SANTA CLARA UNIVERSITY

Department of Bioengineering

Department of Computer Science and Engineering

I HEREBY RECOMMEND THAT THE THESIS PREPARED

UNDER MY SUPERVISION BY

Maria Esquivel, Johann Fernando, Anna Fisher, Cameron Leong, & Adam Weaver

ENTITLED

BIOAI FOR ANTI-INFECTIVE DRUG DISCOVERY

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

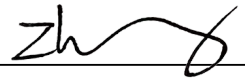
FOR THE DEGREE OF

BACHELOR OF SCIENCE

IN

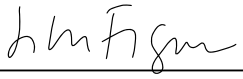
BIOENGINEERING

COMPUTER SCIENCE AND ENGINEERING



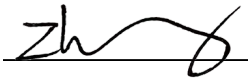
06/10/2022

Thesis Advisor — date



6/9/2022

Thesis Advisor — date



06/10/2022

Department Chair — date

N. Ling
N. Ling (Jun 13, 2022 14:23 PDT)

Department Chair — date

BIOAI FOR ANTI-INFECTIVE DRUG DISCOVERY

By

Maria Esquivel, Johann Fernando, Anna Fisher, Cameron Leong, & Adam Weaver

SENIOR DESIGN PROJECT REPORT

Submitted to
the Department of Bioengineering
the Department of Computer Science and Engineering

of

SANTA CLARA UNIVERSITY

in Partial Fulfillment of the Requirements
for the degree of
Bachelor of Science in Bioengineering
Bachelor of Science in Computer Science and Engineering

Santa Clara, California

Spring 2022

Abstract

Modern antibacterial drugs are quickly becoming insufficient for medical, agricultural and veterinary use, and the drug design techniques used to supplement the supply are not able to adequately and quickly produce enough effective candidates. We took a multi-pronged approach to remedying this problem, including shifting the target mechanism from antibiotic drugs in favor of anti-infective drugs, developing an AI which aimed to predict the interactions of small molecule fragments with the target protein, and refining wet lab testing protocols in order to make the drug design process holistically more efficient and effective. We have shown that there is potential in this design space, as the design techniques are faster and more effective than those previously used, and the drug candidates are promising as they undergo further testing. Research into broader applications of this technology to other target proteins, and expansion of the AI's scope in specific applications will allow for fast and safe design of life-saving drugs.

Acknowledgements

We would like to acknowledge our advisor Dr. Zhiwen Zhang and thank him for his leadership and guidance throughout the project. We would also like to acknowledge Dr. Silvia Figueria for her support and advice on the Computer Engineering aspects of the project. We would like to thank Dr. Emily Park for her guidance in designing our qPCR assay, as well as Lisa Jin and Melina Huang Xia for their training and support with the invasion assay. We would also like to acknowledge Daryn Baker, Victoria Walton, and Wendy Raposa for their support in the lab, as well as Vivian Zhang and Cathy Chen for their help with mammalian cell culture. Finally, we would like to acknowledge the School of Engineering at SCU for their financial support of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Artificial Intelligence	AI
Bioengineering	BIOE
Carboxyfluorescein succinimidyl ester	CFSE
Chinese Hamster Ovary	CHO
Computer Science and Engineering	COEN
Dulbecco's Modified Eagle Medium	DMEM
Dimethyl Sulfoxide	DMSO
Ethylenediaminetetraacetic acid	EDTA
Fetal Bovine Serum	FBS
Fluorescence-Activated Cell Sorter	FACS
Graphical User Interface	GUI
Methicillin-Resistant *Staphylococcus Aureus*	MRSA
National Center for Biotechnology Information	NCBI
No template control	NTC
Paraformaldehyde	PFA
Phosphate-Buffered Saline	PBS
Quantitative Real-Time Polymerase Chain Reaction	qPCR
Quantification cycle	C_q
Sodium Lauryl Sulfate	SDS
Tris Base, Acetic Acid, and EDTA	TAE
Tryptic Soy Broth	TSB

Chapter 1: Introduction

1.1 Background and Motivation

1.1.1 Antibiotic Resistance

Antibiotics are known as one of the most effective and successful drug therapies in all of human history. They are a type of antimicrobial chemical that is meant to destroy and/or inhibit the growth of bacteria to prevent and treat harmful infections. Their discovery has resulted in the development of countless subsequent medical treatments, countless lives saved, and even some diseases eradicated, such as smallpox. Paul Ehrlich and Alexander Fleming are credited with the development of the first natural and synthetic antibacterial drugs, Salvarsan and Penicillin. Ehrlich's "magic bullet" theory is what guided his search for a drug that could target disease-causing organisms without harming the rest of the body. His theory was based on synthetic laboratory dyes that were used to only stain microbes and not other tissues (Aminov et al, 2010). These revolutionary drugs have since been used as a baseline for furthering research on and development of bacterial infection treatment, our project will take a new approach.

While the development of antibiotics has saved millions of lives and improved overall quality of life on Earth, there is a growing threat in the form of antibiotic-resistant bacteria. Antibiotic-resistant infections are hard to treat and hard to prevent because antibiotic drugs introduce a pressure that causes bacterial cells within a population to mutate their targeted sites in order to survive the attack. This can happen with any kind of antibiotic and the rate of mutation has begun to outpace the rate at which antibiotics are being tested and created, as the process is both time and labor intensive.

One especially prominent type of antibiotic-resistant infection is infection by methicillin-resistant strains of *Staphylococcus aureus* (MRSA) (Taylor, 2022). *Staphylococcus aureus* is a gram positive bacteria that is commonly found on the skin and mucous membranes of humans. When it enters deep tissue or the bloodstream, it can cause a variety of infections such as gastroenteritis, meningitis, skin infections, and toxic shock syndrome which are typically treated with first-line antibiotics: methicillin, amoxicillin and penicillin. When left untreated, MRSA can cause severe complications and sepsis with a mortality rate between 20%-50% (CDC, 2019). In 2017, there were nearly 120,000 MRSA bloodstream infections in the United states with 20,000 associated deaths (CDC, 2019). While the number of overall bacterial infections is decreasing

due to improved healthcare access and other factors, the number of antibiotic-resistant strains is growing rapidly.

Attempts have been made to develop an alternative treatment for *S. aureus*, with little progress so far. Some challenges for developing a stronger treatment against *S. aureus* include poor understanding of pathogenic mechanisms, its multiple virulence factors, and the wide variety of diseases it can cause (Proctor et al, 2012). Thus, there is an urgent need for a new approach to treating *S. aureus* infections.

1.1.2 Sortase A

Gram-positive bacteria like *S.aureus* have a multitude of proteins displayed on the cell surface that are very important to successful cell function and pathogenicity. Sortase A is one of these proteins that anchors virulence factors to the outer membrane of *S. aureus*. It is a 206 amino acid, membrane-bound cysteine transpeptidase (Lu et al, 2007). Virulence factors are molecules responsible for the infection of host tissues and for retaining bacteria cells within. When Sortase A is in a dimer conformation, it is inactive. Therefore, the bacteria is unable to express virulence factors and can't invade the mammalian cell. When an activating molecule, released by mammalian cells, binds to sortase A, it undergoes a conformational change from a dimer into a monomer (Zhu et al, 2016). In this active monomer conformation, the bacteria can now express virulence factors and invade mammalian cells.

Typically, surface virulence proteins contain an N-terminal signal peptide and a C-terminal sorting signal, and Sortase A is capable of catalyzing a transpeptidation reaction, which is a reaction that links two protein chains together (Zhu et al, 2016). The LPXTG motif is part of the sorting signal that will be first cleaved by the Sortase A. The resulting surface protein intermediate is then covalently attached to the cell wall, a characteristic structure of Gram-positive bacteria, displaying the protein on the microbial surface.

By inhibiting the dimerization of sortase A with a small molecule drug, it may be possible to remove the anchored virulence factors from the surface of the cell and prevent its entry into mammalian host cells.

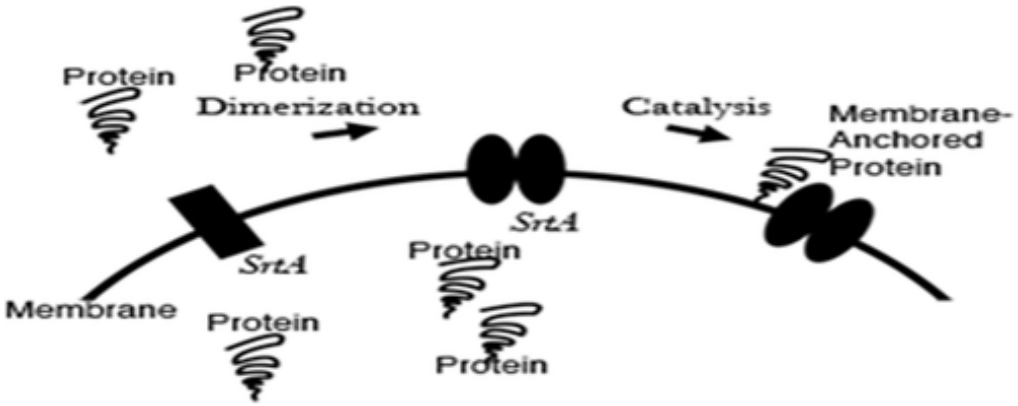


Figure 1. The sortase A protein on the surface of *S. aureus* exists as a dimer in its active form, which allows for the anchoring of proteins to the outer surface of the bacterial cell wall (Lu et al, 2007).

1.1.3 Current Drug Design Approaches

Current drug discovery methods are known to be time-intensive, labor-intensive, and expensive. They are not capable of keeping up with the rapidly growing number of antibiotic-resistant bacteria strains. There are two major schools of drug design, physical mass chemical testing and simulation based testing. The former has historically been more common, though its presence is fading to more modern approaches (Mahdal et al, 2009). Modern testing takes years to establish candidates, before even beginning clinical testing. There are several firms that have established deep learning approaches for drug design have been successful in some aspects but still have significant room to improve (Urbina et al, 2022). Both in terms of efficacy and safety, the industry must progress in order to effectively combat the problem of antibiotic drug resistance. Researchers from the Massachusetts Institute of Technology have shown notable results using AI-driven drug screening for antibiotics rather than anti-infectives. They trained a deep neural network model to analyze *E.coli* inhibition by introducing an established FDA-approved drug library supplemented with another natural compound library (Stokes et al, 2020). Next, the researchers had the model sort through known antibiotic compounds from the Drug Repurposing Hub to find a suitable broad-spectrum antibiotic for *E.coli*. This network model performed well with a 55% accuracy given its limited parameters (Stokes et al, 2020).

1.1.4 Flow cytometry for cell analysis

The invasion assay developed by Dr. Zhang's lab performs cellular fluorescence quantification using flow cytometry in order to establish a baseline quantification of *S. aureus* infection of CHO cells before testing our target molecule. Flow cytometry measures the intensity of fluorescently-stained cells and provides an absolute value for said measured light intensity. This technology is very popular in molecular research and has numerous applications in immunology, virology, molecular biology, cancer biology, and infectious disease monitoring. Stained and preserved cell solutions are suspended in a buffered saline solution that runs through the fluidics system within the cytometer where lasers measure the light intensity coming from each individual particle (McKinnon et al, 2018).

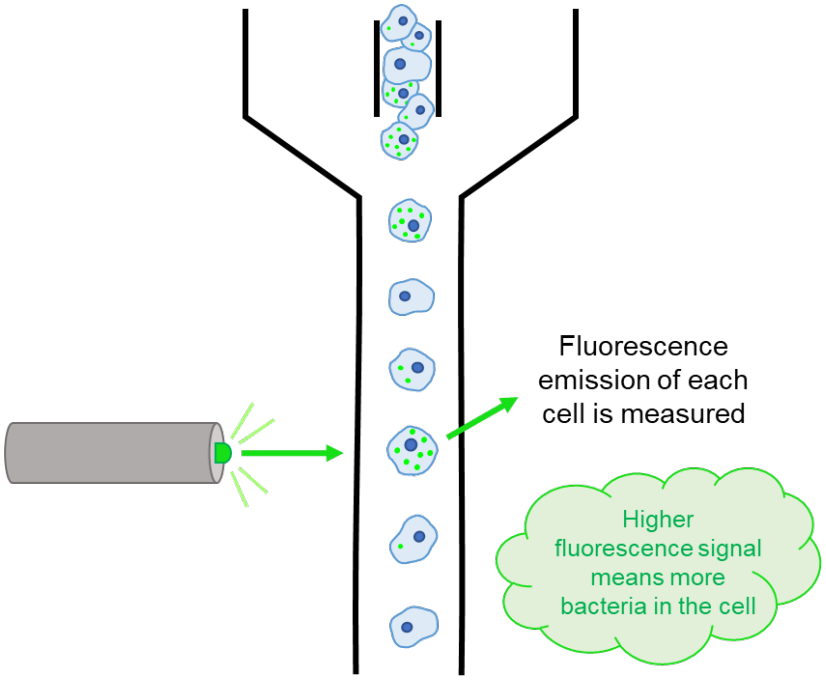


Figure 2. Simplified diagram of flow cytometry analysis used for quantification of bacterial invasion.

1.1.5 Real-Time PCR (qPCR)

Real-Time PCR (qPCR) is a nucleic acid amplification and detection technology that is currently one of the most valuable tools in modern biological research as it is used in a wide range of applications including diagnostics, research, biotechnology and medicine. This method quantifies the presence and amount of a gene of interest within a sample well in real time.

Quantification is achieved through the inclusion of fluorescence reporter molecules and the gene of interest itself is identified and amplified through the use of specific oligonucleotide primers.

For successful measurement, each sample well must have a certain amount of target DNA, primers, fluorescent dye, dNTPs and taq DNA polymerase and each of these ingredients play an important role. The target DNA sequence can be added in any form. Many biological research applications will extract and purify this sample before PCR. Our project intended to recognize a target sequence in *S.aureus* within infected chinese hamster ovary (CHO) cells without purification.

Primers are short fragments of single-stranded DNA that are complementary to sequences that flank the gene of interest. Taq DNA polymerase is an enzyme that attaches free floating nucleotides (dNTPs) to growing DNA strands, the fluorescent dye is often an intercalating dye that attaches to the newly formed sequence, and the qPCR machine will detect and quantify the levels of fluorescence. A higher fluorescence signal means there are more copies of target DNA successfully forming within the sample (Bio-Rad, "What is Real-Time PCR?").

A qPCR machine amplifies the labeled target sequences by cycling between specific temperatures over specific periods of time. The first step is denaturing where the machine will heat the sample up to about 95°C which is when double-stranded DNA will separate into single strands. The second step is primer annealing and extension. The temperature is brought down to 50-60°C where the primers will attach to both ends of the single-strand target sequence, and taq DNA polymerase will add complementary dNTPs (free floating nucleotides) (Bio-Rad, "What is Real-Time PCR?").

The average qPCR run will have between 30-40 amplification cycles, each time the amount of target DNA is amplified by 2^n copies. Fluorescence levels are measured during each cycle but are only detected after a certain amount of DNA is reached (Cq value) and the fluorescence is no longer background noise, after which the amplification happens exponentially. The Cq value is inversely related to the initial amount of target DNA within a sample and correlates to a number of target DNA copies in the sample. Low Cq values such as those around 29 cycles indicate larger amounts of the target sequence, while high Cq values indicate small amounts (Ruiz-Villalba, 2021).

1.2 Literature Review

1.2.1 AI for drug design

Motivated by the slow and expensive process of traditional drug design, researchers are eagerly pursuing AI and machine learning solutions. Most of the research is focused on the De Novo drug design, with two primary goals: distribution-learning and goal-directed generation (Blaschke, 2020). Distribution-learning is focused on creating molecules similar to given target molecules, and goal-directed learning seeks to design molecules that satisfy the users' specific objectives. Many ML methods are used for drug design, to varying degrees of success. A promising application is using recurrent neural networks to generate large amounts of molecules and then using transfer learning and reinforcement learning as a way to filter results (Arús-Pous, 2020). Conditional Graph generation has also shown success, especially in its ability to consistently generate valid outputs (Li, 2018). Because AI drug design is still in its infancy, we will surely see many more methods tried in the coming years.

1.2.2 Drug testing

High-throughput screening (HTS) is one of the most popular and well-established methods of drug design and testing. This method focuses on assaying and screening established chemical, genomic, protein and peptide libraries with biological models at a large scale which involves the use of robotics, liquid handling and advanced processing software. Cell-based assays are the main way of testing these libraries and high-density arrays of microreaction wells such as 96- or 384-well microtiter plates are typically how they are structured. Each microcreaction well will contain a 2D cell monolayer culture and the addition of one chemical from a library to be tested; the automated system in the form of robotic liquid handling technology will then grow the culture and monitor the reactions within each well. "Hits" are known as compounds that have produced the desired reaction within the microreaction well and those are taken through more specific and quantitative testing. Using this method, it is possible to screen up to 10,000 compounds a day (Szymanski et al, 2011) with the right amount of space and financial support.

Small molecules refer to chemical compounds with "small" molecular weights, ranging under 500 g/mol and 100 g/mol. These compounds and the vast libraries they make up are routinely used in HTS due to their ability to easily cross the cell membrane (Yip et al, 2017). Small molecule drug screening is a more specific type of HTS and can be split into two

categories: phenotype-oriented and target-based. Compounds in phenotype-oriented screens are typically tested for their ability to induce a certain phenotype in biological models. Things such as cell division alteration, metabolism, adhesion and viability are monitored using this type of screening method. Target-based screening involves more direct targeting, for example a purified compound would be tested for its ability to inhibit the function of certain proteins. Antibiotics are tested mainly through target-based screens for toxicity and for every "hit" there will be several counter and cross screens to validate the effectiveness of the compound (Yip et al, 2017).

1.3 Critiques of current approaches

1.3.1 AI for drug design

Most current applications of AI in this field are based around the De Novo drug design. The AI is given free reign to design the drug from scratch. The danger in this is that the drug being designed could potentially be toxic, or even fatal (Urbina, 2022). We argue that the best use of AI would be using it to build off of a scaffold that has been developed beforehand through rigorous wet lab testing, thus reducing the risk of severe side effects.

Furthermore it is critical that AI processes are carefully assembled. Current research is largely into narrow applications of AI which must later be integrated into an effective unit. There is immense danger here that ethical safety precautions which existed through human intervention are simply unthinkingly removed as they are combined. The current approaches for a synthesis of AI techniques must be interrogated to ensure that as they are brought together, the contexts which make them acceptably safe in isolation are not broken.

1.3.2 Drug testing

High-throughput small molecule screening is one of the most commonly used methods for developing drugs, but it can be expensive and time consuming especially considering the growing threat of antibiotic resistance. This threat is not new, but is quickly becoming a crisis as the inappropriate use of antibiotics in medical, veterinary and agricultural sectors within the last few decades has pushed microbe populations to adapt and develop drug resistance at astonishing rates(O'Neill, 2014). This is costly, current epidemiologic studies have shown that the economic burden associated with antibiotic resistance is significant, each year in Europe and the United

States such infections lead to 11 million additional hospitalization stays and over 20 billion dollars in additional social and healthcare costs (David et al, 2021).

The rate at which new antibiotic drugs are being developed is not enough to keep up with the ever-changing microbe populations that they target, researchers typically put in more effort than it is worth and financial support for such experiments is decreasing in favor of alternative drug testing methods. While HTS is known for its large scale testing methods and easy programmability, it has several drawbacks such as cost, prolonged testing/implementation and period of synthesis, and requires large amounts of space and human resources for maintenance (David et al, 2021).

1.3.3 Zhang lab invasion assay

Though measurement of fluorescence intensity is a useful and accurate method of quantifying the number of bacteria that were able to enter the CHO cells during the invasion assay, it has some limitations. For example, adding fluorescent dye to bacteria is a relatively labor-intensive step that is not guaranteed to label all bacterial cells properly. It is also difficult to ensure that all bacteria have been labeled permanently. The instrumentation required to quantify fluorescence intensity is also expensive and requires a significant technical background to operate, making the invasion assay more costly and difficult to run (Jian et al, 2020).

1.4 Project Overview

Anti-infective drug approaches may be the answer to this call for a new bacterial infection treatment approach. By preventing *S. aureus* from interacting with and invading mammalian cells in the first place, they would no longer be able to proliferate using the host cell's machinery- thus, preventing severe disease (Alksne, 2000). Previous research has shown that the protein called sortase A is integral to the entry of *S. aureus* into mammalian host cells. Sortase A anchors surface proteins, such as virulence factors, to the bacterial cell wall that interact with host tissues and activate bacterial invasion (Ton-That, 1999). By inhibiting sortase A, surface proteins that interact with mammalian host cells will no longer be anchored to the bacterial cell wall. Without these invasion proteins on its outer surface, *S. aureus* can no longer enter host tissue, preventing infection (Bierne, 2002). Our goal is to drastically reduce the time

required to design drugs that are capable of preventing bacterial infection, and to do this we attacked the problem from multiple directions.

First we chose to focus on an anti-infective approach and develop a small molecule drug using AI instead of manual human labor that could effectively inhibit sortase A and prevent bacterial infection *in vitro*. And secondly, we aimed to improve the physical testing mechanisms for these candidates, so that subsequent testing would be as efficient as possible. Existing computing solutions are inefficient and require significant manual input, and thus are only suitable for testing individual molecules. We sought to design a program that can be fully automated and test large libraries of fragments in a relatively short amount of time. Our program showed us the best performing small molecule fragments, which we then manually attached to our scaffold using Chemdraw to create our proposed drug candidate.

The goal of our project is to test these predicted molecules *in vitro* using an invasion assay established by Dr. Zhang's lab (Lu et al, 2007), and also optimize this assay through the implementation of quantitative polymerase chain reaction (qPCR) for analysis of drug efficacy. We want to build a protocol that would decrease the amount of time needed to run the invasion assay, allowing future researchers to assess drug efficacy faster and more easily. Our data so far is promising but needs more testing to reach a definitive conclusion.

By using qPCR instead of fluorescence detection, we hope to completely eliminate the need for a time-consuming fluorescence labeling step. Doing so would reduce the cost of the instrumentation needed for quantification too, as qPCR machines are relatively inexpensive compared to FACS machines and have often been purchased by many laboratories already anyways (Takamatsu, 2017). Finally, we want our protocol to increase the sensitivity of the invasion assay by detecting bacteria at the transcriptomic level.

1.5 Backup Plan

1.5.1 AI

If our program had given us inaccurate results in our testing, we would have re-evaluated the tools we were using and either modify the implementation, or simply use different tools. We would look first to other tools in the Autodock family. Autodock 4 is a well-established system for simulating ligand-protein interactions, and implementing it instead of Vina would not require a significant change to our architecture (Morris, 2009). We could also have reduced the scope of our AI in order to have a smaller search space. Though this would have had affected the quality of our results, it would still have acted as a strong proof of concept for our AI model.

1.5.2 Invasion assay optimization

In the event that we are unable to successfully implement qPCR into the invasion assay, we planned to investigate the possibility of quantifying the amount of bacteria that invaded the CHO cells using a gram-positive-specific dye. We would follow the original invasion assay protocol, but rather than using a fluorescent stain before the invasion, we would use a UV-Vis stain after the invasion was complete. CHO cells would be lysed after the invasion using a gentle homogenizer, while the *S. aureus* cells would be kept intact. The remaining *S. aureus* cells would be spun down, stained using a form of gram stain, and quantified using a plate reader.

1.6 Significance

Anti-infective drug design is a fairly new term for the process of developing drugs that inhibit the infectious mechanisms of bacterial cells and other harmful microbes. This type of drug design addresses the growing threat of antibiotic-resistant bacterial infections which are responsible for roughly 700,000 deaths worldwide (O'Neill, 2014) and is showing no signs of slowing down without a major shift in prevention tactics. Our project aims to address this need and even speed up the process. Currently, the average amount of time needed to push a new drug from the discovery phase to market authorization is 12 years, computer-aided drug design is capable of cutting this number down by several years because it greatly reduces the need for costly drug design and testing methods such as high-throughput screening (David et al, 2021).

1.7 Team and Management

1.7.1 Interdisciplinary team management

Our team was made up of three computer science and engineering (COEN) and two bioengineering (BIOE) students. The COEN half of the group was tasked with designing the AI program, while the BIOE half was tasked with optimizing the invasion assay used to test the drug molecules generated by the AI program. Though the two end products were separate, they were unified in the goal of making the process of drug design more efficient, and design choices of each were influenced by information from the other.

1.7.2 Timeline

Table 1. Project timeline for the academic year

	Fall Quarter	**Winter Quarter**	**Spring Quarter**
COEN	● Molecular docking ● AI brainstorming	● Developing AI program ● Becoming familiar with Chemdraw and AMDock	● Finish AI program ● Predict fragments
BIOE	● Invasion assay optimization brainstorming ● Begin CHO cell culture	● Zero experiment ● qPCR primer design ● Initial qPCR testing	● Establish and optimize qPCR protocol ● Test AI-generated molecule
Both	● Literature review	-	● Conference presentation ● Thesis writing

1.7.3 Materials and Budget

Our team was granted a total of $2,500 for this project ($1,000 for bioengineering materials and $1,500 for computer science and engineering materials). We also utilized materials already available in Dr. Zhang's lab.

Table 2. Project budget breakdown

Material	**Quantity**	**Price**

Laptops (Lenovo IdeaPad 3)	3	$500
ChemDraw	1 license	In storage
0.25% Trypsin-EDTA (Gibco)	3	$46.98
FBS, Heat Inactivated (Gibco)	500 mL	$654
10x PBS, pH 7.4 (Gibco)	500 mL	In storage
DMEM, High glucose, GlutaMAX Supplement (Cytiva)	10 x 500 mL	In storage
Trypan blue, 0.4% (Rs' Science)	1 x 10 mL	In Storage
DMSO (Fisher Scientific)	1 x 15 mL	In storage
TSB Media Powder	100 g	In storage
Gentamicin	1 x 15 mL	In storage
Penicillin/Streptomycin	1 x 15 mL	In storage
CHO cells	5 x 1 mL vials	In storage
RN4220 cells	1 x 1 mL	In storage
Triton X-100 (Bio Basic)	500 mL	In storage
Nuclease-free water, DEPC-treated (Ambion)	1 L	In storage
SYBR Green qPCR Master Mix (Bio-Rad)	2 x 1 mL	$152.29
Lysozyme (Bio-Rad)	1 x 500 µL	In storage
16S rRNA primers (IDT)	2 x ~200 µL	$21.50
S. aureus primers, SaQuant (IDT)	2 x ~200 µL	$21.50
Sortase A primers (Elim Bio)	2 x 50 µL	$7.86
Sortase B primers (Elim Bio)	2 x 50 µL	$7.86
Agarose powder (Bio-Rad)	1 x 125 g	In storage
Gelite™ Safe DNA Gel Stain (AAT Bioquest)	1 x 500 µL	In storage
Hemocytometer (Rs' Science)	1	In storage
Conical vials, 15 mL (CELLTREAT)	10 x 50/PK	In storage
Conical vials, 50 mL	500 tubes	$369

10 cm petri dishes, tissue-culture treated	25 x 20/PK	In storage
Pipette tips, 1000 µL, United Scientific Supplies	2 x 500/PK	$41.08
Pipette tips, 20 µL, United Scientific Supplies	1 x 1000/PK	$27.48
Pipette tips, 10 µL, United Scientific Supplies	1 x 1000/PK	$20.66
Serological pipette tips, 5 mL	1 x 200/PK	$36.70
Serological aspirating pipette tips, 5 mL	8 x 20/PK	In storage
96-well PCR plates (ThermoFisher)	1 x 25/PK	$138
Optical plate covers (Applied Biosystems)	1 x 25/PK	$89.50
Cryovials, 2 mL ()	2 x 50 tubes	$27.10
Eppendorf Tubes, 1.5 mL	200 tubes	In storage
CS&T RUO Beads	50 tests	$124
Globe Scientific Glass Culture Tubes	250/box	In storage
CellTreat Cell Strainer	70 µm	In storage

Chapter 2: Artificial Intelligence Platform Development

2.1 Design Goals

The cornerstone of the design philosophy of this project is anchoring the machine computation in human biological understandings. Both the target protein Sortase-A and the "scaffolding" drug candidates are built from products of traditional molecular biological engineering research. This has two-fold benefits: first it minimizes necessary research space as the AI is able to more efficiently predict drug candidates. Given the expense of computing time this is a worthwhile investment, both in time and money. Second, by building the drug candidates from scaffolds of known safety and effectiveness, the chances of compounds with unpredictable side effects can be reduced. This philosophy arises from the collected experience of our advisors in drug design.

Most AI systems are inherently uninterpretable, and therefore difficult to prove the reliability of. This software is no exception. Both of these traits carry ethical risks which our system must be designed around. The limited interpretability of internal AI mechanisms and strategies means that the AI cannot have its processes verified for thoroughness or freedom from biases. While a simpler system can be manually checked by a human, ML operations are too deeply abstracted for this to be possible. Biases which creep into the system are therefore difficult to detect. This is the danger of De Novo drug design, which is the unfettered design of drugs by AI systems. While some degree of De Novo design is necessary to create an AI, minimizing this will create the safest system possible.

Biases in the training data and docking software are likely to creep into the system. Testing in other analogous BioAI designs has shown the ability for dangerous scaffolds to produce similarly dangerous results (Urbina, 2022). However, the pursuit of perfection cannot be allowed to get in the way of progress. It is relevant here that there are likely a myriad of compounds which would be effective, and it is not necessary to find them all, or the optimal candidate. A sufficiently effective anti-infective drug is the goal of this project. So while a set of candidates to be tested for side effects is needed, biases which prevent a subset of candidates from being detected will have a negligible impact. With sufficient wet-lab testing and care in the post-AI design of drugs, the risks will therefore be acceptably minimized.

2.2 Software Architecture

The AI operates on the input data of small-molecule fragments. As shown in figure 3, the first step is the selection of an input library. Several preprocessing steps are undertaken in order to extract the files from their compressed library file. Then they are converted into a workable format. Finally, the real work can begin. Each individual fragment is scored on its stability and undergoes simulated attachment to the target protein. The best performing fragments are outputted, which is then manually sent to chemdraw to be combined and further tested.

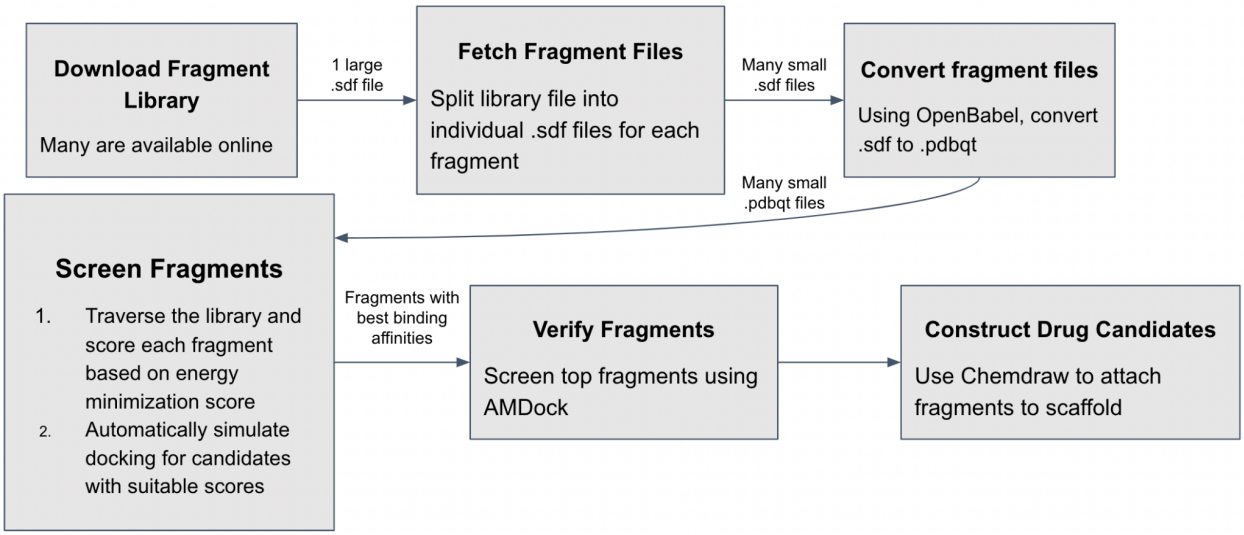


Figure 3. High level architecture of our software implementation

2.3 Software Components

There were several software modules used in the construction of the BioAI. Autodock Vina is the backbone of our project, as it is the framework we use to simulate the docking of fragments with Sortase A. OpenBabel is the tool which we used in order to convert input chemical files to workable types, allowing the rest of the process to occur.

Several helper pieces of software were important to the design process of the AI, although they are not directly included in the platform. PyMol was critical, as it allowed the viewing of 3-dimensional representations of the molecular structures. Although this view is not directly tied to the computation of binding, it allows for useful human checks of the AI's output. AMDock is a popular docking simulator, which we used to verify our results. Chemdraw allowed the creation of virtual representations of drug candidates from fragments and scaffold for further testing.

We used several drug fragment libraries, selecting those which were recommended by our advisors. The Maybridge Corporation's Screening fragments (~ 300 fragments) and General fragment library (~ 400 fragments) as well as Life Chemicals' General fragment library (~ 500 fragments) were our main sources of data.

2.4 BioAI Results

After screening fragments from the Maybridge Screening fragments library, the Maybridge General fragment library, and the Life Chemicals General fragment library, we selected the top 3 fragments and verified their affinities with Sortase A using AMDock. Then, using Chemdraw, we constructed our drug candidate, exported the chemical structure file, and simulated the chemical docking with Sortase A. Our goal for our drug candidate was for it to have a higher affinity than the candidate discovered by previous groups through wet lab testing. Our result was that our drug candidate's affinity exceeded the previous group's by 30.4%. Our goal in terms of efficiency was to have our program be able to screen large libraries of fragments in a relatively short amount of time compared to existing programs. We found that our platform can test a fragments' binding affinity with Sortase A in an average of 30 seconds, a huge improvement compared to AMDock, which takes about 4 minutes on average to perform comparable tests.

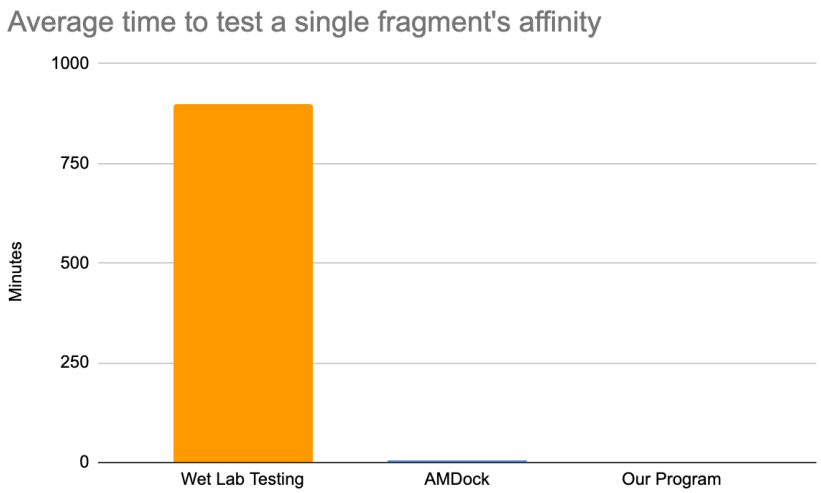


Figure 4. The time required to test the binding affinity of a single fragment with the target protein using 3 different methods: Wet Lab testing, AMDock, and our program

As shown in figure 4, testing fragments using wet lab testing takes a minimum of 15 hours. Testing fragments using AMDock takes an average of 4 minutes. Our program takes an average of only 30 seconds.

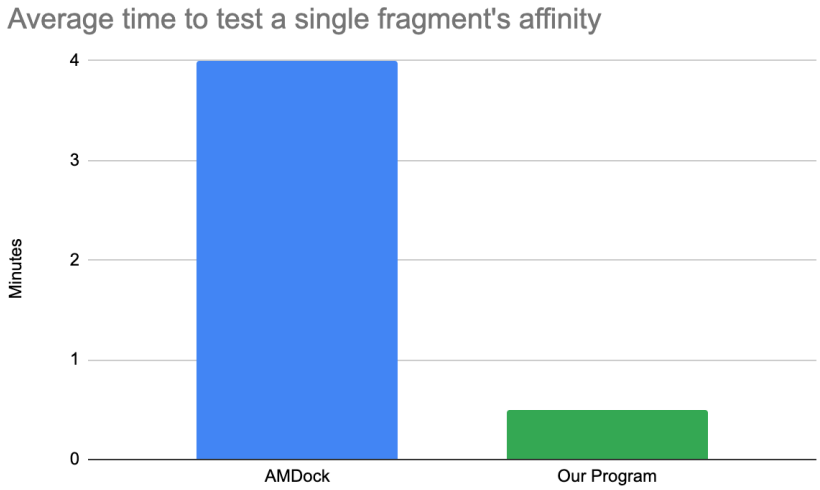


Figure 5. The time required to test the binding affinity of a single fragment with the target protein using AMDock as compared with our program

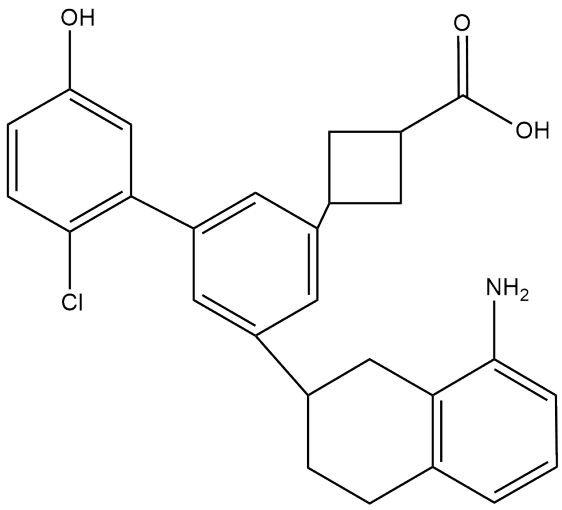


Figure. 6. The chemical structure of our best performing molecule

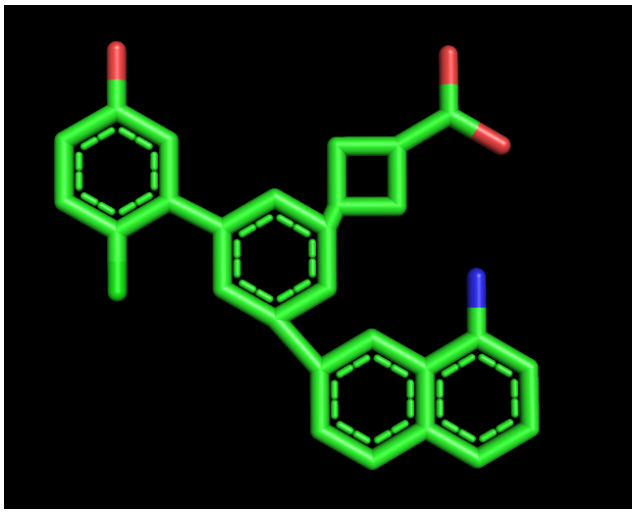


Figure. 7. Visualization via Pymol of our best performing molecule

Figure. 8. A 3-dimensional representation of the best performing molecule docked with Sortase A. The molecule can be seen at the bottom right corner of the image.

Chapter 3: Invasion Assay Optimization

3.1 *S. aureus* culture

A strain of *S. aureus* called RN4220 was used for infection of CHO cells in the invasion assay. This strain has been genetically modified through the use of chemicals and radiation so that it can accept foreign plasmids for cloning (Nair et al, 2011). RN4220 was obtained from Dr. Zhang's lab, and was grown on petri dishes coated in tryptic soy broth (TSB) agar. The RN4220 stock was removed from the -80°C freezer and briefly thawed, then streaked on an agar plate in a zigzag pattern using an inoculation loop. The plate was wrapped in parafilm and incubated upside down at 37°C overnight, and then stored in a 4°C fridge.

3.2 CHO cell culture

3.2.1 Recovery

CHO cells were recovered from liquid nitrogen storage before performing the invasion assay. They were thawed by warming for 60 seconds in a 37°C water bath. It is critical to ensure that the time spent in the 37°C water bath is kept to a minimum, as the dimethyl sulfoxide (DMSO) present in the cryopreservation media is toxic to cells. Thawed cells were slowly added to a 15 mL falcon tube containing 5 mL of complete growth medium (DMEM containing 10% fetal bovine serum (FBS)). The tube was then centrifuged at 3000 rpm for 1 minute to collect the cells at the bottom. The supernatant was removed, and the cell pellet was resuspended in 2 mL of growth medium. The resulting cell suspension was then added to a 10 cm plate containing 5 mL of growth medium.

3.2.2 Passaging

CHO cells were cultured in a growing medium consisting of DMEM with 10% FBS. FBS concentration was changed multiple times throughout the project with the goal of optimizing cell growth, but ultimately 10% proved to be the optimal concentration. While antibiotics such as penicillin and streptomycin are typically added to mammalian cell culture media, we intentionally left them out of our media since they would prevent bacterial infection during the invasion assay. Prior to performing the invasion assay, CHO cells were incubated at 37°C and 5% CO_2, and passaged every 3-4 days after reaching ~80% confluence.

To passage a 10 cm plate of CHO cells, the old growth medium was first removed through aspiration. 4 mL of PBS was then added to the plate, which was subsequently tilted five times to remove leftover serum and dead cells. The PBS was then aspirated and replaced with 2 mL of 0.25% Trypsin-EDTA. The plate was incubated at 37°C for 3 minutes to allow sufficient time for cell detachment. After incubation, 5 mL of growth medium was added to the plate to neutralize the trypsin and help wash off cells that may not have fully detached yet. The resulting cell suspension was collected and centrifuged in a 15 mL falcon tube at 3000 rpm for one minute. The cell pellet was resuspended in 3 mL of growth medium and split evenly between three 10 cm plates each containing 6 mL of growth medium.

Due to the lack of antibiotics in our growth medium, we did see occasional bacterial contamination in our cell culture plates (**Appendix A**). To address this, we began adding penicillin and streptomycin to a final concentration of 50 µg/mL during the spring quarter. The growth medium containing antibiotics was removed approximately one passage before performing the invasion assay and replaced with growth medium containing no antibiotics.

3.2.3 Cryopreservation

To ensure that we always had cells throughout the project, we routinely expanded and froze down aliquots of our CHO cells. To freeze one 10 cm plate of cells, the cells were trypsinized and collected using the same protocol used for passaging. However, the cell pellet was instead resuspended in 1 mL of freezing medium (DMEM with 10% FBS and 20% DMSO). DMSO decreases the electrolyte concentration in the medium surrounding the cells to help prevent crystallization during freezing, which can lead to apoptosis (Jang et al., 2017). This cell suspension was then transferred to a 2 mL cryovial, which was frozen overnight at -80°C and subsequently transferred to liquid nitrogen storage. The cryovial was kept in liquid nitrogen storage until required for the invasion assay.

3.3 Original invasion assay (zero experiment)

3.3.1 Zero experiment

This invasion assay protocol is intended to be run with the addition of a drug molecule expected to prevent bacterial infection. However, for our purposes we ran the assay without adding a drug in order to obtain a baseline quantification of *S. aureus* infection of CHO cells as

well as gain practice running the assay. Running the original assay also aided while we brainstormed ideas for optimization. The zero experiment consisted of six samples: three with fluorescent-stained bacteria, and three with unstained bacteria. The unstained bacteria samples serve as a control to ensure that the fluorescent staining worked as expected, as well as to check for possible cross-contamination. If successful, there will be a distinct difference between the two sample conditions during flow cytometry analysis (Lu et al, 2007).

3.3.2 CHO cell plating and quantification

CHO cells were prepared the night before the invasion assay. Depending on the number of samples needed for the assay, 2-4 10 cm plates of CHO cells at ~70-90% confluence were trypsinized and collected in one tube. Cell suspension concentration was quantified using a hemocytometer. For hemocytometer quantification, 15 µL of the cell suspension was added to an Eppendorf tube containing 15 µL of 0.5% trypan blue solution, which helps assess cell viability. The cells in each square of the hemocytometer were counted using a microscope set to 10X magnification. CHO cell concentration was calculated using the following equation:

$$Total\ cells/well = \frac{Total\ cell\ count * initial\ volume\ of\ CHO\ cells * dilution\ factor * 10^4}{4}$$

The cell suspension was split equally into the wells of one or more six–well tissue culture plates, each of which contained 1 mL of growth medium. Each well received 0.5 mL of cell suspension to bring the total volume to 1.5 mL. The plate(s) were incubated at 37ºC overnight. On the day of the invasion, the growth medium was removed from each well and replaced with 1 mL of DMEM.

3.3.3 RN4220 culture and quantification

S. aureus strain RN4220 was cultured in suspension the day before the invasion. [%] TSB media was prepared by adding 1.5 g of TSB to 50 mL of DI water, and then autoclaving at 121ºC for 15 minutes. 4 mL of TSB media was added to each of four culture tubes, and one RN4220 colony was added to each using an inoculation loop. The tubes were incubated on a shaker at 225 rpm and 37ºC for approximately 6 hours, then stored at 4ºC overnight until the invasion. The tubes were incubated under the same conditions for one more hour the morning of the invasion to increase cell viability again after refrigeration. Each tube's absorbance at 600 nm (OD600) was

measured, and ideally between ~0.4-1.0 to ensure that the bacteria are in the exponential phase of their growth curve.

3.3.4 CFSE labeling of RN4220

Once confirmed to be within the optimal OD600 range, the *S. aureus* suspensions were consolidated into two 15 mL falcon tubes and centrifuged at 5000 rpm for 10 minutes. The supernatant was removed and each pellet was resuspended in 2 mL of PBS. Next, a 5 mM carboxyfluorescein succinimidyl ester (CFSE) stock solution was prepared by adding 18µL of DMSO to a separate vial of CFSE. This vial and all other vials containing the light-sensitive fluorescent dye were wrapped in aluminum foil to protect the integrity of the dye and help prevent photobleaching. One tube of bacteria was stained with 8µL of fluorescent dye and the other tube was left unstained as the negative control. Both tubes were then incubated for one hour on a shaker at 225 rpm and 37ºC.

After incubation, the bacterial cell culture tubes were centrifuged at 5000 rpm for 10 minutes. The unstained *S. aureus* pellet was resuspended with 2 mL of PBS, while the CFSE-stained bacteria were washed three times with 2 mL of PBS using a cycle of centrifugation and PBS resuspension. Finally, the OD600 of both tubes was measured to determine bacterial concentration. The concentration of RN4220 in each tube was calculated using the following equation:

$$Total\,bacteria\,cells/mL = OD\,600 \times 8 \times 10^{8}$$

3.3.5 Invasion

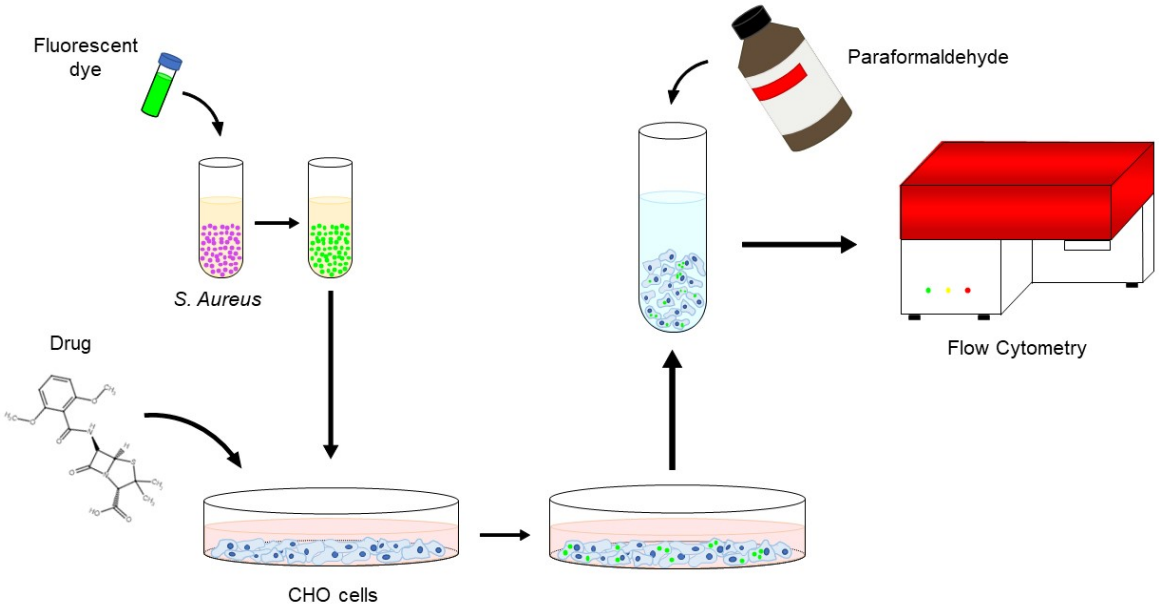


Figure 9. Original invasion assay developed by Dr. Zhang's lab (Lu et al, 2007).

In order to achieve notable and equal bacterial invasion, we calculated the volume of bacteria cells that needed to be added to each well of CHO cells. To achieve a multiplicity of infection (MOI) of 40, we used the equation below to calculate the volume of bacteria needed from each tube.

$$Volume\ of\ bacteria\ needed = \frac{Total\ CHO\ cells/well}{Total\ bacteria\ cells/mL}$$

The specified volume of unstained bacteria was added to three wells of the six-well plate, while the other three wells received the specified volume of stained bacteria. The six-well plate was then wrapped in tin foil and incubated at 37ºC for one hour. After being given sufficient time to infect the CHO cells, the bacterial invasion was terminated by adding 6µL of gentamicin to each well and incubating at 37ºC for 45 minutes.

After termination, the CHO cells were washed with 2mL of PBS three times to remove any debris and bacteria that didn't invade the cells. The remaining CHO cells were then detached using 0.5 mL of 0.25% Trypsin-EDTA and collected using 2 mL of growth medium. The cell suspension collected from each well was placed in its own tube, with the tubes containing stained samples being wrapped in aluminum foil. These six tubes were centrifuged at 1340 rpm for 5

minutes, and the pellets were resuspended in 4% paraformaldehyde (PFA). PFA initiates cross-linking of the molecules within the infected cells, hardening the cell surface and preserving its shape and contents (Kim et al., 2017). PFA fixation helps preserve cell structure before and during flow cytometry analysis. Cells were resuspended in a volume of PFA solution that resulted in a final concentration of 10^6 cells/mL and incubated at room temperature for 15 minutes. The samples were then spun down at 1340 rpm for 5 minutes. The supernatant was discarded, and each sample was resuspended in 1 mL of PBS.

3.3.6 Flow cytometry

Flow cytometry analysis was performed using the BD™ Accuri C6 Plus flow cytometer. The instrument was set to 10,000 events for all settings, then cleaned using 2 mL of FACS clean solution (BD) and calibrated using 2 mL of DI water containing two drops of CS&T RUO beads. Once successfully calibrated, 1 mL of each sample was pipetted through a CellTreat cell strainer into a glass culture tube. The software was set to generate plots with FITC-A on the x-axis, and the graphs were gated appropriately after running the first sample. A SIP clean step was performed before running each sample to limit potential cross-contamination.

3.4 Optimized invasion assay

3.4.1 Invasion

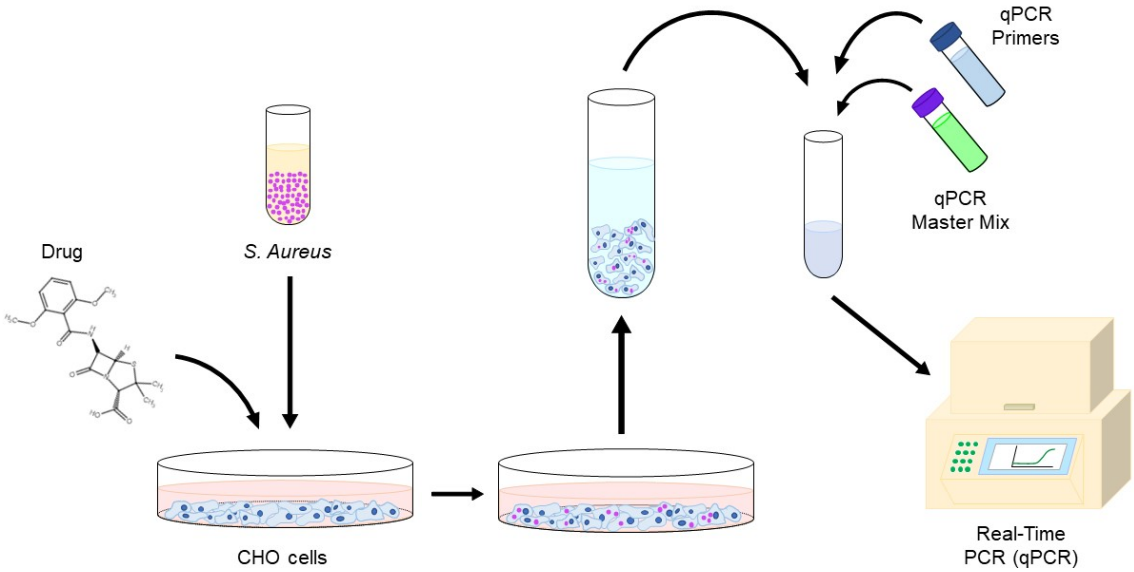


Figure 10. Our proposed optimization of the invasion assay.

The invasion portion of the experiment is intended to be performed almost exactly as described in 3.3.4, with the elimination of a few steps. The CFSE labeling step is not necessary in our optimized protocol, since we are analyzing genomic DNA rather than whole cells. The PFA fixation step is also unnecessary for this same reason, since we want to release the nucleic acids from inside the cells rather than keeping everything preserved inside. After collection and centrifugation at the end of the invasion, the pellets will be resuspended in 1 mL of PBS and can be stored at -80°C until use without significantly compromising DNA integrity (Kim et al, 2011).

3.4.2 qPCR reaction

Our optimized invasion assay uses quantitative polymerase chain reaction (qPCR) to analyze infection efficiency rather than flow cytometry. The qPCR reaction consisted of the reagents listed in **Table 3** and followed the thermal cycler program described in **Table 4**. A Bio-Rad CFX Connect thermal cycler was used for all experiments, and each sample was run in triplicate to ensure consistent results.

Table 3. qPCR reaction components

Reagent	Volume	Final concentration
SYBR Green Master Mix (2x)	5 µL	1x
Forward primer (10 µM)	1 µL	1 µM
Reverse primer (10 µM)	1 µL	1 µM
Sample	3 µL	Variable

.

Table 4. qPCR thermal cycler program.

Cycles	Temperature (°C)	Time
1	95	3 min
40	95	10 sec
	60	20 sec
1	65	5 sec

3.4.3 Primer design

Multiple sets of primers were tested in order to troubleshoot non-specific amplification. Primer sequences were either obtained from published work (Wood et al, 2021) or designed using an online primer design program. We designed primers that individually targeted the 16S rRNA, sortase A, and sortase B genes using Primer-BLAST (NCBI, "Primer BLAST"). The specificity of these primer sequences for RN4220 was confirmed using Nucleotide BLAST (NCBI, "Standard Nucleotide BLAST"), as well as the nonspecificity for the Chinese Hamster genome.

3.4.4 Agarose gel electrophoresis

A 2% agarose gel was used for electrophoretic analysis in order to obtain better resolution of smaller DNA fragments, as our expected PCR products were less than 300 bp in length. 2% agarose was made by mixing 1 g of molecular biology grade agarose with 50 mL of

1x TAE buffer. The solution was microwaved for 1 minute, and then again in 15 second intervals until the agarose powder was completely dissolved. Once cool to the touch, 5 µL of Gelite™ Safe DNA Gel Stain was added to the solution. The solution was then poured into a gel mold, and an 8-sample comb was inserted. The gel was allowed to sit for approximately 20 minutes, after which the comb was removed and the gel was placed in the electrophoresis tank.

The tank was filled to the max line with a 1x TAE buffer, making sure that the agarose gel was completely covered. Samples of interest were selected and mixed with 2 µL of 6x Purple Gel Loading Dye (no SDS), then 10 µL of each were loaded into the gel. 10 µl of Quick-Load Purple 1 kb Plus DNA Ladder was added to the first lane of the gel and used to determine approximate fragment size. The electrophoresis tank was connected to a power source set to output 120V, and was run for approximately 45 minutes or until the loading dye was ~2-3 cm from the end of the gel. Gels were imaged using a UV gel imager from Azure Biosystems set to 302 nm, programmed for a 10 sec exposure.

3.4.5 qPCR data analysis

qPCR data was analyzed by looking at the C_q values of each sample. Data was prepared in graph form by taking the reciprocal of each C_q value, so that increases and decreases in value would be shown in a way that was easier to interpret visually.

3.5 Results

3.5.1 Zero experiment

Before beginning qPCR experiments, a "zero experiment" was run using the original invasion assay protocol to establish a baseline for bacterial invasion in the absence of an inhibiting drug molecule. The zero experiment protocol was established by Dr. Zhang's lab in 2015. For this experiment, the BD Accuri™ C6 Plus Flow Cytometer was used to measure the fluorescence intensity in samples of RN4220 *S.aureus* stained with CFSE dye and added to a six-well plate containing CHO cells. Three wells represented our negative control which had CHO cells infected with non-stained bacteria and the other three represented our positive control with CHO cells infected with CFSE-stained bacteria.

We performed this experiment four times throughout the school year and were able to gain success and valuable information from our final trial. The negative control seen in **Fig. 11**

showed that the peak on the left and the gating indicate that 95.8% of the cells did not fluoresce, which is expected of the non-stained bacteria. **Fig. 12** indicates 52.1% of the bacteria in the positive control was stained and had successfully invaded the mammalian cells. In order to better understand the amount of bacteria within the CHO populations, we calculated the average uptake index of each triplicate group and normalized it. These calculations were consistent with what the flow cytometry graphs showed. Uptake index is a measurement of the bacterial internalization and an indicator of sortase A activity on the cell membrane (Zhu et al, 2016); uptake index is calculated by:

*Uptake Index (UI) = Mean FITC-A * Percentage of Fluorescent Molecules*

Table 5: Zero Experiment Calculations

		Raw Data			Processed Data		
	Wells (Six-well Plate)	**Mean FITC-A**	**FITC-A Stained Percentage**	**Percentage of area (%)**	**Uptake Index (UI)**	**Average UI**	**Normalization**
Negative Control, non-stained	A1	565.68	73.24	0.46	414.304032	917.138518	0.005092325188
	A2	3,387.42	40.91	5.14	1385.793522		
	A3	2,676.00	35.55	0.82	951.318		
Positive Control, stained	B1	54,971.66	306.7	52.07	168598.0812	180102.1113	1
	B2	141,801.80	145.68	51.88	206576.8622		
	B3	53,722.23	307.38	51.98	165131.3906		

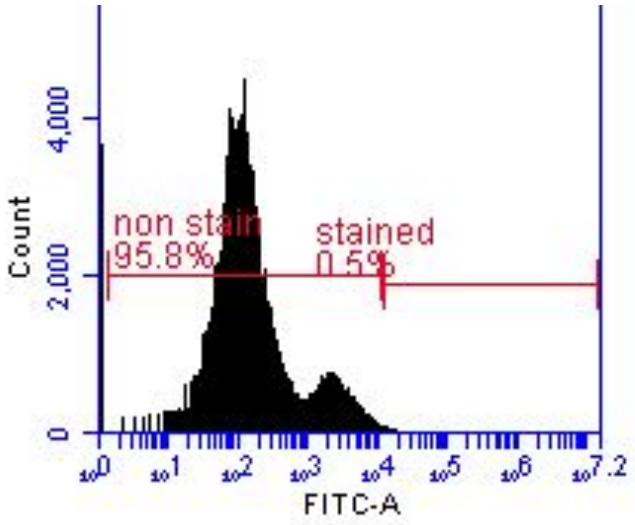


Figure 11: Negative Control: Non-stained bacteria. The x-axis, FITC-A, represents the fluorescent level from low to high. The y-axis represents the cell count.

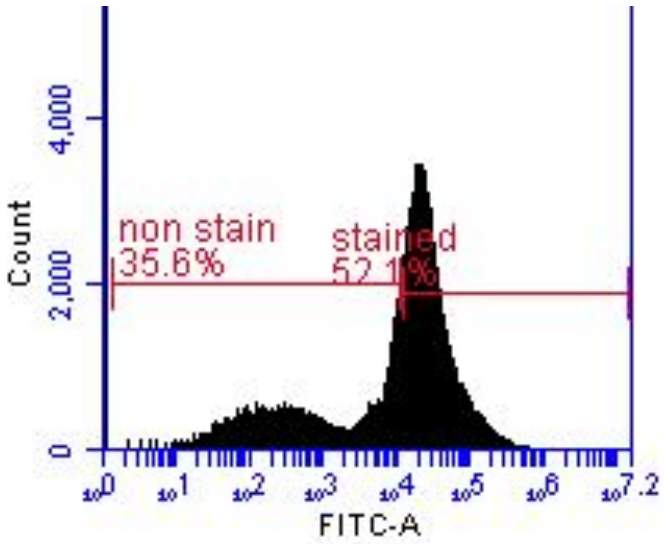


Figure 12: Positive control: Stained bacteria. The x-axis, FITC-A, represents the fluorescent level from low to high. The y-axis represents the cell count.

3.5.2 Primer optimization

3.5.2.1 16S rRNA gene primers

Initial experiments used primers targeting the 16S rRNA gene of *S. aureus*. The 16S rRNA gene was selected because it is a popular target for bacterial quantification, and there are several well-established qPCR assays for targeting it (Galazzo et al, 2020). While these primers

did produce amplification in RN4220 samples, they also produced amplification in NTC samples containing water instead of RN4220. This is likely because the 16S rRNA gene is present in all bacterial genomes, and is relatively similar between species (Galazzo et al, 2020). Because of this, any amount of bacterial contamination in the NTC samples would likely result in amplification.

3.5.2.2 SaQuant primers

Because of the amplification seen in the NTC samples when using the 16S rRNA gene primers, we decided to test primers specifically targeting *S. aureus*. We used primer sequences from a published method for bacterial quantification using qPCR developed by Wood et al, titled SaQuant. These primers had been optimized to be as specific for *S. aureus* as possible, which we hoped would help reduce the amount of optimization time needed for our assay. The product they produce is expected to be 73 bp long (Wood et al, 2021). While we used the primer sequences presented in this paper, we did not use the same thermal cycler conditions; we utilized a program confirmed to work by others in our lab. We also used SYBR Green for fluorescent labeling of nucleic acids instead of probes.

These primers successfully amplified *S. aureus* genes without producing any amplification in NTC samples. qPCR was also found to be sensitive enough to distinguish between serial dilutions of *S. aureus* when using these primers (**Fig.13**). Previously prepared invasion assay samples were also tested with these primers, and were found to have amplification. However, amplification of a non-specific product was also produced in samples containing only CHO cells (**Fig 14.**). Performing gel electrophoresis on these three samples revealed a product in the samples containing CHO cells that was smaller than expected (**Fig. 15**). These results raised questions about the specificity of our primers, and how well they were able to successfully target *S. aureus* without amplifying mammalian genomes. It is critical that our qPCR assay be able to distinguish between bacterial and mammalian genomes when analyzing invasion assay samples, so further optimization was required.

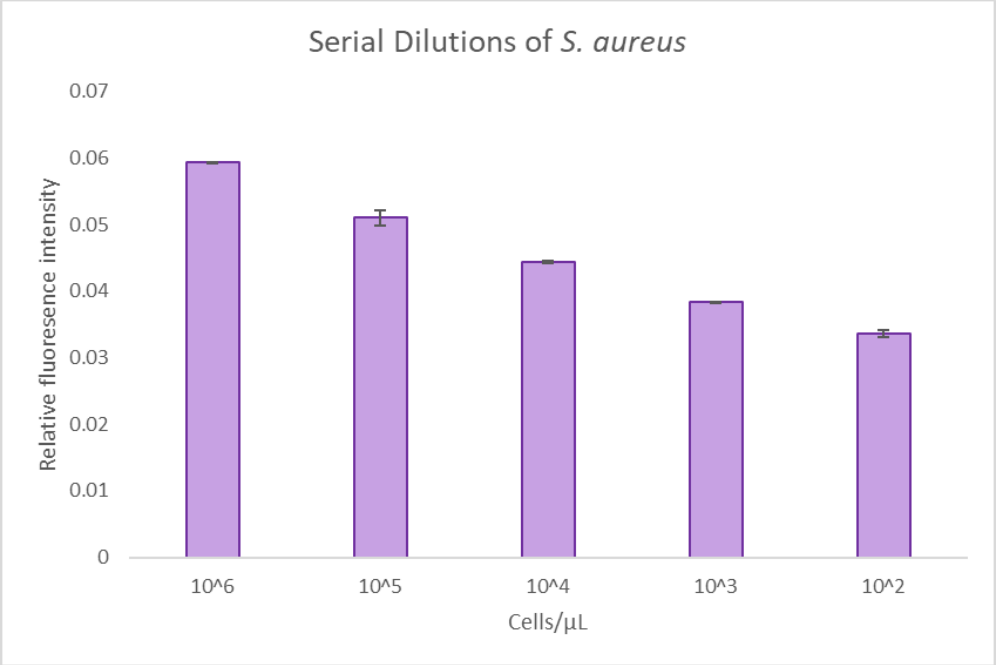


Figure 13. Relative amplification of serial dilutions of *S. aureus*.

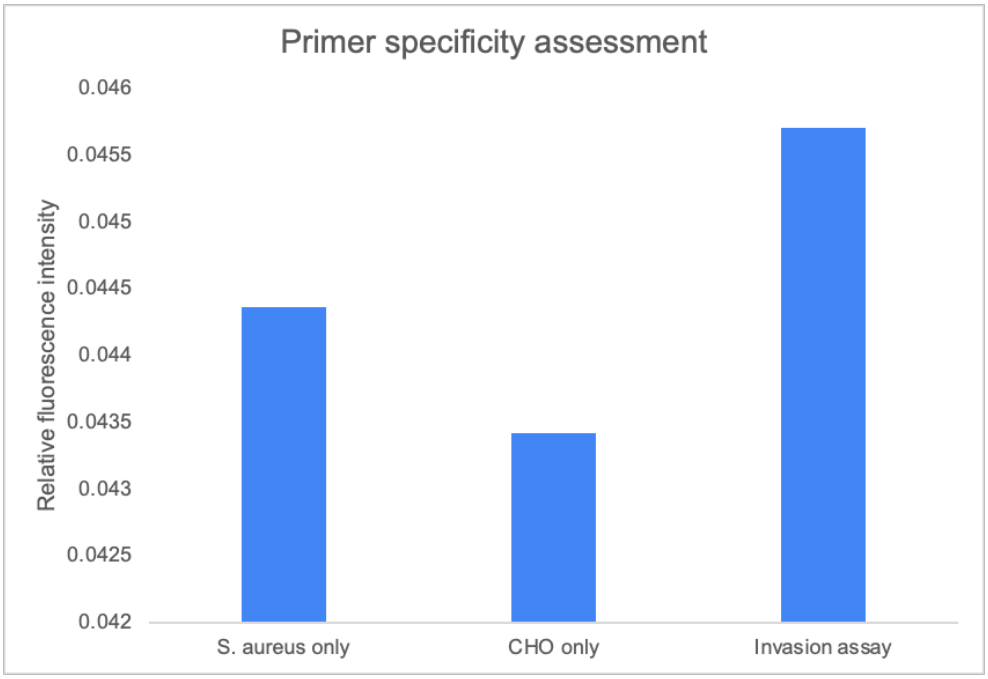


Figure 14. Relative amplification of *S. aureus*, CHO cell, and invasion assay samples.

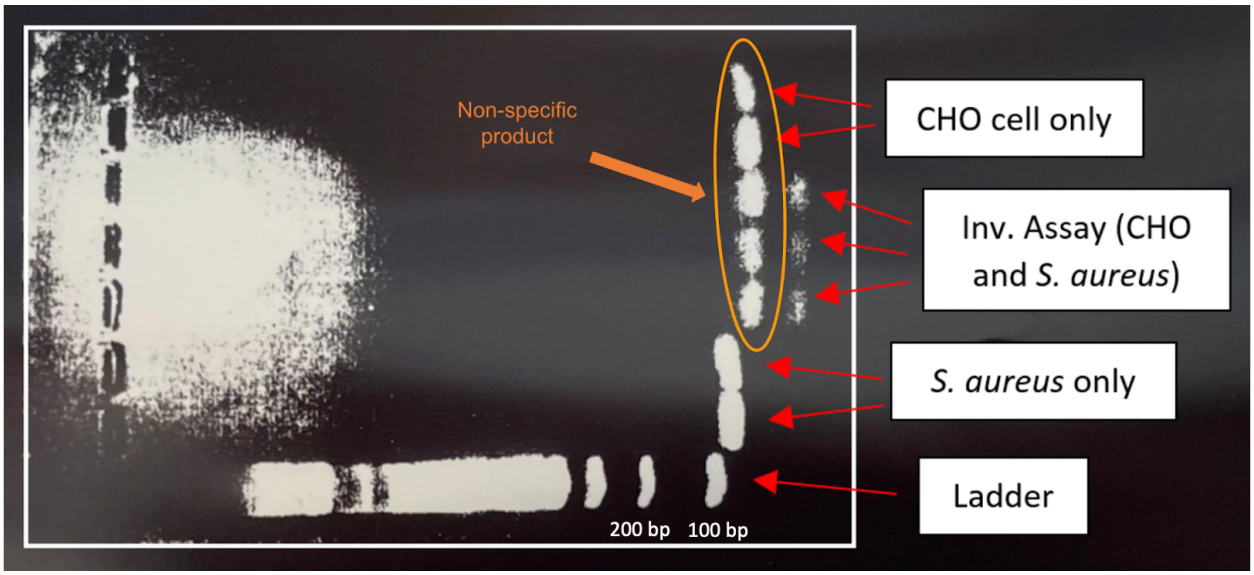


Figure 15. Agarose gel showing PCR product sizes of *S. aureu*s, CHO cell, and invasion assay samples.

Primers were aligned with both the RN4220 genome and the Chinese Hamster genome using Nucleotide BLAST (NCBI, "Standard Nucleotide BLAST") to check for possible specificity issues. Specificity was evaluated by comparing the E-values of each alignment. A lower E-value correlates with a higher alignment, while a higher E-value correlates with a worse alignment. Both the forward and reverse SaQuant primers were found to have high alignment with the RN4220 genome and relatively low alignment with the Chinese hamster genome (**Table 6**), indicating that these primers should be specific for RN4220. This is contradictory to the results we achieved during wet lab testing, and requires further investigation. However, it should also be considered that the genome of the CHO cells we grow in our lab will likely be different from the CHO cell genome recorded in online databases like NCBI.

Table 6. SaQuant primer alignment with the RN4220 and Chinese hamster genomes. Specificity was evaluated by looking at the E-values of each alignment.

Primer	**Genome**	**E-value**
SaQuant (F)	RN4220	10^{-5}
	Chinese Hamster	3.6

SaQuant (R)	RN4220	10^{-7}
	Chinese Hamster	7.3

3.5.2.3 Sortase A and Sortase B primers

New primer sets were designed in the hopes of eliminating the non-specific amplification seen in samples containing CHO cells. One set targeted the sortase A gene and produced a 232 bp product, while the other targeted the sortase B gene and produced a 224 bp product. We chose to target the sortase A gene because of its specificity to gram-positive bacteria and to get a closer look at the potential inhibition mechanism of our small molecule drug. The sortase B primers were chosen because they indicated a lower alignment to CHO cell lines than the sortase A primers and we wanted to see if there was a significant difference in testing.

Table 7. Sortase A & sortase B primer alignment with RN4220 and CHO cell genomes. Specificity was evaluated by analyzing the E-values of each alignment.

Primer	**Genome**	**E-value**
Sortase A (F)	RN4220	10^{-5}
	Chinese Hamster	4.6
Sortase A (R)	RN4220	10^{-6}
	Chinese Hamster	5.5
Sortase B (F)	RN4220	10^{-5}
	Chinese Hamster	18
Sortase B (R)	RN4220	10^{-5}
	Chinese Hamster	4.6

Both sets of primers were tested on samples *S. aureus,* CHO, or invasion assay samples. qPCR was run as before, and then samples were selected for gel electrophoresis analysis (**Fig. 16**). While the expected products were present in the *S. aureus* samples, they were not present in the CHO cell or invasion assay samples. While this result makes sense for the CHO sample, there are still nonspecific bands present on the gel. This is a problem because these amplicons will show up as fluorescent during qPCR (due to the SYBR green), and make it difficult to differentiate between CHO and *S. aureus* based on qPCR results alone. The invasion assay samples also do not show the expected *S. aureus* product, which raises questions about whether or not *S. aureus* was actually present in these samples to begin with. The success of the invasion assay was not verified with fluorescence microscopy the day these samples were collected.

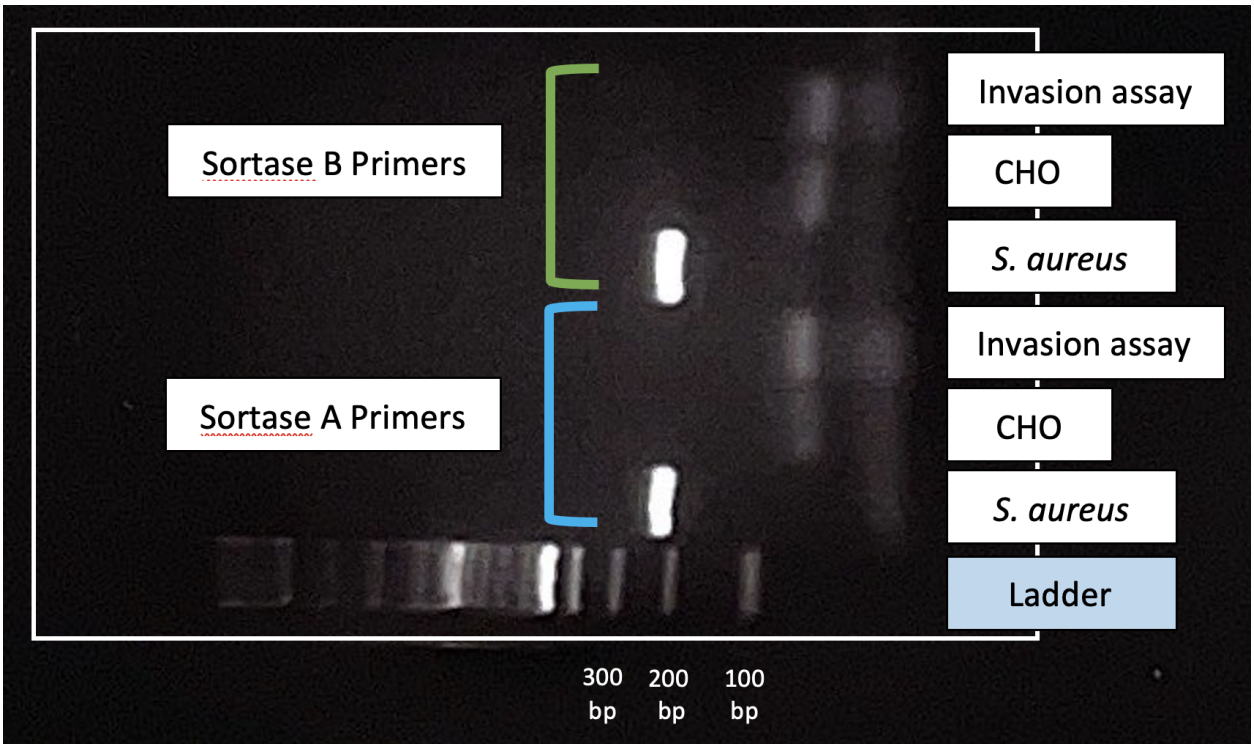


Figure 16. Gel electrophoresis analysis of samples amplified with either sortase A or sortase B primers. Samples are compared to a 1 kbp Plus ladder.

While not successful at completely eliminating nonspecific amplification, both sets of primers do work in *S. aureus* and produce the expected product. These preliminary results

suggest that our primer design may not be the issue, and that we should examine our reaction conditions and experimental samples.

3.5.3 Cell lysis optimization

Cell lysis conditions were optimized to ensure enough DNA was present for the iTaq polymerase to amplify. Initial experiments compared samples of RN4220 at a concentration of 10^6 cells/mL with no preparation, heat treatment, a freeze-thaw cycle, and lysozyme treatment plus freeze-thaw cycle. Heat treatment was performed by heating 20 µL of RN4220 at 95ºC for 10 minutes, while freeze-thaw cycle was performed by freezing 20 µL of RN4220 at -80ºC for 10 minutes and then immediately heating at 95ºC for 10 minutes. Lysozyme treatment involved the addition of 2 µL of lyophilized lysozyme (Bio-Rad, "PCR Troubleshooting") to 20 µL of RN4220, then the completion of a freeze-thaw cycle as described above. The samples treated with lysozyme did not show any amplification, which we hypothesized could be due to interference of the enzyme with the PCR reaction. There were no significant differences in amplification between the baseline samples, the heat treated samples, and the freeze-thaw cycle samples when analyzing our preliminary data (unpaired *t*-test, $p>0.05$) (**Fig. 17**).

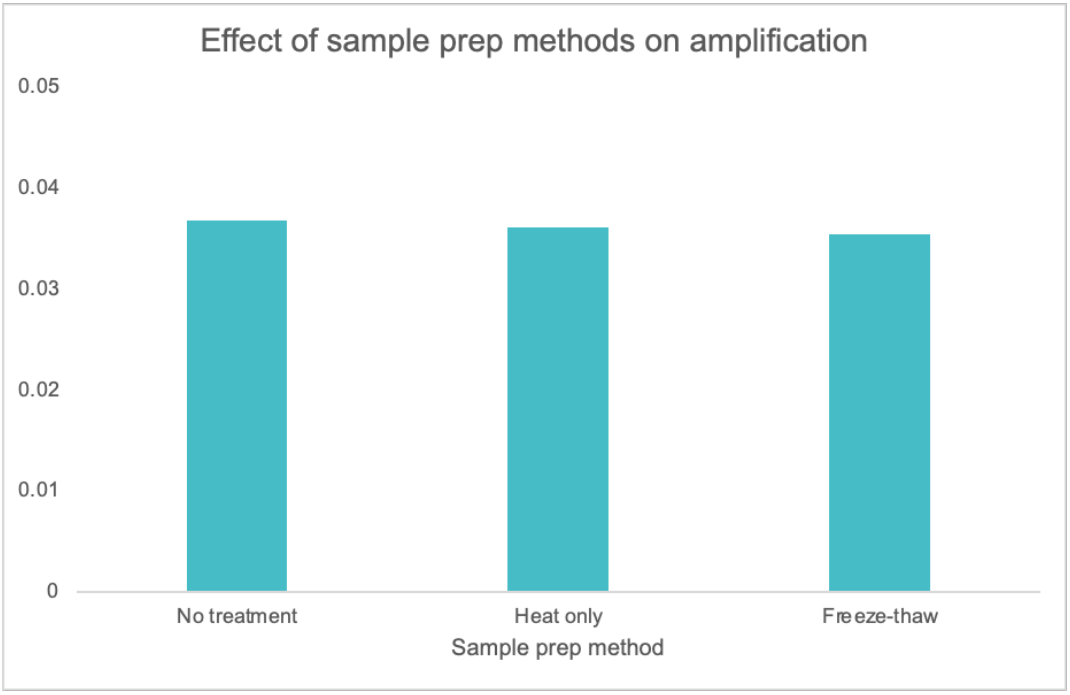


Figure 17. Effect of sample preparation methods on relative amplification. Differences between C_q values of the different treatment methods were not significant (unpaired *t*-test, $p>0.05$).

A detergent-based lysis buffer was also tested as part of this optimization. 0.5%, 1%, 2%, 5%, and 10% solutions of Triton X-100 were prepared by diluting with nuclease-free water (Ambion). 1 ml aliquots of RN4220 were spun down at 5000 RPM for 10 min, then resuspended in one of the prepared lysis buffers. Triton X-100 lysis buffer was found to produce inconsistent C_q values as well as abnormal-looking amplification curves in some samples (**Fig. 18**). These results suggested that may Triton X-100 interfere with the qPCR reaction, and that it may be better to use samples of RN4220 that had not been prepared with detergent-based lysis buffers. More aggressive lysis methods were not attempted because sufficient amplification was obtained without them, and also because these methods are more time-consuming. The goal of this optimization was to reduce the time needed to perform the invasion assay.

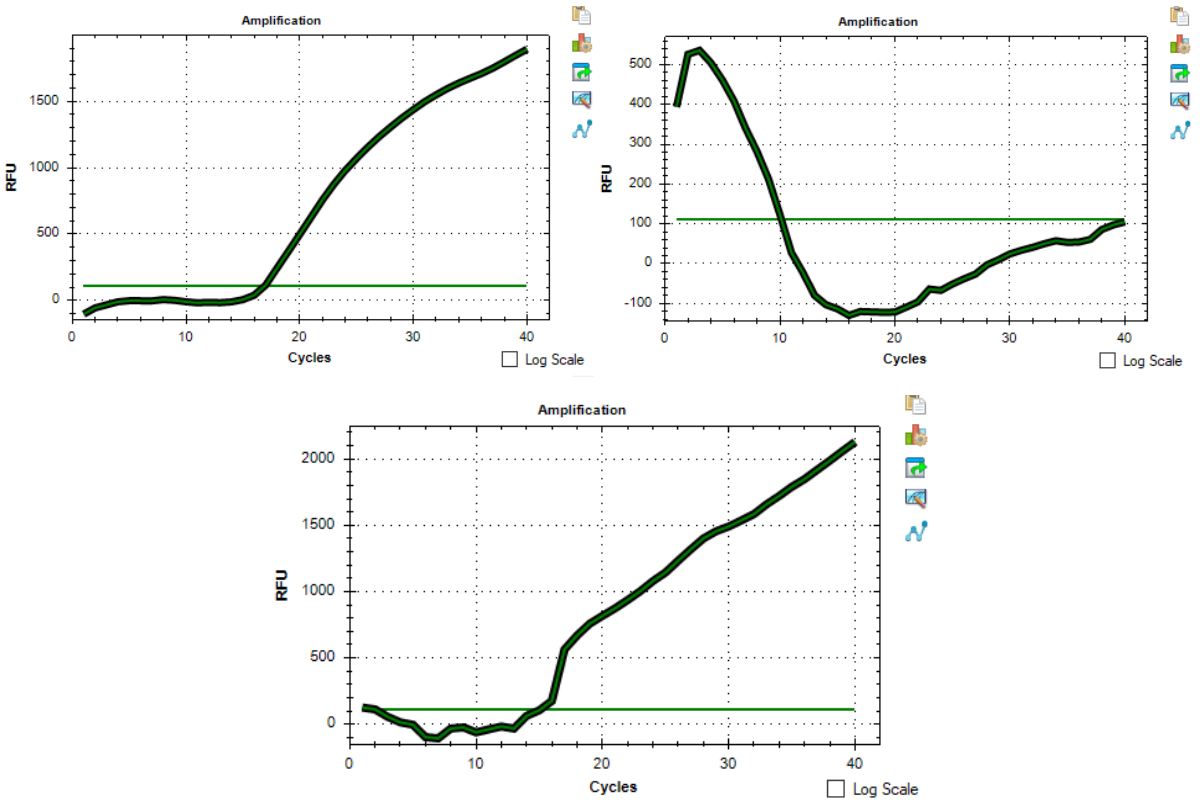


Figure 18. Comparison of amplification curves with 0% (top left), 1% (top right), and 10% (bottom) Triton X-100. These graphs are representative of a majority of the samples from this experiment.

Chapter 4: Discussion and Future Directions

4.1 AI

There are many components and facets of the AI which have room for growth. Due to the youth of the field and expense of software tools, there were significant barriers to the development and research of the system. While our tool did provide advantages in a limited scope, a more complete system which automates the next step of fragment combination would significantly raise the value of the system. This was infeasible due largely to a lack of training data: while data on the structures of small molecule fragments was available, reliable ground truth data on how they would interact with sufficient complexity was sparse. This left a gap in the research which should be investigated by a later project. However, our research was meant to establish the potential for this method of drug design. So while there are some aspects for which further research is required, that goal was accomplished.

In terms of future directions, there are three major directions of future software research we would recommend pursuing. These are automatic fragment combination, new target selection, and the creation of a GUI. Each of these would yield enormous benefits to the software's usage, in different ways.

Automation of fragment combination is the logical next step in the AI's progression. The industry software tools which would enable this were beyond our budget scope, but it is theoretically possible and would be another huge step forward in accelerating the drug design process. An equally potent step forward would be attempting to replicate the process we undertook with targets other than Sortase A. There exist hundreds if not thousands of dangerous proteins which an effective binding molecule could defang, and approaches like the one we tested here could be the solution.

4.2 Invasion assay optimization

Despite being specific for the RN4220 genome and not for Chinese hamster genome during Nucleotide BLAST analysis, both the SaQuant primers and the sortase A/B primer sets appeared to produce non-specific amplicons when CHO cells were present in the sample. One possible reason for this is contamination of our CHO cell samples, though this seems unlikely since the CHO cells we used appeared to be clean at the time of trypsinization. It is also possible that there is a sequence in the Chinese hamster genome that is similar to the region of *S. aureus*

that our primers target. The next question to answer is how to decrease the nonspecific binding of primers to and amplification of this Chinese hamster sequence.

Perhaps the next logical step to troubleshoot this issue is to run the same set of experiments using a different mammalian cell line, such as HEK293. This would help us determine whether or not the nonspecific amplification was the result of a sequence in the CHO cell line genome. Another possible step could be to vary primer concentration, since the Bio-Rad troubleshooting page suggests that using a high concentration of primers can increase the chances of nonspecific amplification. They suggest that primer concentration be in the range of 0.2-1 μM, and we are at the high end of the range with a concentration of 1 μM (Bio-Rad, "PCR Troubleshooting").

One last way to troubleshoot could be the addition of common PCR additives such as DMSO, bovine serum albumin (BSA), and/or glycerol, which have been shown to help improve the sensitivity and specificity of qPCR reactions (Wang et al, 2015; Sedgley et al, 2005). If neither primer dilution nor the addition of common PCR additives are effective, two slightly more complex directions could be using a commercially-available hot-start DNA polymerase (Green et al, 2018; ThermoFisher Scientific, "How is Hot-Start Technology Beneficial For Your PCR?") or performing a nested PCR reaction (van Pelt-Verkuil et al, 2008; Tran et al, 2014).

Once optimal specificity has been achieved and optimized for detecting *S. aureus* in samples containing CHO cells, we hope to implement our protocol with the invasion assay. We would first run a baseline experiment with no drug molecule to ensure that *S. aureus* can be detected as expected. This baseline experiment would still involve the fluorescent staining of RN4220 before the invasion, so that successful infection can be confirmed via fluorescent microscope. The next experiment would be one that mimics a successful drug without adding another layer of complexity to the protocol. It would follow the same protocol as the baseline experiment, except each condition would get a different number of *S. aureus* cells. The original protocol calls for 40x the number of CHO cells in the sample, so we may try 40x, 30x, 20x, and 10x. This experiment would allow us to determine whether or not our assay is sensitive enough to detect differences in the amount of bacteria present in different samples. The final step would be to test our protocol with a drug molecule that is already known to inhibit sortase A and prevent infection, to see if our assay can detect a difference between samples treated with

different concentrations of this drug. These results could be compared to flow cytometry results to give us an idea of how accurate our quantification is.

Though we still have a ways to go in terms of optimizing the invasion assay, we have made good progress on primer optimization and gained important insights on possible ways to improve our protocol. If successful, this optimized protocol would significantly improve the amount of time and labor required for testing potential drug candidates. This project has made small but important steps towards the development of successful anti-infective drugs, an alternative treatment method against bacterial infections that addresses the growing crisis of microbial antibiotic resistance.

Chapter 5: Engineering Standards and Realistic Constraints

5.1 Science, Technology, and Society

With decreasing effectiveness of antibiotic drugs due to the increasing resistance of bacteria, anti-infective drugs will become much more important and urgent in the coming years as they deal with drug-resistant strains of bacteria (O'Neill, 2014). Antibiotics only work to kill off bacteria inside a person who has already been infected, which allows any bacteria which survive the drugs to become a drug resistant strain of bacteria. Anti-infective drugs prevent this issue entirely by preventing bacteria from infecting a host, which also prevents the bacteria from developing a resistance to the drugs by disabling their ability to reproduce.

An issue of public concern is the decreasing effectiveness of existing drugs to combat bacteria. As part of our engineering responsibility to civic engagement, we chose to tackle this project. The SARS-CoV-2 pandemic and the increasing frequency of pandemics and epidemics in general have demonstrated the degree to which pandemics strain resources which are important to the development of drugs to combat the outbreaks. Research into BioAI will make the urgent development of drugs much faster, easier, and cheaper, potentially allowing for the mitigation of future pandemics.

5.2 Ethical Considerations

While there are numerous goods which could potentially come from this project and from the design process it follows, there is a strong potential for misuse. The ability to design a compound for interaction with a specific protein could lead to the creation of poisonous and toxic compounds which could cause severe and widespread harm if left in the wrong hands. There are several realities that must be considered in evaluating the danger of this process to limit the danger of this model of BioAI.

For the vast majority of malicious actors, this project would be entirely irrelevant. First and foremost, this model requires chemical and biological knowledge to develop drugs, which are not commonly held by amateurs. This BioAI requires a singular protein to target, so even if a group of malevolent actors got a hold of this technology, without experienced molecular biology experts, lab technicians, and expansive laboratories, they would not be able to put it to effective use. Without those resources and specialists, it would be impossible to select an effective protein target. The significant danger from this BioAI then, is exclusively large-scale organizations who

are seeking to target particular proteins and have access to huge research, testing, and manufacturing capabilities. While such organizations exist, they almost certainly have existing weaponry which makes the risk of this project causing additional harm minimal. For example, if the BioAI is capable of producing a compound which is equally harmful as the nerve agent VX, a prospect which sounds terrifying, it would actually cause little to no harm. The group using the chemical agent could have simply produced VX, which has a publicly known structure, and been no less dangerous for it. Therefore the BioAI would not increase the harm of those actors.

A large number of toxins and chemical weapons are already public knowledge just like VX, and those would be faster and cheaper to produce as the design phase would be entirely skipped. While it is possible that the BioAI could produce more toxic compounds than VX, it is unlikely that malicious actors would choose to waste so many resources for incremental gains. Thus this AI pragmatically is of little use to dangerous actors, given the scale of publicly available toxins. So while the BioAI being able to create unique chemical threats is an extant possibility, it is an unlikely case which is outweighed by the consistent good anti-infective drugs it would create in medical care. So while security of data and toxins ought to be maintained as much as possible to control the risk as much as possible, the potential for misuse is not a reason to stop research in this field.

5.3 Usability and Sustainability

The physical testing of chemicals consumes chemical reagents, as well as often safety equipment and lab time. It is far more energy and resource efficient to instead test them *in silica*. Both in terms of ease for technologists and for reducing chemical waste, AI drug development presents a clear improvement in usability and sustainability as compared to traditional drug design. Furthermore, optimization of the invasion assay used for testing anti-infective drugs *in vitro* to use qPCR instead of flow cytometry improves the assay's usability, making it significantly easier to run and decreasing the amount of time required for troubleshooting.

5.4 Health and Safety

An important thing to consider is that drug candidates will not escape the normal scrutiny of drug proposals just because they are designed by a computer. The usual multi-year process of testing and clinical trials will be enacted, and this will severely limit the potential damage which

drug candidates algorithmically generated will be able to produce. Drug candidates will undergo in vitro testing with immortalized cell lines, and then will be tested in vivo using mouse or other animal models. If the drug candidate does well in both of these tests, it can advance to the clinical trial phase where its safety and efficacy will be tested in human patients. The drugs produced by the BioAI program do not have an innately higher risk than those designed by typical human testing methods, so once the drug candidate reaches this stage, it is ethically indistinguishable from classically developed drug candidates.

Another important ethical aspect of drug development and human testing is upholding informed consent. One of the most important aims for our research is not only to create an AI program and wet lab protocol that is safe, but also transparent and easy to understand. Our advisor and entire research group take the saying "Make sure your grandma can understand!" into our research, especially when it comes to ethical considerations. If human subjects are unaware of the process that is responsible for their experimental treatment and the risks involved, there is no point in testing and the ultimate goal of our research is considered failed. Great strides in medicine have almost always been made with informed consent and by ethical standards based on the ultimate goal of improving health and quality of life. Although the process of initial design is different, the same standards and principles of drug testing such as accessibility, which apply to all drug design, can be implemented here, ensuring continued ethical practice and as little medical risk as possible.

References

Alksne, L. E., & Projan, S. J. (2000). Bacterial virulence as a target for antimicrobial chemotherapy. Current Opinion in Biotechnology, 11(6), 625–636. https://doi.org/10.1016/s0958-1669(00)00155-5

Aminov R. I. (2010). A brief history of the antibiotic era: lessons learned and challenges for the future. Frontiers in microbiology, 1, 134. https://doi.org/10.3389/fmicb.2010.00134

Arús-Pous, J., Patronov, A., Bjerrum, E.J. et al. SMILES-based deep generative scaffold decorator for de-novo drug design. J Cheminform 12, 38 (2020). https://doi.org/10.1186/s13321-020-00441-8

Bierne, H., Mazmanian, S.K., Trost, M., Pucciarelli, M.G., Liu, G., Dehoux, P., , Jänsch, L., Portillo, F.G.-d., Schneewind, O. and Cossart, P. (2002), Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. Molecular Microbiology, 43: 869-881. https://doi.org/10.1046/j.1365-2958.2002.02798.x

Bio-Rad. (2022). PCR troubleshooting. Bio-Rad. Retrieved June 8, 2022, from https://www.bio-rad.com/en-us/applications-technologies/pcr-troubleshooting?ID=LUSO3HC4S#:~:text=Primer%20concentration%20was%20too%20high,µM%20in%20the%20final%20reaction

Bio-Rad. (2022). *What is real-time PCR (qPCR)?.* Bio-Rad. Retrieved June 9, 2022, from https://www.bio-rad.com/en-us/applications-technologies/what-real-time-pcr-qpcr?ID=LUSO4W8UU

Blaschke, T., Arús-Pous, J., Chen, H., Margreitter, C., Tyrchan, C., Engkvist, O., Papadopoulos, K., & Patronov, A. (2020). Reinvent 2.0 – an AI tool for de Novo Drug Design. https://doi.org/10.26434/chemrxiv.12058026

David, L., Brata, A. M., Mogosan, C., Pop, C., Czako, Z., Muresan, L., Ismaiel, A., Dumitrașcu, D. I., Leucuta, D. C., Stanculete, M. F., Iaru, I., & Popa, S. L. (2021). Artificial Intelligence and Antibiotic Discovery. Antibiotics (Basel, Switzerland), 10(11), 1376. https://doi.org/10.3390/antibiotics10111376

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016, February 12). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics. Retrieved June 8, 2022, from https://www.sciencedirect.com/science/article/pii/S0167629616000291?via%3Dihub

Galazzo G, van Best N, Benedikter BJ, Janssen K, Bervoets L, Driessen C, Oomen M, Lucchesi M, van Eijck PH, Becker HEF, Hornef MW, Savelkoul PH, Stassen FRM, Wolffs PF and Penders J (2020) How to Count Our Microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. Front. Cell. Infect. Microbiol. 10:403. doi: 10.3389/fcimb.2020.00403

Green, R.M., Sambrook, J. (1970, January 1). Hot Start Polymerase Chain Reaction (PCR). Cold Spring Harbor Protocols. Retrieved June 8, 2022, from http://cshprotocols.cshlp.org/content/2018/5/pdb.prot095125.short

Jang, T. H., Park, S. C., Yang, J. H., Kim, J. Y., Seok, J. H., Park, U. S., Choi, C. W., Lee, S. R., & Han, J. (2017). Cryopreservation and its clinical applications. Integrative medicine research, 6(1), 12–18. https://doi.org/10.1016/j.imr.2016.12.001

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., & Korpela, K. (2020, January). *Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling*. PLOS ONE. Retrieved June 9, 2022, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0227285#references

Kim, S. O., Kim, J., Okajima, T., & Cho, N. J. (2017). Mechanical properties of paraformaldehyde-treated individual cells investigated by atomic force microscopy and scanning ion conductance microscopy. Nano convergence, 4(1), 5. https://doi.org/10.1186/s40580-017-0099-9

Kim, Y.-T., Choi, E.-H., Son, B.-K., Seo, E.-H., Lee, E.-K., Ryu, J.-K., Ha, G.-W., Kim, J.-S., Kwon, M.-R., Nam, J.-H., Kim, Y.-J., & Lee, K.-R. (2011). Effects of Storage Buffer and Temperature on the Integrity of Human DNA. Korea Science. Retrieved June 2022, from https://www.koreascience.or.kr/article/JAKO201203337761724.pdf

Kourtis AP, Hatfield K, Baggs J, et al. Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections — United States. MMWR Morb Mortal Wkly Rep 2019;68:214–219. DOI: http://dx.doi.org/10.15585/mmwr.mm6809e1

Li, Y., Zhang, L. & Liu, Z. Multi-objective de novo drug design with conditional graph generative model. J Cheminform 10, 33 (2018). https://doi.org/10.1186/s13321-018-0287-6

Lu, C., Zhu, J., Wang, Y., Umeda, A., Cowmeadow, R. B., Lai, E., Moreno, G. N., Person, M. D., & Zhang, Z. (2007). Staphylococcus aureus Sortase A exists as a dimeric protein in vitro. Biochemistry, 46(32), 9346–9354. https://doi.org/10.1021/bi700519w

Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/

McKinnon K. M. (2018). Flow Cytometry: An Overview. Current protocols in immunology, 120, 5.1.1–5.1.11. https://doi.org/10.1002/cpim.40

Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. Alzheimer's & dementia (New York, N. Y.), 3(4), 651–657. https://doi.org/10.1016/j.trci.2017.10.005

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry, 30(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., Francois, P., & Cheung, A. L. (2011, April 11). Complete genome sequence and immunoproteomic analyses of the bacterial ... Journal of Bacteriology. Retrieved June 9, 2022, from https://journals.asm.org/doi/10.1128/JB.00182-11

O'Neill, J. (Ed.). (2014). Antimicrobial resistance: Tackling a crisis for the health and wealth of nations. Antimicrobial Resistance Review. Retrieved June 9, 2022, from https://amr-review.org/sites/default/files/AMR%20Review

Proctor, R. A. (2012). Challenges for a universal staphylococcus aureus vaccine. Clinical Infectious Diseases, 54(8), 1179–1186. https://doi.org/10.1093/cid/cis033

Ruiz-Villalba, A., Ruijter, J. M., & van den Hoff, M. (2021). Use and Misuse of Cq in qPCR Data Analysis and Reporting. Life (Basel, Switzerland), 11(6), 496. https://doi.org/10.3390/life11060496

Sedgley, C. M., Nagel, A. C., Shelburne, C. E., Clewell, D. B., Appelbe, O., & Molander, A. (2004, December 15). Quantitative real-time PCR detection of oral enterococcus faecalis

in humans. Archives of Oral Biology. Retrieved June 8, 2022, from https://www.sciencedirect.com/science/article/pii/S000399690400264X

Soma Mandal, Mee'nal Moudgil, Sanat K. Mandal, Rational drug design, European Journal of Pharmacology, Volume 625, Issues 1–3, 2009, Pages 90-100, ISSN 0014-2999, https://doi.org/10.1016/j.ejphar.2009.06.065

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020). A Deep Learning Approach to Antibiotic Discovery. Cell, 180(4), 688–702.e13. https://doi.org/10.1016/j.cell.2020.01.021

Szymański, P., Markowicz, M., & Mikiciuk-Olasik, E. (2012). Adaptation of high-throughput screening in drug discovery-toxicological screening tests. International journal of molecular sciences, 13(1), 427–452. https://doi.org/10.3390/ijms13010427

Takamatsu H. (2017). Comparison of Minimal Residual Disease Detection by Multiparameter Flow Cytometry, ASO-qPCR, Droplet Digital PCR, and Deep Sequencing in Patients with Multiple Myeloma Who Underwent Autologous Stem Cell Transplantation. *Journal of clinical medicine*, *6*(10), 91. https://doi.org/10.3390/jcm6100091

Taylor TA, Unakal CG. Staphylococcus Aureus. [Updated 2022 Feb 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441868/

Thermo Fisher Scientific. (n.d.). How is hot-start technology beneficial for your PCR: Thermo Fisher Scientific - US. Thermo Fisher Scientific - US. Retrieved June 8, 2022, from https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/spotlight-articles/hot-start-technology-benefits-PCR.html

Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., & Schneewind, O. (1999). Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. Proceedings of the National Academy of Sciences of the United States of America, 96(22), 12424–12429. https://doi.org/10.1073/pnas.96.22.12424

Tran, T. M., Aghili, A., Li, S., Ongoiba, A., Kayentao, K., Doumbo, S., Traore, B., & Crompton, P. D. (2014). A nested real-time PCR assay for the quantification of Plasmodium falciparum DNA extracted from dried blood spots. Malaria journal, 13, 393. https://doi.org/10.1186/1475-2875-13-393

van Pelt-Verkuil, E., van Belkum, A., & Hays, J. P. (2008). Principles and technical aspects of PCR amplification. Retrieved June 2022, from https://books.google.com/books?id=-AMuhy2Nwb0C&pg=PA190#v=onepage&q&f=false

Verhoeven, P. O., Grattard, F., Carricajo, A., Lucht, F., Cazorla, C., Garraud, O., Pozzetto, B., & Berthelot, P. (2012). Quantification by real-time PCR assay of Staphylococcus aureus load: a useful tool for rapidly identifying persistent nasal carriers. Journal of clinical microbiology, 50(6), 2063–2065. https://doi.org/10.1128/JCM.00157-12

Wang, HY., Lu, JJ., Chang, CY. et al. Development of a high sensitivity TaqMan-based PCR assay for the specific detection of Mycobacterium tuberculosis complex in both pulmonary and extrapulmonary specimens. Sci Rep 9, 113 (2019). https://doi.org/10.1038/s41598-018-33804-1

Wood, C., Sahl, J., Maltinsky, S. et al. SaQuant: a real-time PCR assay for quantitative assessment of Staphylococcus aureus. BMC Microbiol 21, 174 (2021). https://doi.org/10.1186/s12866-021-02247-6

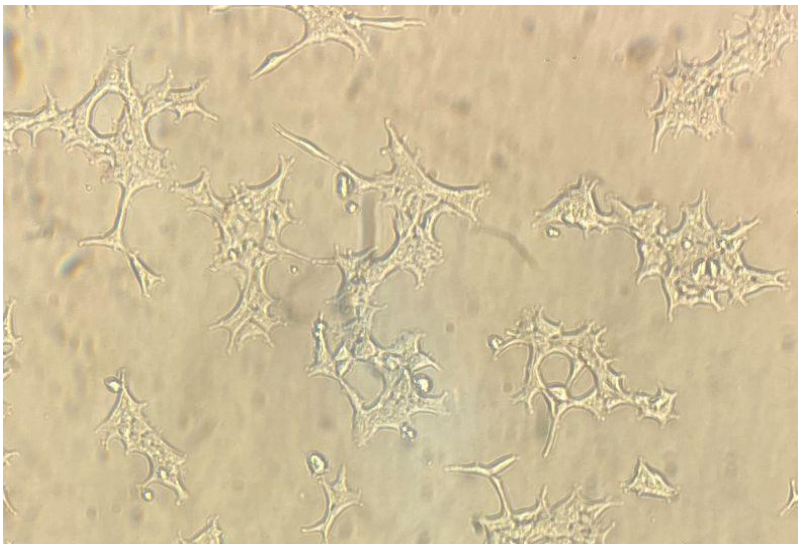
Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, & Madden TL. (2012) "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC Bioinformatics 13:134. PubMed

Yip K.W., Liu FF. (2011) Small Molecule Screens. In: Schwab M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16483-5_5376

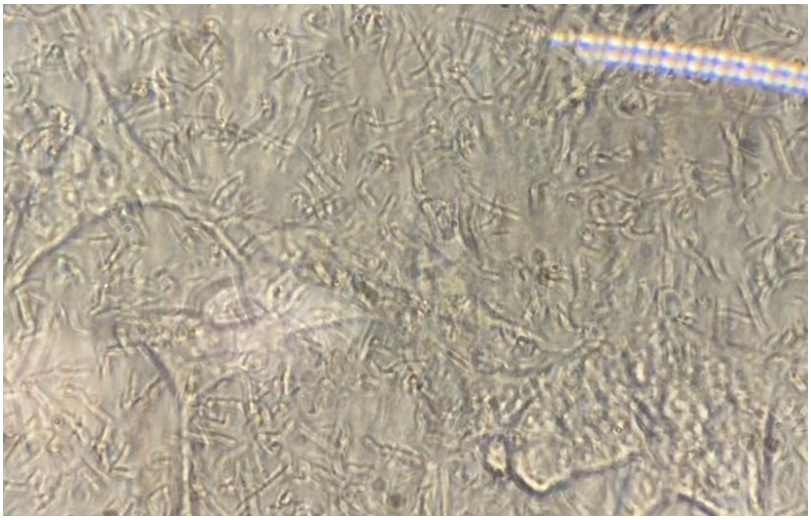
Zhu, J., Xiang, L., Jiang, F., & Zhang, Z. J. (2016). Equilibrium of sortase A dimerization on Staphylococcus aureus cell surface mediates its cell wall sorting activity. Experimental biology and medicine (Maywood, N.J.), 241(1), 90–100. https://doi.org/10.1177/1535370215592122

Appendix A: Healthy vs. contaminated CHO cell culture

(A) Healthy CHO cells



(B) CHO cells contaminated with bacteria



Appendix B: qPCR Primer Sequences

Primer	Sequence
16S rRNA (F)	CTGGAACTGAGACACGGTCC
16S rRNA (R)	TGCCACCTACGTATTACCGC
SaQuant (F)	AACTACTAGGGGAGCCTAATRAT
SqQuant (R)	GGTACTAACCAAATCAGGTCATAA
Sortase A (F)	CGTTTTTCCCAAACGCCTGT
Sortase A (R)	ACACTTTCATTGACCGTCCGA
Sortase B (F)	ACAAGCAAGAACGCGCAAAT
Sortase B (R)	ACGTCGATGTTCTCGCTCAA