



DEEMED TO BE UNIVERSITY

22BIO211: Intelligence of Biological Systems - 2

MASS SPECTROMETRY

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

Mass Spectrometry

- The mass spectrometer measures the mass of a molecule in daltons (Da);
 - 1 Da is approximately equal to the mass of a single nuclear particle (i.e., a proton or neutron).
 - 1 Da is not exactly equal to the mass of a proton/neutron, and we may need to account for different naturally occurring isotopes of each atom when weighing a molecule
- Integer Mass
 - amino acid glycine - Chemical formula C_2H_3ON
 - $2 \cdot 12 + 3 \cdot 1 + 1 \cdot 16 + 1 \cdot 14 = 57.$
 - amino acids typically have non-integer masses (e.g., glycine has total integer mass equal to approximately 57.02 Da)

Mass Spectrometry

- Integer Mass of Aminoacids

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	113	114	115	128	128	129	131	137	147	156	163	186

Daltons = {

'A': 71, 'C': 103, 'D': 115, 'E': 129,
'F': 147, 'G': 57, 'H': 137, 'I': 113,
'K': 128, 'L': 113, 'M': 131, 'N': 114,
'P': 97, 'Q': 128, 'R': 156, 'S': 87,
'T': 101, 'V': 99, 'W': 186, 'Y': 163

}

Mass Spectrometry

- Integer Mass of Tyrocidine B1

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

- *Amino acid sequence - VKLFPWFNQY*
- $99 + 128 + 113 + 147 + 97 + 186 + 147 + 114 + 128 + 163 = 1322 \text{ Da}$

Calculate total molecular weight of our target Molecule

- What is the total molecular weight of Tyrocidine B1?

```
# Here's a new dictionary!
Daltons = {
    'A': 71, 'C': 103, 'D': 115, 'E': 129,
    'F': 147, 'G': 57, 'H': 137, 'I': 113,
    'K': 128, 'L': 113, 'M': 131, 'N': 114,
    'P': 97, 'Q': 128, 'R': 156, 'S': 87,
    'T': 101, 'V': 99, 'W': 186, 'Y': 163
}
```

```
TyrocidineB1 = "VKLFPWFNQY"

# The weight of Tyrocidine B1
print(sum([Daltons[res] for res in TyrocidineB1]))
```

Mass Spectrometry - Theoretical Spectrum

- Mass spectrometer can break each molecule of Tyrocidine B1 into two linear fragments.
- One copy may break into LFP and WFNQYVK (with respective masses 357 and 965), whereas another may break into PWFN and QYVKLF.
- The collection of all the fragment masses generated by the mass spectrometer is called an **experimental spectrum**.
- **subpeptides**- possible linear fragments of the peptide.
- The **theoretical spectrum** of a cyclic peptide Peptide, denoted Cyclospectrum(Peptide), is the collection of all of the masses of its subpeptides
 - *in addition to the mass zero and the mass of the entire peptide*
 - *with masses ordered from smallest to largest.*

Theoretical Spectrum

- Eg: Cyclic peptide NQEL
 - Cyclic Peptide : ELEL
 - 12 subpeptides:
 - *E, L, E, L*
 - *EL, LE, EL, LE*
 - *ELE, LEL, ELE, LEL*
 - *Nil and ELEL*
- The number of possible subpeptides from a peptide of length n will always be $n^*(n-1)$
- Including zero and the full peptide, there will be $n^*(n-1) + 2$ entries

Theoretical Spectrum of NQEL

- Theoretical Spectrum of NQEL

L	N	Q	E	LN	NQ	EL	QE	LNQ	ELN	QEL	NQE	NQEL	
0	113	114	128	129	227	242	242	257	355	356	370	371	484

- Generating Theoretical Spectrum Problem:

Generate the theoretical spectrum of a cyclic peptide.

- *Input: An amino acid string Peptide.*
- *Output: Cyclospectrum(Peptide).*

- Step1: Construct a list containing all subpeptides of Peptide

- Step2: Find the mass of each subpeptide by adding the integer masses of its constituent amino acids.

Generating Theoretical Spectrum Problem

- Our approach to generating theoretical spectrum is based on the following assumption
 - *mass of any subpeptide is equal to the difference between the masses of two prefixes of Peptide.*
- peptide - NQEL
 - L N Q E LN QE LNQ ELN QEL NQEL
 - 0 113 114 128 129 227 242 257 355 356 370 371 484
- Assume Linear Peptide : Calculate the prefix masses
 - *NQEL*
 - N NQ NQE NQEL
 - 114 242 371 484
 - PREFIXMASS = (0, 114, 242, 371, 484).
 - $MASS(QE) = PREFIXMASS(NQE) - PREFIXMASS(N)$ $= PREFIXMASS(3) - PREFIXMASS(1)$ $= 371 - 114 = 257$

Generating Theoretical Spectrum Problem

■ Calculate the prefix masses

For peptide = "NQEL"

```
Daltons = {'A': 71, 'C': 103, 'D': 115, 'E': 129,
          'F': 147, 'G': 57, 'H': 137, 'I': 113,
          'K': 128, 'L': 113, 'M': 131, 'N': 114,
          'P': 97, 'Q': 128, 'R': 156, 'S': 87,
          'T': 101, 'V': 99, 'W': 186, 'Y': 163 }

# Function to calculate Prefix Mass of a cyclic peptide
def CalcPrefixMass(peptide):
    n = len(peptide)
    prefix_mass = [0] * (n + 1)
    frag = ""
    for i in range(1, n + 1):
        prefix_mass[i] = prefix_mass[i - 1] + Daltons[peptide[i-1]]
        frag = peptide[0:i]
        print(frag, prefix_mass[i])
    peptide_mass = prefix_mass[n]

peptide = "NQEL"
CalcPrefixMass(peptide)
```

For peptide = "VKLFPWFNQY"

```
V 99
VK 227
VKL 340
VKLF 487
VKLFP 584
VKLFPW 770
VKLFPWF 917
VKLFPWFN 1031
VKLFPWFNQ 1159
VKLFPWFNQY 1322
```

Generating Theoretical Spectrum Problem

- Assume Cyclic Peptide, then the masses in its theoretical spectrum can be divided into two
 - Those found by Linearspectrum
 - E.g., $\text{MASS}(\text{QE}) = \text{MASS}(\text{NQE}) - \text{MASS}(\text{N})$
 $= 371 - 114 = 257$
 - Those found by sub peptides wrapping around the end of the linearized version of Peptide
 - E.g., $\text{MASS}(\text{LN}) = \text{MASS}(\text{NQEL}) - \text{Mass}(\text{QE})$
 $= 484 - 257 = 227$

Generating Theoretical Spectrum Problem

- For Liner Spectrum - the mass of the sub peptide beginning at position $i+1$ and ending at position j can be computed as $\text{PREFIXMASS}(j) - \text{PREFIXMASS}(i-1)$
- $\text{Mass}(\text{QE})$ is at $\text{index}(2,3) - > \text{Mass}(3) - \text{Mass}(1)$
- For cyclic Spectrum – the subpeptide has a mass equal to the difference between $\text{MASS}(\text{Peptide})$ and a subpeptide mass identified by Linearspectrum.
- $\text{MASS}(\text{LN}) = \text{MASS}(\text{NQEL}) - \text{Mass}(\text{QE}) = 484 - 257 = 227$

Generating Theoretical Spectrum Problem

```
CYCLICSPECTRUM(Peptide, AMINOACID, AMINOACIDMASS)
PREFIXMASS(0) ← 0
for i ← 1 to Peptide
    for j ← 1 to 20
        if AMINOACID(j) = i-th amino acid in Peptide
            PREFIXMASS(i) ← PREFIXMASS(i - 1) + AMINOACIDMASS(j)
peptideMass ← PREFIXMASS(|Peptide|)

CyclicSpectrum ← a list consisting of the single integer 0
for i ← 0 to |Peptide| - 1
    for j ← i + 1 to |Peptide|
        add PREFIXMASS(j) - PREFIXMASS(i) to CyclicSpectrum
        if i > 0 and j < |Peptide|
            add peptideMass - (PREFIXMASS(j) - PREFIXMASS(i)) to CyclicSpectrum

return sorted list CyclicSpectrum
```

Calculate the prefix
masses

Generating Theoretical Spectrum Problem

```
# To calculate theoretical spectrum of a cyclic peptide
def theoretical_spectrum(peptide):
    prefix_mass = CalcPrefixMass(peptide)
    n=len(peptide)
    peptide_mass = prefix_mass[n]
    spectrum = list()
    spectrum.append(0)
    subpeptide =list()
    subpeptide.append(" ")
    for i in range(0, n):
        for j in range(i+1,n+1):
            subpeptide_mass = prefix_mass[j] - prefix_mass[i]
            spectrum.append(subpeptide_mass)
            subpeptide.append(peptide[i:j])
            if i > 0 and j < n:
                subpeptide_mass = peptide_mass -(prefix_mass[j] - prefix_mass[i])
                spectrum.append(subpeptide_mass)
            spectrum.append(subpeptide_mass)
    print (len(subpeptide),": ",subpeptide)
    print (len(spectrum),": ",spectrum)
    return(sorted(spectrum))
```

Generating Theoretical Spectrum Problem

NQEL - Linear sub peptides

- i = 0, j = 1,2,3,4
N, NQ, NQE, NQEL
- i = 1, j = 2,3,4
Q, QE, QEL
- i = 2, j = 3,4
E, EL
- i = 3,j=4
L

NQEL - Cyclic Subpeptides

- i = 1, j = 2
peptide - Q -> ELN
- i = 1, j = 3
peptide - QE -> LN
- i = 2, j = 3
peptide - E -> LNQ

```
peptide = "NQEL"
spectrum = theoretical_spectrum(peptide)
print ("The theoretical spectrum of peptide ", peptide, "is : ")
print(spectrum)
```

N 114

NQ 242

NQE 371

NQEL 484

14 : [' ', 'N', 'NQ', 'NQE', 'NQEL', 'Q', 'ELN', 'QE', 'LN', 'QEL', 'E', 'LNQ', 'EL', 'L']
14 : [0, 114, 242, 371, 484, 128, 356, 257, 227, 370, 129, 355, 242, 113]

The theoretical spectrum of peptide NQEL is :

[0, 113, 114, 128, 129, 227, 242, 242, 257, 355, 356, 370, 371, 484]

Theoretical Spectrum

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.

Summary

- Mass Spectrometry
- Theoretical Spectrum
- Generating Theoretical Spectrum