



DEEMED TO BE UNIVERSITY

22BIO211: Intelligence of Biological Systems - 2

CYCLOPEPTIDE SCORING PROBLEM

Dr. Manjusha Nair M
Amrita School of Computing, Amritapuri
Email : manjushanair@am.amrita.edu
Contact No: 9447745519

Is Cyclopeptide Sequencing Successful always?

- Cyclopeptide Sequencing successfully reconstructed Tyrocidine B₁
 - *But this algorithm only works in the case of an ideal spectrum, i.e., when the experimental spectrum of a peptide coincides exactly with its theoretical spectrum.*
- Mass spectrometers generate "noisy" spectra that are far from ideal – produce both **false masses** and **missing masses**.
 - *A false mass is present in the experimental spectrum but absent from the theoretical spectrum;*
 - *A missing mass is present in the theoretical spectrum but absent from the experimental spectrum.*

Is Cyclopeptide Sequencing Successful always?

- Theoretical and (simulated) experimental spectra of the cyclic peptide NQEIT
 - Theoretical: 0 113 114 128 **129** 227 **242** 257 355 356 370 371 484
 - Experimental: 0 **99** 113 114 128 227 257 **299** 355 356 370 371 484
- missing masses from the experimental spectrum are shown in blue,
- false masses in the experimental spectrum are shown in green.
- mass of the amino acid E (129) is missing, and the mass of the amino acid V (99) is false
- As a result, the first step of Cyclopeptide Sequencing would establish {V, L, N, Q} as the amino acid composition of our candidate peptides, which is incorrect.

Adapting cyclopeptide sequencing for spectra with errors

- Any false or missing mass will cause Cyclopeptide Sequencing to throw out the correct peptide
 - *because its theoretical spectrum differs from the experimental spectrum.*
- How would you reformulate the Cyclopeptide Sequencing Problem to handle experimental spectra with errors?

Adapting cyclopeptide sequencing for spectra with errors

- Cyclopeptide Sequencing Problem to handle noisy spectra
 - *relax the requirement that a candidate peptide's theoretical spectrum must match the experimental spectrum exactly*
 - instead incorporate a **scoring function** that will select the peptide whose theoretical spectrum matches the given experimental spectrum **the most closely**.

Cyclopeptide Scoring Problem

- Given a cyclic peptide **Peptide** and a spectrum **Spectrum**, we define **Score(Peptide, Spectrum)** as the number of masses shared between Cyclospectrum(Peptide) and Spectrum.
- **Cyclopeptide Scoring Problem:** Compute the score of a cyclic peptide against a spectrum.
Input: An amino acid string Peptide and a collection of integers Spectrum.
Output: The score of Peptide against Spectrum,
Score(Peptide, Spectrum).

Cyclopeptide Scoring Problem

Theoretical:	0	113	114	128	129	227	242	242	257	355	356	370	371	484
Experimental:	0	99	113	114	128	227		257	299	355	356	370	371	484

- If *Spectrum* is the experimental spectrum in the bottom row of this figure, then $\text{Score}(\text{NQEL}, \text{Spectrum}) = 11$ (the number of columns shared by the theoretical and experimental spectra, shown in black).
- The scoring function should take into account the multiplicities of shared masses
 - *For Eg., If 242 has multiplicity 2 in the experimental spectrum of Peptide, then 242 contributes 2 to $\text{Score}(\text{Peptide}, \text{Spectrum})$*

Cyclopeptide Sequencing Problem Redefined

- We can redefine the Cyclopeptide Sequencing Problem for noisy spectra.
 - Our goal is to adapt the Cyclopeptide Sequencing algorithm to find a peptide with maximum score.
 - **Cyclopeptide Sequencing Problem (for spectra with errors):** Find a cyclic peptide having maximum score against an experimental spectrum.
- Input:** A collection of integers Spectrum.*
- Output:** A cyclic peptide Peptide maximizing Score(Peptide, Spectrum) over all peptides Peptide with mass equal to ParentMass(Spectrum).*

Cyclopeptide Sequencing Problem Redefined

- In our Branch and Bound algorithm, all candidate linear peptides having inconsistent spectra were thrown out.
- We need to revise this step to include more candidate linear peptides, while still ensuring that the number of peptides that we consider does not grow out of control.
 - *For example, we saw that the linear peptide VKF is inconsistent with the theoretical spectrum of the cyclic peptide Tyrocidine B1.*
 - *In the new approach, we should not ban VKF in the case of experimental spectra, since they can have missing masses.*

Cyclopeptide Sequencing Problem Redefined

How can we limit the growth of the list of candidate linear peptides in the case of experimental spectra?

Summary

- Is Cyclopeptide Sequencing Successful always?
- Adapting cyclopeptide sequencing for spectra with errors.
- Cyclopeptide Scoring Problem
- Cyclopeptide Sequencing Problem Redefined