[Methods to Determine Protein Structure: Traditional vs Machine Learning]
Total Number of Lessons: [3 lessons]
Total Estimated Time for Module: [number of minutes]

SECTION 1: MODULE OVERVIEW AND CURRICULAR CONNECTIONS – REQUIRED

- **Module goal**: [At the end of this module, SWBAT…; try to make it one big goal]
  To help students better understand protein secondary and tertiary structure and their relationship to protein function, students will compare traditional x-ray crystallography methods and model predicting/identifying segments of protein structures using a simple machine learning approach.

- **Connection to standards**: [List standards that will be addressed by the module. Should include Indiana standards; NGSS and CCSS are strongly recommended; CSTA could also be useful for the Computing RET.]

- **RET materials/ideas to be leveraged**: [Describe what elements of your summer experience you will use as part of the module for students.]
  Concepts related to machine learning and deep neural networks; teachable machine as an example/useable model for students to gain a conceptual understanding of how these networks can be used.

SECTION 2: OVERVIEW OF MODULE FRAMEWORK – REQUIRED

- **Real-world context**: [Describe in 1-2 paragraphs]

  - **Possible Lesson Ideas**: [What are some ideas for a lesson or two focused on the real-world context and framing the project for students?]
    The module will start off with a reading about drug discovery. In the reading, the students will learn about how current advances in computer science and molecular modeling are providing key information about protein structural models that can be used to predict drug structures specific to a protein. They will also be introduced to the idea of x-ray crystallography as the most prominent method of protein structure determination and its current limitations. This will then set up for the concept of using machine learning as new method to get around some of these limitations. This will help ground the module in why it is important to protein structures.

- **Background STEM content**: [Describe in 1-2 paragraphs]
  Prior to this module, students should be well versed in the central dogma of molecular biology including all terminology, chemistry, and process surrounding protein synthesis. They should be able to differentiate between and describe the functionality of

UNIVERSITY OF NOTRE DAME

Center *for* STEM Education

107 Carole Sandner Hall, Notre Dame, IN 46556   http://STEMeducation.nd.edu

communities, populations, individuals, organ systems, tissues, cells, and molecules. Students should know:

1. Monomer/polymer reactions in DNA replication, transcription, and translation
2. Monomer/polymer structures of DNA, RNA, and proteins
3. Chemical properties/identities of the 20 amino acids
4. Relationship between structure and function and the process of denaturation of proteins
5. Basics of molecular cloning and protein purification
   o **Possible Lesson Ideas**: [What are some ideas for a lesson or two focused on the background STEM content?]

- **Description of Main Activity/Project**: [Describe in 1-3 paragraphs]

  o **Possible Activity/Project Ideas**: [What is the goal? What will students do? What is the timeline? At what point would students work in groups or individually?]

  o This module is split into two projects that students will work on side-by-side. In the first, they will use purified albumin protein to produce a protein crystal in order to understand the crystallization process. They will then use The Interactive Structure Factor Tutorial (http://www.ysbl.york.ac.uk/~cowtan/sfapplet/sfintro.html) to examine how x-ray crystallography data can be used to determine a 2-D protein structure.

  o While their crystal is forming, students will explore the different levels of protein structure with hands on manipulative to relate amino acid structures and chemical properties (hydrophobic, hydrophilic, positive and negative charges) with secondary structure (alpha helix and beta sheet) formation. In addition, students will predict possible effects on these structures when the amino acids have been altered. Students then apply what they learned to a real-world situation: sickle cell anemia and how a single amino acid change alters the structure and function of the hemoglobin protein.

  o Once students have a firm background on the basics of secondary structure, they will be introduced to the concepts (not programming) of machine learning through a presentation and then using Google's teachable machine (https://teachablemachine.withgoogle.com/). First, they will explore training the machine using simple images of their choosing with two different categories. Then, there will be there will be model building activities related to protein structure. First, they will train the model to identify alpha helices vs beta sheets using downloaded and cropped x-ray crystal structures to see if they can train the model to pick them out in full structures. If that works, the students would then download and crop proteins with similar functional domains (such as a nucleotide binding domain) to see if they can train the model to pick out when other proteins

have similar domains with the underlying understanding that similar structure should result in similar function.

SECTION 3: MODULE SEQUENCING AND ASSESSMENT – REQUIRED

- **Brief description of sequenced learning objectives**: [How do learning objectives build from lesson to lesson to meet the module goal?]
Prior to beginning the module, students will gain knowledge of the structure and function of the molecules involved in gene expression and how proteins are formed. The beginning of this project will reintroduce the idea of structure-function relationships and the difficulties and relevancies of determining protein structures and the possible ways that machine learning and computer science can help solve some of these problems. Then students investigate and develop machine learning protocols for predicting/determining protein structures.

- **Brief description of formative and summative assessment approaches**: [When/what techniques will you use to assess student progress towards objectives during and at the end of the module?]

  Students will be assessed with formative quizzes on protein structure, amino acid types, and other basic content.
  Student groups will produce a narrated PowerPoint discussing:
  o Why protein structures are important.
  o How protein structures are currently determined and the limitations.
  o How machine learning works
  o Their thought process in developing a training program for all three models.
  o How they evaluated all three trained models
  o Their analysis their models' ability to recognize the objects presented

- **Recommendations for implementation**: [Describe any "safety tips" or advice you have for other educators who might implement this module, including logistical considerations.]

SECTION 4: MODULE FILES – OPTIONAL

- **List of attached files**: [should include any lesson plans and handouts relevant to the Module, if you have them ready]

  o [File name 1 and one line description]

  o [File name 2 and one line description]

UNIVERSITY OF NOTRE DAME

Center *for* STEM Education

107 Carole Sandner Hall, Notre Dame, IN 46556   http://STEMeducation.nd.edu

- o [File name 3 and one line description]

- o [File name 4 and one line description]

- o [File name 5 and one line description]