Characterization of a novel protease in *Drosera Capensis*

Background

Enzymes play a role in nearly all biological processes. There are many different types of enzymes, each with their own unique function. A protease is a type of enzyme that helps break down proteins. Proteases have many different uses in research, including a wide variety of applications in the medical field. Due to their high specificity, proteases can be used for targeted degradation of protein aggregates, which if not broken down may lead to diseases (1). Another application of proteases is that they are ingredients in a variety of laundry detergents. Enzymes have been thoroughly studied in a variety of different organisms. However, enzymes have been less studied in carnivorous plants than in, for example, mammalian systems. Diverse enzymes have evolved to function under a number of different conditions. The more diverse enzymes that are characterized, the more applicable they become in various different industries. Carnivorous plants contain a variety of enzymes that enable them to break down proteins in insect tissue. *Drosera capensis* is a carnivorous plant that contains many proteins that have not yet been identified. This makes the carnivorous plant an amazing organism to study with the hope of identifying novel proteases that can be useful in the medical field.

Figure 1 : Novel aspartic protease identified in transcriptome

A *D. capensis* draft genome was assembled and annotated by the Martin and Butts labs in 2016, which helped identify proteins that may be expressed by the plant. This led to the discovery of numerous novel proteins within the plant (2). There are limitations on genome annotation, which is why looking at the transcriptome is also important. The genome consists of both introns and exons, however, the transcriptome is solely exons. Annotating the genome aids in identifying novel proteins, however, the transcriptome provides a better insight as to what is getting expressed. This is because all of the introns are spliced out, leaving only the coding regions. The transcript codes

for a protein that will be expressed. Following this, the transcriptome was assembled in order to catch anything that was missed by the original genome annotation. Using the transcriptome of *D. capensis,* we have been able to identify a novel protein in this plant that was not identified in the genome. The goal of this research is to characterize this identified protein. We want to verify if this protein contains any protease activity. This will be done by expressing the protein in the lab, and testing the protein for protease activity.

Approach

The first step in verifying if this protein has protease activity is to design a plasmid. My lab and I will need to decide what parts of the protease to include in the plasmid. For example some proteins contain propeptides that need to be cleaved off prior to activation. In order to determine if the sequence needs to be cleaved off, we need to compare our novel protein sequence to a sequence of a known aspartic protease in *D. capensis.* After the plasmid is designed, it will be ordered from Twist Bioscience and transformed into *E. coli.* Once transformed, the protein will need to be expressed and purified. A classic way to do this is via a His-tag. Histidine binds well with a nickel column, so adding a His-tag to our protein during plasmid design will allow our protein to bind to the column, while all of the other proteins wash away. There are many other techniques that can be used if the nickel column fails, such as size exclusion chromatography (separation due to size of protein). Following this, we have to verify that it is our novel protease. There are a variety of methods to verify the identity of a pure protein, including western blots with a His-tag antibody, and mass spectrometry. We can then test protease activity by introducing peptides to the protein and using mass spectrometry in order to determine if the peptides are cleaved. If peptides are cleaved, we will have shown that this novel protein is indeed a functional protease.

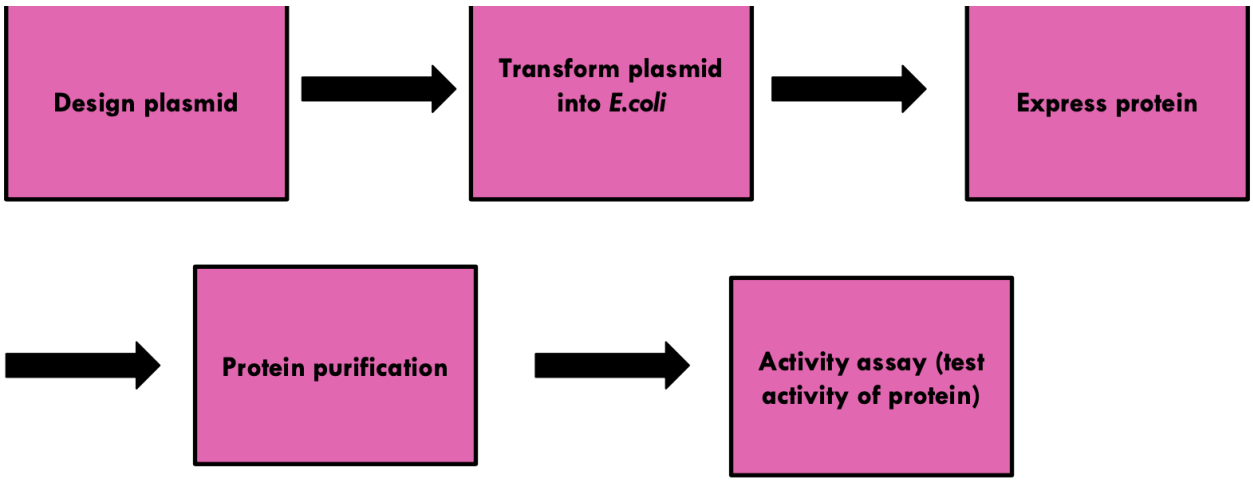

Figure 3: Outline of my research approach.

Objective

Short term goal 1: Successfully transform plasmid into *E. coli.* This plasmid will contain an antibiotic resistance gene, and if this transformation is successful, *E. coli* will grow.

Short term goal 2: Successfully express and purify protein. Using techniques such as mass spectrometry will help us determine if we were successful in purifying this protein.

Short term goal 3: Demonstrate protease activity in a novel putative protease identified by Omar Akbari.

Long term goal: Characterize the function of novel enzymes in *D. capensis.*

Responsibilities

- Coming to lab meetings once a week to discuss any updates I may have, and to ask any questions that come up.
- Communicate with lab members, including [Redacted] and grad student mentor(s). I am constantly communicating with both [Redacted] and grad students.
- Take care of the plants. Watering, fertilizing, and pruning the plants when necessary.
- Perform the hands on lab work as well as data analysis in order to successfully demonstrate protease activity

Timeline

May 2022: Design and order the plasmid that will be used in *E. coli* transformation.
June 2022: Transform the plasmid into *E. coli* and create glycerol stocks of transformed cells.
July 2022: Express and purify protein. Refine expression protocols, purification techniques, and plasmid design as necessary in order to maximize yield and purity.
August 2022: Perform activity assays on purified protein to verify protease activity

Itemized budget

Item	**Why is it necessary?**	**Cost**
Plasmid from Twist Bioscience	Plasmid used to express putative novel aspartic protease	~$200
IPTG (1g) (Catalog #: 15529019)	Used to induce expression in transformed *E. coli*	$76.75
HPLC peptide standard mixture (Catalog #: H2016-1VL)	Peptides used for activity assays	$33.50
New England Biolabs, Inc. LongAmp Taq 2X Master Mix (Catalog #: M0287S)	DNA polymerase mix required for PCR - necessary for library preparation	$110.40

LSK109 (Catalog #: SQK-LSK109)	Ligation sequencing kit used to add adapter sequences to ends of DNA strands - necessary for library preparation	$599
PCR barcoding expansion 1-12 (Catalog #: EXP-PBC001)	Barcoding kit used to add barcodes to DNA sequences for multiplexing - necessary for library preparation	$288
Total		$2,713.40

References

1. Butts, Carter T., Jan C. Bierma, and Rachel W. Martin. “Novel Proteases from the Genome of the Carnivorous Plant *Drosera Capensis* : Structural Prediction and Comparative Analysis.” *Proteins: Structure, Function, and Bioinformatics* 84, no. 10 (October 2016): 1517–33. https://doi.org/10.1002/prot.25095.
2. Butts, Carter T., Xuhong Zhang, John E. Kelly, Kyle W. Roskamp, Megha H. Unhelkar, J. Alfredo Freites, Seemal Tahir, and Rachel W. Martin. “Sequence Comparison, Molecular Modeling, and Network Analysis Predict Structural Diversity in Cysteine Proteases from the Cape Sundew, Drosera Capensis.” *Computational and Structural Biotechnology Journal* 14 (2016): 271–82. https://doi.org/10.1016/j.csbj.2016.05.003.