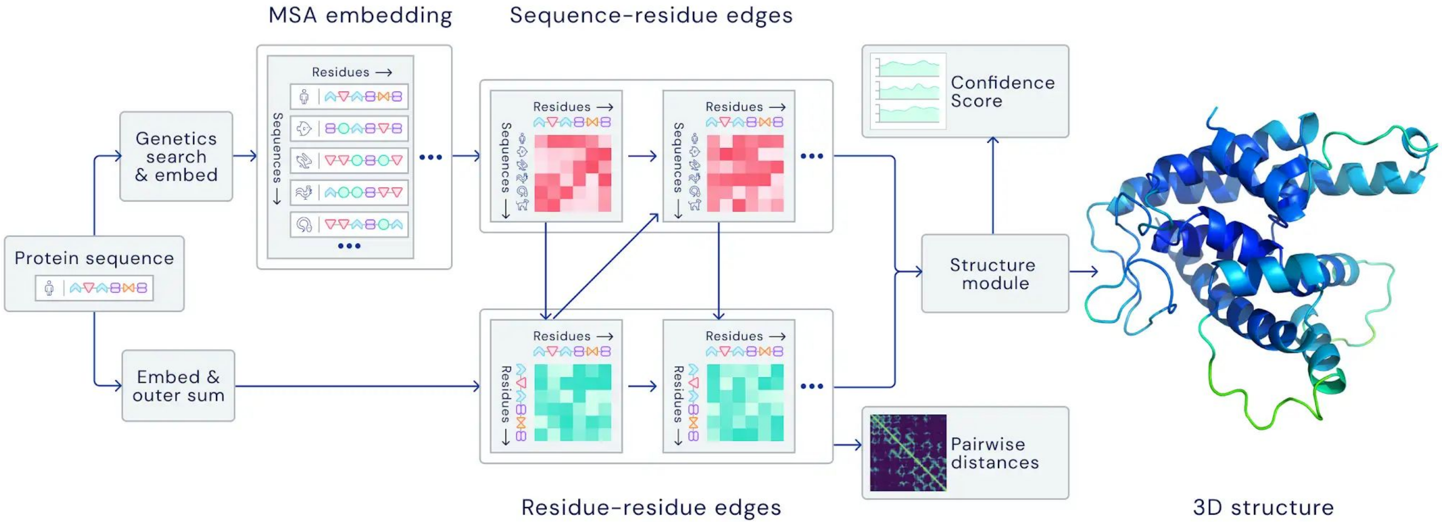
Why “Bio-” in data structures?

Young-suk Lee

Young Laboratory of Quantitative and Computational Biology

The advancement in artificial intelligence

AlphaFold: Near well-defined system



Spectrum of bioinformatics research

Software engineering

Mathematical modeling

Contributions to biology

Software engineering

Mathematical modeling

- Sequence alignment
- Genome assembly
- Phylogenetic tree reconstruction
- Big biological data
- Protein folding

Modeling biological systems

Software engineering

Mathematical modeling

- Sequence alignment
- Genome assembly
- Phylogenetic tree reconstruction
- Big biological data
- Protein folding

- Sequence analysis
- Differential gene expression
- Network analysis
- System biology
- Next strain prediction

Focus on the basics before branching out

Software engineering

Mathematical modeling

- **Sequence alignment**
- Genome assembly
- Phylogenetic tree reconstruction
- Big biological data
- Protein folding

- **Sequence analysis**
- Differential gene expression
- Network analysis
- System biology
- Next strain prediction

Our approach for this course

1. Start from a specific biological question

Our approach for this course

1. Start from a specific biological question
2. Reduce it into a concrete computational problem

Our approach for this course

1. Start from a specific biological question
2. Reduce it into a concrete computational problem
3. Design algorithms that solve the problem
4. Build data structures that improve time and/or space complexity

Our approach for this course

1. Start from a specific biological question
2. Reduce it into a concrete computational problem
3. Design algorithms that solve the problem
4. Build data structures that improve time and/or space complexity
5. Interpret the solution in the context of the original biological question

Textbook (mostly Ch1 and Ch5)

BIOINFORMATICS ALGORITHMS
An Active Learning Approach
3rd Edition
by Phillip Compeau & Pavel Pevzner

What you will learn

- Basic (python) programming for bioengineering
- Some discrete mathematics
- Nontrivial algorithms
- Clean code
- Algorithmic thinking
- Some background in (molecular) biology
- A little time and space complexity
- Dynamic programming
- Heuristics and big biological data

Course logistics

Office hour: MW 3:30 pm - 4:30 pm (Rm#1113, E16)

Recommended Prerequisites:

- CS101 Introduction to Programming
- BS120 General Biology

Grading: Programming assignments (30%), problem sets (10%), midterm exam (25%), final project (25%), and participation (10%).

More about grading

- Start early
- Visit office hours
- Do not consult the internet
- Do not use third-party modules or libraries
- Include comments when necessary
- Check if your code is readable and clean
- Test your implementation on the class computing server
- Exercises are designed to help you better understand the material and will not be graded.

Late policy

- You have three late days. Notify the TA before the deadline and use them wisely.
- Note cannot use late day for final project and midterm.
- 24 hr late submission results in 50% and will not accept assignments submitted after that.
- Again start early!

Special tutorials (tentative schedule)

- Week 2: ssh, unix, vscode, and remote servers
- Week 3: python programming #1 (random variables, control flow, function)
- Week 4: python programming #2 (list, objects, modules, file handling)
- Week 9: python programming #3 (Nested arrays, custom classes, and more)

Academic honor code for BiS232

- All assignments must be completed individually. There won't be any collaborative assignments.
- Looking at someone else's solution is an infraction of the Honor Code.
- You should not leave copies of your work on public computers nor post your solution on a public website.
- If you received aid while producing your solution, you should indicate from whom you got help (if that person is not a TA, or instructor for this class) and what help you received.
- All submissions are subject to automated plagiarism detection.
- The full document is available on KLMS.