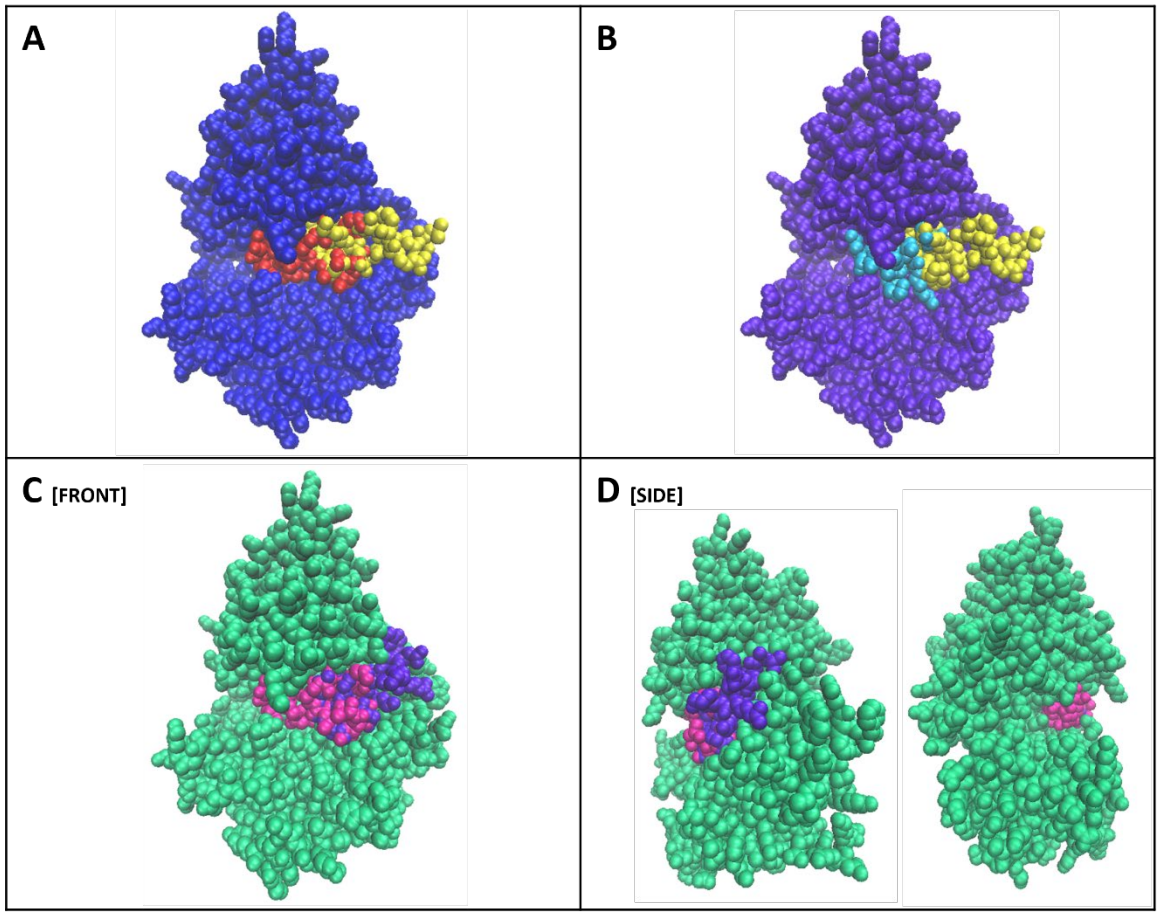


Figure 8: Comparison of MDockPep Docked Fragments
Multiple *Francisella* OppA models with peptides superimposed for comparison
(A) MD_model_5_SM_3tchopen_LL20 (yellow) + MD_mod1_3tchopen_VQ17 (red)
(B) MD_model_5_SM_3tchopen_LL20 (yellow) + MD_mod2_3tchopen_VQ17 (blue)
(C) MD_model_1_SM_3tchopen_LL20 (purple) + MD_mod1_3tchopen_VQ17 (pink) Front View
(D) MD_model_1_SM_3tchopen_LL20 (purple) + MD_mod1_3tchopen_VQ17 (pink) Side Views

DISCUSSION

Francisella OppA Computational Models

The computational creation of a *Francisella* OppA protein was very successful due to the large homologous structures available in PDB. Consistent models were obtained from various tools as demonstrated in the pairwise alignments supporting the current view that conserved structural homologs, when available, still provide better templates than non-homologous computational methods. The multiple web-based homology servers each provided a different strength, but the agreement of these models speaks to the highly conserved structure of this OppA transporter protein across many species.

Binding Prediction Takeaways

The SPOT-peptide binding predictions demonstrated that known peptide bindings occur within the large ABC-transporter cavity on the OppA protein. This essentially provided a local docking parameter for this study and visible non-cavity docking was ignored in open conformation models. This was extremely critical to the immediate recognition of the docking patterns in open vs closed conformation models.

Secondly, the SPOT-peptide binding prediction returned two different favorable docking models for **SM_hlog_3tch_open** (**Table 5, Figure 2**). This is indicative of the

many favorable biding possibilities and low specificity seen during *Francisella*'s OppA peptide docking, which was confirmed during this study.

The Effect of Open and Closed Conformations

The majority of homologous structures of the OppA transporter protein available in PDB were crystallizations of OppA proteins docked with ligands or small peptides. Therefore, the available templates for modeling an initial *Francisella* OppA structure were all closed conformations except for one open OppA *E. coli* template (PDB:3TCH). As seen in SPOT-peptide prediction, the computational program GalaxyPepDock only demonstrated "in-cavity" binding due to its use of template-based algorithms on the small ligand templates available. This was in complete contrast to the HPEPDOCK and MDockPep programs that use amino interactions or energy-optimization to predict binding. In both of these programs, all top 10 complexed models for each *Francisella* OppA input protein of a closed conformation demonstrated non-cavity binding of KR-12 and LL-20. The figures from this study show that *Francisella* OppA transporter proteins undergo a significant structural change when the large, spacious cavity is open vs closed. In-cavity peptide docking is not truly available in a closed conformation state. While only molecular dynamic simulations can account for real-time conformation changes in the docking process, it is important to address that successful computational modeling of OppA docking can be achieved with an open conformation model.

The use of open conformation *Francisella* OppA protein models allowed for distinctive trends of large peptide binding to be observed, including (1) partial binding in the cavity with exterior wrapping and binding to protein surface, (2) partial binding

within the cavity with free-floating ends, (3) nose-in cavity binding with the forward extension binding to exterior surface or free-floating, and (4) single-side docking to maximize cavity space to accommodate large proteins (**Figure 7**).

Favorable Docking with LL-37 fragments

The results from *Francisella* OppA model **SM_hlog_3tch_open**, showed favorable docking of the KR-12 peptide and LL-20 peptide both in docking scores and binding within the protein cavity. Despite the limited ability to computationally dock a peptide larger than 20aa, the complexed images were extremely supportive of a large protein cavity that could easily fit a peptide of double size for KR-12 and LL-20. This supports the experimental evidence that LL-37 is binding to OppA for transport across the membrane. Furthermore, the lack of specificity to amino acids and the multiple docking positions showing favorable binding, illustrates the many ways in which peptide binding can occur. This supports the experimental data that the highly conserved and widely used transporter protein is capable of transporting LL-37, regardless of the peptides' ability to fit within the whole cavity.

Future Directions

Computational protein modeling and peptide docking methods are excellent ways to understand specific aspects of a protein's structure and valid favorable bindings from experimental studies. Detailed understanding of *Francisella's* OppA and LL-37 binding could be further explored in additional wet-lab binding experiments, small fragment molecular dynamic simulations or binding parameters that address factors such as LL-37's +6 peptide charge.

More importantly, the validation of this protein's spacious cavity and flexible binding properties in open conformation can be applied to other large peptides, often unexplored due to the logistics of size. Furthermore, computational modeling and docking could be used to enhance databases for new or synthetic AMPs that are being increasingly researched due to the growing trends of antibiotic resistant bacteria. The computational knowledge from this study provides another part to the understanding of the role of OppA in LL-37's mechanism of action across the *Francisella novicida* cellular membrane which may lead to new therapeutic options against *Francisella* infections.

REFERENCES

1. Zhang, L.-j. and R.L. Gallo, *Antimicrobial peptides.* Current Biology, 2016. **26**(1): p. R14-R19.
2. Xhindoli, D., et al., *The human cathelicidin LL-37 — A pore-forming antibacterial peptide and host-cell modulator.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2016. **1858**(3): p. 546-566.
3. Chung, M.-C., Scott N. Dean, and Monique L. van Hoek, *Acyl carrier protein is a bacterial cytoplasmic target of cationic antimicrobial peptide LL-37.* Biochemical Journal, 2015. **470**(2): p. 243-253.
4. Ellis, J., et al., *Tularemia.* Clinical microbiology reviews, 2002. **15**: p. 631-46.
5. van Hoek, M.L., *Biofilms: an advancement in our understanding of Francisella species.* Virulence, 2013. **4**(8): p. 833-846.
6. Slamti, L. and D. Lereclus, *The oligopeptide ABC-importers are essential communication channels in Gram-positive bacteria.* Res Microbiol, 2019. **170**(8): p. 338-344.
7. Tanabe, M., et al., *The ABC Transporter Protein OppA Provides Protection against Experimental Yersinia pestis Infection.* Infection and Immunity, 2006. **74**(6): p. 3687-3691.
8. Norcross, S., A. Sunderraj, and M. Tantama, *pH- and Temperature-Dependent Peptide Binding to the Lactococcus lactis Oligopeptide-Binding Protein A Measured with a Fluorescence Anisotropy Assay.* ACS Omega, 2019. **4**(2): p. 2812-2822.
9. Berntsson, R.P., et al., *The structural basis for peptide selection by the transport receptor OppA.* Embo j, 2009. **28**(9): p. 1332-40.
10. Tanaka, K.J. and H.W. Pinkett, *Oligopeptide-binding protein from nontypeable Haemophilus influenzae has ligand-specific sites to accommodate peptides and heme in the binding pocket.* J Biol Chem, 2019. **294**(3): p. 1070-1082.
11. Dorn, M., et al., *Three-dimensional protein structure prediction: Methods and computational strategies.* Computational Biology and Chemistry, 2014. **53**: p. 251-276.
12. de Boer, M., et al., *Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers.* Elife, 2019. **8**.
13. Tame, J.R., et al., *The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands.* Structure, 1995. **3**(12): p. 1395-406.

14. Detmers, F.J.M., F.C. Lanfermeijer, and B. Poolman, *Peptides and ATP binding cassette peptide transporters.* Research in Microbiology, 2001. **152**(3): p. 245-258.
15. Weng, G., et al., *Comprehensive Evaluation of Fourteen Docking Programs on Protein-Peptide Complexes.* J Chem Theory Comput, 2020. **16**(6): p. 3959-3969.
16. Källberg, M., et al., *Template-based protein structure modeling using the RaptorX web server.* Nature Protocols, 2012. **7**(8): p. 1511-1522.
17. Ma, J., et al., *Protein threading using context-specific alignment potential.* Bioinformatics, 2013. **29**(13): p. i257-i265.
18. Peng, J. and J. Xu, *A multiple-template approach to protein threading.* Proteins, 2011. **79**(6): p. 1930-9.
19. Peng, J. and J. Xu, *Raptorx: Exploiting structure information for protein alignment by statistical inference.* Proteins: Structure, Function, and Bioinformatics, 2011. **79**(S10): p. 161-171.
20. Wang, S., et al., *Folding Membrane Proteins by Deep Transfer Learning.* Cell Systems, 2017. **5**(3): p. 202-211.e3.
21. Wang, S., et al., *Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model.* PLOS Computational Biology, 2017. **13**(1): p. e1005324.
22. Benkert, P., M. Biasini, and T. Schwede, *Toward the estimation of the absolute quality of individual protein structure models.* Bioinformatics, 2011. **27**(3): p. 343-50.
23. Bienert, S., et al., *The SWISS-MODEL Repository—new features and functionality.* Nucleic Acids Research, 2017. **45**(D1): p. D313-D319.
24. Studer, G., et al., *QMEANDisCo-distance constraints applied on model quality estimation.* Bioinformatics, 2020. **36**(6): p. 1765-1771.
25. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Research, 2018. **46**(W1): p. W296-W303.
26. Studer, G., et al., *ProMod3—A versatile homology modelling toolbox.* PLOS Computational Biology, 2021. **17**(1): p. e1008667.
27. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nature Protocols, 2010. **5**(4): p. 725-738.
28. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction.* Nature Methods, 2015. **12**(1): p. 7-8.
29. Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and function predictions.* Nucleic Acids Research, 2015. **43**(W1): p. W174-W181.
30. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models and associated resources.* Nucleic Acids Research, 2014. **42**(D1): p. D336-D346.
31. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic acids research, 2005. **33**: p. 2302-9.
32. Berman, H.M., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

33. Huang, S.-Y. and X. Zou, *Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking.* Proteins: Structure, Function, and Bioinformatics, 2006. **66**(2): p. 399-421.
34. Huang, S.-Y. and X. Zou, *An iterative knowledge-based scoring function for protein-protein recognition.* Proteins: Structure, Function, and Bioinformatics, 2008. **72**(2): p. 557-579.
35. Martí-Renom, M.A., et al., *Comparative Protein Structure Modeling of Genes and Genomes.* Annual Review of Biophysics and Biomolecular Structure, 2000. **29**(1): p. 291-325.
36. Pearson, W.R. and D.J. Lipman, *Improved tools for biological sequence comparison.* Proceedings of the National Academy of Sciences, 1988. **85**(8): p. 2444-2448.
37. Remmert, M., et al., *HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.* Nature Methods, 2012. **9**(2): p. 173-175.
38. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular Systems Biology, 2011. **7**(1): p. 539.
39. Yan, Y., D. Zhang, and S.-Y. Huang, *Efficient conformational ensemble generation of protein-bound peptides.* Journal of Cheminformatics, 2017. **9**(1).
40. Zhou, P., et al., *HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm.* Nucleic Acids Research, 2018. **46**(W1): p. W443-W450.
41. Zhou, P., et al., *Hierarchical Flexible Peptide Docking by Conformer Generation and Ensemble Docking of Peptides.* Journal of Chemical Information and Modeling, 2018. **58**(6): p. 1292-1302.
42. Lee, H., et al., *GalaxyPepDock: a protein–peptide docking tool based on interaction similarity and energy optimization.* Nucleic Acids Research, 2015. **43**(W1): p. W431-W435.
43. Xu, X., C. Yan, and X. Zou, *MDockPeP: An ab-initio protein–peptide docking server.* Journal of Computational Chemistry, 2018. **39**(28): p. 2409-2413.
44. Yan, C., X. Xu, and X. Zou, *Fully Blind Docking at the Atomic Level for Protein-Peptide Complex Structure Prediction.* Structure, 2016. **24**(10): p. 1842-1853.
45. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.

BIOGRAPHY

Sarah Parron graduated from Herndon High School, Herndon, Virginia, in 2005. She received her Bachelor of Science in Microbiology from Virginia Polytechnic and State University in 2009. She received her Master of Arts in Curriculum and Instruction from Virginia Polytechnic and State University in 2011. She was employed as a teacher in several years in Virginia before returning to school to complete her Master of Science in Bioinformatics from George Mason University.