Computational Tools for Developing Peptide-based Viral Inhibitors

Austin Seamann – Biomedical Informatics Graduate Student

A. Project Description

Every nation has been impacted by the viral pandemic that took shape in 2019. The SARS-CoV-2 virus has to-date been the cause of more than 5 million deaths world-wide [1]. To combat the virus, scientists around the world have developed and thoroughly tested safe and effective vaccines [2], [3]. Vaccines serve as a primary defense against severe infection from the virus. While vaccines are preventative for severe infection, they do not serve as a therapeutic that can combat on-going infection in an individual. Thus, my mentor, our collaborators, and I, developed a small peptide decoy inhibitor to block the entry mechanism of SARS-CoV-2 to human cells. The small peptide was developed based on computational analysis of the SARS-CoV-2 Spike protein and the human protein ACE-2 interaction and confirmed by cell culture. Through this project, I will expand this approach beyond the viral entry of SARS-CoV-2 to additional viruses and to develop tools that will be broadly applicable. **The goal of this project is to develop tools to rapidly analyze viral entry interactions and provide grounds for potential biologically based protein inhibitors to block viral infection.**

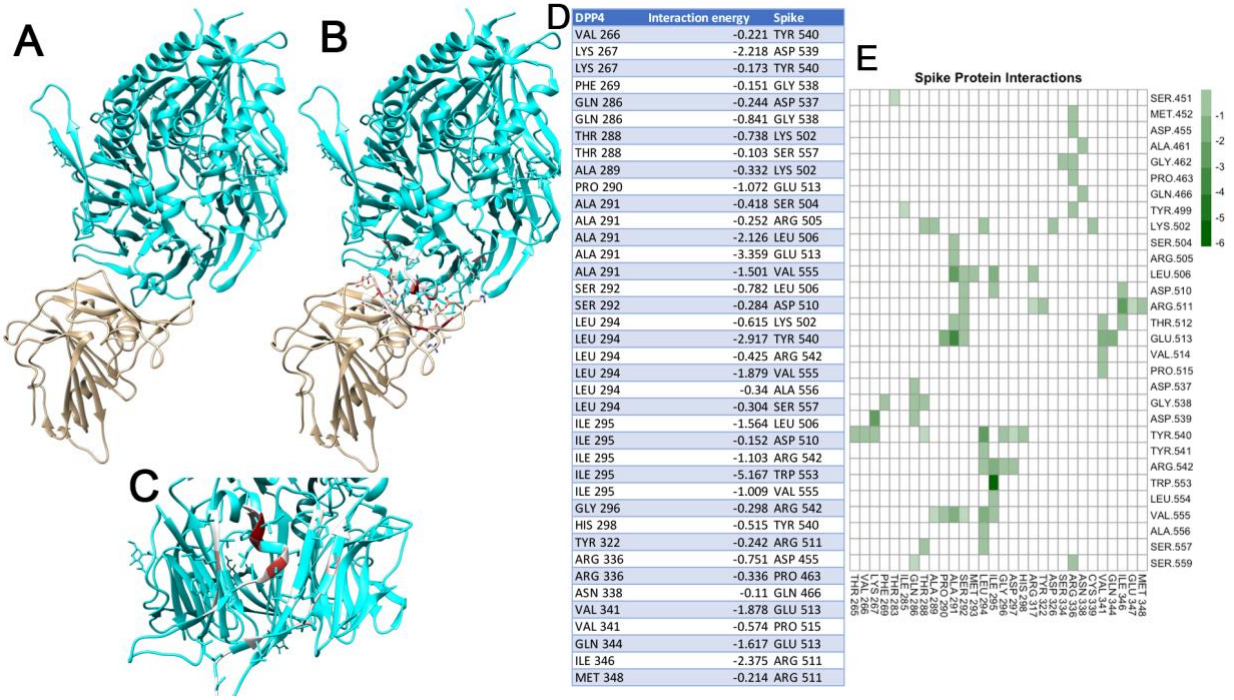
Proteins serve as the functional tools of a cell. Proteins are expressed all throughout the cell and the cell's exterior membrane. The proteins on the exterior of the cell serve several roles. These roles include communication, cell-cell recognition, energy production, defense, and cellular trafficking [4]. Viruses exploit these surface proteins for entry into the host's cells. In the case of SARS-CoV-2, the human protein angiotensin converting enzyme 2 (ACE-2) is the exploited protein. The protein also facilitates the entry of SARS-CoV and HCoV-NL63 [5]. Due to the size of proteins, traditional light microscopes cannot provide the information of the structure and composition of proteins, and so shorter wavelengths such as x-ray are needed. With the use of a technology called x-ray crystallography, we can deduce the structure of these proteins and their interactions [4]. Analyzing these structures can provide the details of the key components of the entry mechanism. My mentor's initial work has been published as a preprint and influenced the further analysis conducted to understand the SARS-CoV-2 Spike protein to ACE-2 interaction [6]. In the preprint, Dr. Ghersi and his collaborator identified key interactions between the two proteins and provided the groundwork for utilizing a piece of the ACE-2 protein that could be mass-produced to competitively inhibit SARS-CoV-2 entry into host cells.

A key component of the paper was looking at the stability of the utilized piece of the ACE-2 receptor. The proteins we are investigating are most stable in their final structure. However, when broken down, proteins will be degraded rapidly in the cellular environment. Proteins are scaffold by substructures called secondary structures. If these substructures are retained as part of the piece selected, the ability for it to survive in the environment is increased. Thus, to produce a potential inhibitor based on the exploited human protein, we want a short fragment (easily produced) that is stable and has the strongest interactions with the viral entry protein.

I determined that combining two components of the human protein would increase the number of strong interactions while maintaining a short fragment. Upon further research, another lab from the University of Michigan implemented a similar method with alternative interactions between the proteins, unique from ours [7]. Our fragment was validated to inhibit the Spike protein and prevent Spike/ACE-2 interaction by our collaborators from UNMC.

This project would be to expand from just looking at the SARS-CoV-2 interaction to other viruses. Following the same process, I'd investigate other virus entry mechanisms and identify key components to the interactions and ways to potentially automate the process of fragment selection. This would serve as an important project due to the increasing potential of future epidemics and viral spillover [8], [9]. A potential candidate would be the epidemic virus Middle East respiratory syndrome coronavirus (MERS-CoV). In figure 1, I show how my current tool can display the key interactions occurring at the virus/human interaction. While this showcases the current tools I've already developed, I will be developing new tools that will increase the efficiency of the entire protein development process and its

ability to be applied to most viral interactions. The first new tool will be able to collect as many protein structures as possible of the available viral entry structures. The second tool will then analyze the output of the current tools to suggest components to create the new protein inhibitors. Additional details can be found below in the Research Activities and the Roles of the Student and Facility Member sections.



DPP4	Interaction energy	Spike
VAL 266	-0.221	TYR 540
LYS 267	-2.218	ASP 539
LYS 267	-0.173	TYR 540
PHE 269	-0.151	GLY 538
GLN 286	-0.244	ASP 537
GLN 286	-0.841	GLY 538
THR 288	-0.738	LYS 502
THR 288	-0.103	SER 557
ALA 289	-0.332	LYS 502
PRO 290	-1.072	GLU 513
ALA 291	-0.418	SER 504
ALA 291	-0.252	ARG 505
ALA 291	-2.126	LEU 506
ALA 291	-3.359	GLU 513
ALA 291	-1.501	VAL 555
SER 292	-0.782	LEU 506
SER 292	-0.284	ASP 510
LEU 294	-0.615	LYS 502
LEU 294	-2.917	TYR 540
LEU 294	-0.425	ARG 542
LEU 294	-1.879	VAL 555
LEU 294	-0.34	ALA 556
LEU 294	-0.304	SER 557
ILE 295	-1.564	LEU 506
ILE 295	-0.152	ASP 510
ILE 295	-1.103	ARG 542
ILE 295	-5.167	TRP 553
ILE 295	-1.009	VAL 555
GLY 296	-0.298	ARG 542
HIS 298	-0.515	TYR 540
TYR 322	-0.242	ARG 511
ARG 336	-0.751	ASP 455
ARG 336	-0.336	PRO 463
ASN 338	-0.11	GLN 466
VAL 341	-1.878	GLU 513
VAL 341	-0.574	PRO 515
GLN 344	-1.617	GLU 513
ILE 346	-2.375	ARG 511
MET 348	-0.214	ARG 511

Figure 1: Showcase of current tools: Above is the analysis conducted utilizing the tools I already developed and applied to the viral entry interaction of MERS-CoV to human protein DPP4. (A) The full crystal structure of MERS-CoV Spike Protein (brown) and human protein DPP4 (blue). (B) Visualization of the interacting components of the Spike protein and DPP4 with weak interactions in white and strong interactions in red. (C) A closer look at the strong interactions of human protein DPP4. (D) A table breakdown of the interactions occurring at the Spike/DPP4 interface (more negative = stronger interactions). (E) A heatmap breakdown of the Spike/DPP4 interface.

B. Research Activities

(1) The first step of this project would be to collect structures of viral entry proteins bound to human surface proteins. These can be collected from the RCSB Protein Data Bank utilizing a sequence search of the viral entry protein sequence and collecting files with more than two protein structures that are unique [10]. A tool will be developed to automate this process. (2) Further curation would be needed to ensure the interactions are with human surface proteins. Specific factors can be checked for through the tool being developed to ensure proper structures are selected (3) Then, utilizing the Rosetta suite of programs, analyze the interactions for determining strongest binders [11]. This provides us with the information displayed in Figure 1. (4) To determine what strong interactions are a component of a protein secondary structure, I will utilize the program DSSP in an automated fashion [12], [13]. Components that are within different secondary structures will then be measured for distance to see if any could be attached together. (5) Manual selection of fragments would then be done based on the two analyses. These fragments will then be (6a) modeled and (6b) computationally docked to the viral protein. Modeling provides us with the theoretical structure that the novel protein takes form. Computational docking attempts to simulate the interaction between the viral entry protein and the designed protein inhibitor. (7) If the binding interactions are maintained, further analysis of the interaction utilizing molecular dynamic techniques will be conducted. Molecular dynamic simulations allow us to computationally predict the stability of small peptides. Thus, molecular dynamics can determine which peptides may survive better in

the environment and be better candidates. While it is outside the scope of this project, these selected fragments could then be produced and tested in cell culture through our collaborators at UNMC.

C. Project Timeline

The first month of this project will be structure collection and curation and the development of the first new tool (1, 2). While this is trivial for the interaction between SARS-CoV-2 and ACE-2 due to the massive global impact, it may be less-trivial for less studied pathogens. The second month of the project will be analyzing the virus-human entry interactions between the collected structures and determining good targets (3-5). The third month of the project will be modeling and docking the designed fragments to the viral entry proteins with further analysis (6-7). The final month will be documenting the project and preparing a final report and presentation. These are all minimum thresholds. However, if time allows additional automation of the process will also be a target.

D. Roles of Student and Faculty Member

I will be responsible for conducting all the listed research activities. Some previous work will help with the analysis steps of the project, but several new tools will need to be implemented in a fashion to be easily replicated. A tool for searching for viral entry protein interactions will need to be developed. From our experience with working on SARS-CoV-2, new structures were constantly being submitted, and this would allow us to periodically look for new structures. The second program to be built would be to interpret interaction energies from the output of Residue Energy Breakdown, a part of the Rosetta Suite. In this program, we'd look for portions of the human protein that have strong interactions and are proximal to other groups of interactions. Then depending on time still available, we'll investigate potential peptides that can be produced. Dr. Ghersi will be responsible for overseeing the project and providing guidance. We will meet weekly to discuss project progress.

E. List of Previous Funding

I was a recipient of a FUSE award in 2019 for my project, "Why does age matter? A structural analysis of T-cell receptors in young and elderly individuals". Dr. Ghersi also served as my mentor for this project. I presented the project in the 2021 student research and creativity activity fair and received the Outstanding Undergraduate Oral Presentation award.

The previous project investigated the interactions of adaptive immune system cells, T-cells, and their binding partners. This project has been expanded and is currently close to a publication. The tools being published can help guide other scientist with T-cell interaction research without putting in the large number of hours to adapting available tools. We hope these tools helps make T-cell interaction research more approachable and easier to achieve.

While some of the same tools may be adapted in this project, the types of interactions are distinctly different and with different goals. T-cell interactions that were being investigated have a constant structure appearance and can be complex to interpret. While this project still has complex protein to protein interactions, the interactions will be more distinctly different based on each virus being investigated.

F. Budget Justification

I am requesting $5,000 for the student stipend. For an estimate, at around 15-20 hours a week for 16 weeks, the $5,000 would put me around $18 an hour. This will allow me to have better flexibility in my schedule to expend adequate time towards the project. The project will leverage open-source software such as Python, UCSF Chimera, and Rosetta. No licensing will be needed for these programs, so they will not contribute to any additional budgeting constraints.

References:

[1] "WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data." [Online]. Available: https://covid19.who.int/. [Accessed: 20-Dec-2021].

[2] L. R. Baden *et al.*, "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine," *N. Engl. J. Med.*, vol. 384, no. 5, 2021, doi: 10.1056/nejmoa2035389.

[3] F. P. Polack *et al.*, "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine," *N. Engl. J. Med.*, vol. 383, no. 27, 2020, doi: 10.1056/nejmoa2034577.

[4] A. Kessel and N. Ben-Tal, *Introduction to Proteins: Structure, Function, and Motion*, 1st ed. CRC Press, 2011.

[5] F. Scialo *et al.*, "ACE2: The Major Cell Entry Receptor for SARS-CoV-2," *Lung*, vol. 198, no. 6, p. 1, Dec. 2020, doi: 10.1007/S00408-020-00408-4.

[6] F. Renzi and D. Ghersi, "ACE2 fragment as a decoy for novel SARS-Cov-2 virus," pp. 2–7, 2020, doi: 10.1101/2020.04.06.028647.

[7] X. Huang, R. Pearce, and Y. Zhang, "De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2," *Aging (Albany. NY).*, vol. 12, no. 12, 2020, doi: 10.18632/aging.103416.

[8] Z. L. Grange *et al.*, "Ranking the risk of animal-to-human spillover for newly discovered viruses," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 15, 2021, doi: 10.1073/pnas.2002324118.

[9] R. K. Plowright *et al.*, "Ecological dynamics of emerging bat virus spillover," *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1798. 2014, doi: 10.1098/rspb.2014.2124.

[10] H. Berman, K. Henrick, and H. Nakamura, "Announcing the worldwide Protein Data Bank," *Nature Structural Biology*. 2003, doi: 10.1038/nsb1203-980.

[11] R. F. Alford *et al.*, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *J. Chem. Theory Comput.*, vol. 13, no. 6, 2017, doi: 10.1021/acs.jctc.7b00125.

[12] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, 1983, doi: 10.1002/bip.360221211.

[13] R. P. Joosten *et al.*, "A series of PDB related databases for everyday needs," *Nucleic Acids Res.*, vol. 39, no. SUPPL. 1, 2011, doi: 10.1093/nar/gkq1105.

University of Nebraska Omaha

Jan 21st, 2022

To the GRACA Committee:

It is my great pleasure to write in support of Austin Seamann's GRACA application entitled "Viral Entry Analysis and Developing Potential Biological Based Protein Decoy Inhibitors". I met Austin in the fall of 2019, when he expressed an interest in volunteering in my lab and work on bioinformatics projects. During that time, Austin became familiar with scripting, structural bioinformatics, and several of the tools we routinely use to work on computational biology problems. After this initial exploration, he became very interested in studying T cell recognition, one of the key steps of an immune response. In the spring of 2020, Austin wrote a FUSE proposal whose goal was to study the properties of T cell receptors in young and elderly populations by using structural bioinformatics approaches, including homology modeling and protein docking. While working on this project, Austin had the opportunity to use data provided by my collaborator (Dr. Selin) at the University of Massachusetts Medical School. Dr. Selin was excited about this work and incorporated some of our results in a manuscript. A separate manuscript describing the tool Austin developed for this project and its applications is about to be submitted to a bioinformatics journal. **Austin's presentation for this work received the Outstanding Undergraduate Presentation award at the UNO Research Fair**.

In addition to working on immunological problems, Austin has taken a lead role in another project I started, which focuses on developing a peptide inhibitor for Sars-CoV-2. Austin was able to successfully apply his training in structural bioinformatics to tackle this important problem. **Austin's work has led to the computational design of an inhibitor which has been experimentally validated *in vitro* by our colleagues at UNMC**.

The goal of Austin's current proposal is to take this work to the next level, by developing a computational pipeline that can help us design such inhibitors for other viral infections. As a mentor, I will track Austin's progress with weekly meetings and offer advice about the computational methods and analyses that he is going to employ to tackle this project. **The summer stipend offered by the GRACA grant would enable Austin to dedicate time to this exciting project.**

Austin has all the necessary programming and analytical skills to succeed in this project. He is highly motivated, an independent thinker, and possesses a solid quantitative and computational foundation. His work ethics is also impeccable. **Given his computational skills and his passion for addressing important questions, I have no doubt that Austin will excel in this project**.

Please do not hesitate to get in touch if you need additional information.

Sincerely,

--Dario Ghersi, MD, PhD
Associate Professor of Biomedical Informatics

College of IS&T, PKI 173B, 1110 South 67th Street I Omaha, NE 68182
(402) 554.2097 I http://faculty.ist.unomaha.edu/dghersi