

## 22BIO211: Intelligence of Biological Systems - 2

# CYCLOPEPTIDE SEQUENCING

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

# Cyclopeptide Sequencing Problem

- We can't always use DNA to resolve peptide/protein sequences
- Solution approach
  - Extract and purify a pure sample of the peptide/protein
  - Randomly fracture the peptide
  - Assemble an answer from the pieces
- A Mass Spectrograph can precisely measure the molecular weight (and charge and abundance) of any peptide chain
- Since the molecular weight of each of the possible 20 amino acids is known precisely, which combination of them would give a particular weight?

# Cyclopeptide Sequencing Problem

- How can we use the experimental spectrum to sequence a peptide?
  - Generating the theoretical spectrum of a known peptide is easy
  - But our aim is to solve the reverse problem of reconstructing an unknown peptide from its experimental spectrum.
  - We can start with an Ideal spectrum- An experimental spectrum which is coinciding with the peptide's theoretical spectrum
- Easy Problem:** Peptide Sequence → Spectrum  
**Hard Problem:** Peptide Sequence ← Spectrum

# Cyclopeptide Sequencing Problem

- Cyclopeptide Sequencing Problem:

Given an ideal spectrum, find a cyclic peptide whose theoretical spectrum matches the experimental spectrum.

- *Input: A collection of (possibly repeated) integers Spectrum corresponding to an ideal experimental spectrum.*
- *Output: An amino acid string Peptide such that Cyclospectrum(Peptide) = Spectrum (if such a string exists).*
- *I.e., theoretical spectrum = experimental spectrum*

# Cyclopeptide Sequencing Problem

- Consider the theoretical spectrum for Tyrocidine B1 shown below

0	97	99	113	114	128	128	147	147	163	186	227
241	242	244	260	261	262	283	291	333	340	357	388
389	390	390	405	430	430	447	485	487	503	504	518
543	544	552	575	577	584	631	632	650	651	671	672
690	691	738	745	747	770	778	779	804	818	819	835
837	875	892	892	917	932	932	933	934	965	982	989
1031	1039	1060	1061	1062	1078	1080	1081	1095	1136	1159	1175
1175	1194	1194	1208	1209	1223	1225	1322				

If an experiment produced this spectrum, how would you reconstruct the amino acid sequence of Tyrocidine B1?

# Cyclopeptide Sequencing Problem

- Alphabets of amino acids - 20
- Alphabet of amino acid masses - 18

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W				
57	71	87	97	99	101	101	103	113	113	113	114	115	128	128	129	131	137	147	156	163	186

Two amino acid pairs have the same integer mass

→

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

- Represent a peptide by a sequence of integers denoting the peptide's constituent amino acid masses
- We replaced an alphabet of 20 amino acids with an alphabet of 18 integers

# Cyclopeptide Sequencing Problem

- Amino acid mass representation of peptides
  - NQEL as
  - 114-128-129-113
  - Tyrocidine B<sub>1</sub> (VKLFPWFNQY) as
  - 99-128-113-147-97-186-147-114-128-163

We don't know whether there exists different peptides (in the alphabet of 18 amino acid masses) with identical theoretical spectra – can we find such peptides?  
(E.g., NQEI, NKEL, NKEI)

# A Brute Force Algorithm for Cyclopeptide Sequencing

- Denote the total mass of an amino acid string Peptide as **Mass(Peptide)**.
- In mass spectrometry experiments, the peptide that generated Spectrum is unknown. So Mass(Peptide) can be inferred from Spectrum and is denoted **ParentMass(Spectrum)**.
- For the sake of simplicity, we will assume that for all experimental spectra, ParentMass(Spectrum) is equal to the largest mass in Spectrum.
- *For every peptide sequence with the target peptide's molecular weight*
  - Compute the sequence's Theoretical Spectrum
- *If it matches the one given, report this peptide as a possible solution*

# A Brute Force Algorithm for Cyclopeptide Sequencing

- The brute force cyclopeptide sequencing algorithm
  - generates all possible peptides whose mass is equal to  $\text{ParentMass}(\text{Spectrum})$
  - then checks which of these peptides has theoretical spectra matching Spectrum.

```
BFCyclopeptideSequencing(Spectrum)
for every peptide with Mass(Peptide) equal to ParentMass(Spectrum)
    if Spectrum = Cyclospectrum(Peptide)
        output Peptide
```

## Brute Force Algorithm for Cyclopeptide Sequencing - Example

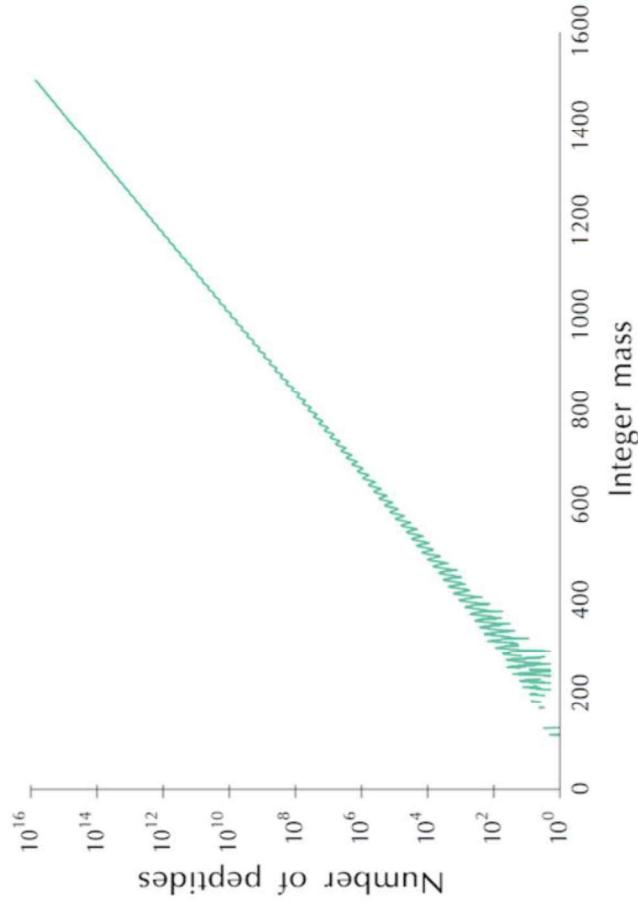
- Let the mass of some sample alphabets be given as
  - A: 2, B: 3, C:5, D:4, E:1
- Let Experimental Spectrum is [ 0 1 4 5 5 6 9 10]
- Parent Mass from Experimental Spectrum = 10
- Generate all substring and Possible candidate strings are(Mass equal to parent mass)
  - ABC
  - CDE
  - ABDE
  - Etc..

## Brute Force Algorithm for Cyclopeptide Sequencing - Example

- Generate theoretical spectrum of each of the above
  - ABC : A(2),B(3),C(5), AB(5),AC(7),BC(8),ABC(10)
  - CDE : C(5), D(4),E(1),CD(9),DE(5),EC(6),CDE(10)
  - Etc..
- Compare Experimental and theoretic spectrum of such peptides, If they match, the corresponding peptide is the solution.
  - *Here, CDE is the solution*

# A Brute Force Algorithm for Cyclopeptide Sequencing

- Guarantees a solution always, but not efficient in terms of running time.
- There are trillions of peptides that have the same integer mass (1322) as Tyrocidine B1.
- Therefore, Brute force approach is completely impractical.
- For large  $m$ , the number of peptides with given integer mass  $m$  can be approximated as  $k \cdot C^m$ , where  $k$  and  $C$  are constants.



# Summary

- Cyclopeptide Sequencing Problem
- Brute Force Algorithm for Cyclopeptide Sequencing Problem
- Counting Peptides with Given mass
- Brute Force Algorithm - Illustration