

CMSC423

Chapter 4 – Proteomics/mass-
spectrometry
Leaderboard searching

Warmup

Assume you sail to a remote island and discover a hidden treasure. You want to bring as much of the treasure back with you, but you cannot overload your boat and return safely. You, thus, decide to pick a set S of artifacts that will fit in your boat, and which, in combination, are the most valuable. Since you are a computer scientist, you decide to model the problem as follows:

Your boat can carry at most W total weight. Each artifact i has weight w_i and value v_i

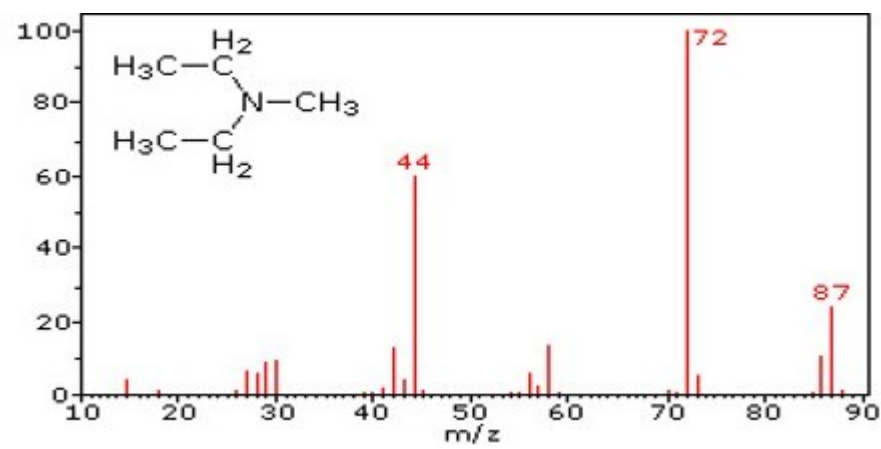
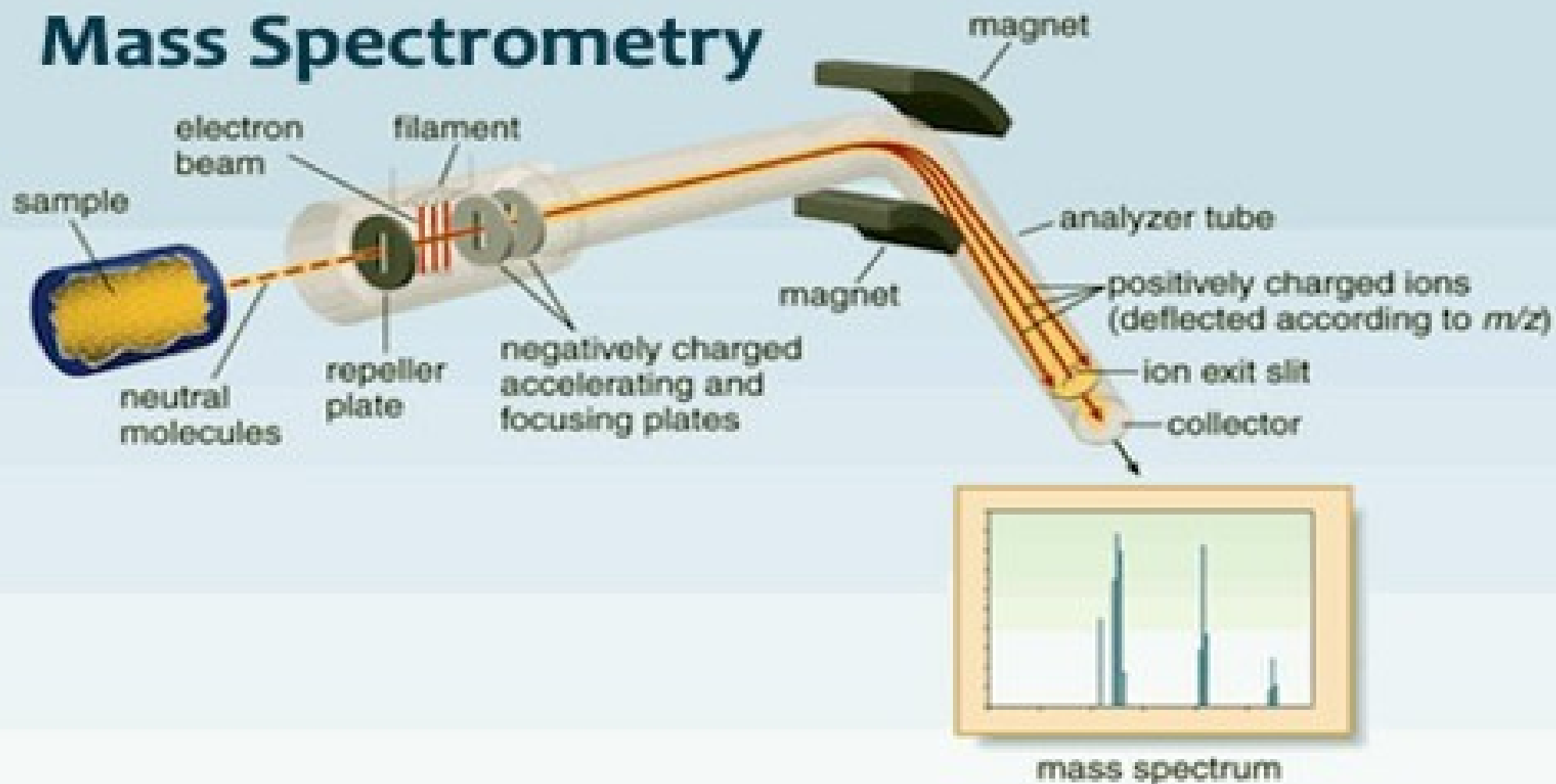
You need to identify S such that

$$\sum_{obj_i \in S} w_i \leq W \quad \text{and} \quad \sum_{obj_i \in S} v_i \quad \text{is maximized over all choices of } S$$

Class so far...

- Deterministic searching (counting, clumps, KMP)
- Randomized searching (Gibbs sampling)
- This week: Branch and bound search

Mass Spectrometry



Random breakage

Glu Leu Val Ile Ser Ile Ser Ala Leu Ile Val Glu

ELVISISALIVE weight (ELVISISALIVE)

E LVISISALIVE + weight(E) + weight (LVISIS...)

EL VISISALIVE + weight (EL) +

ELV ISISALIVE

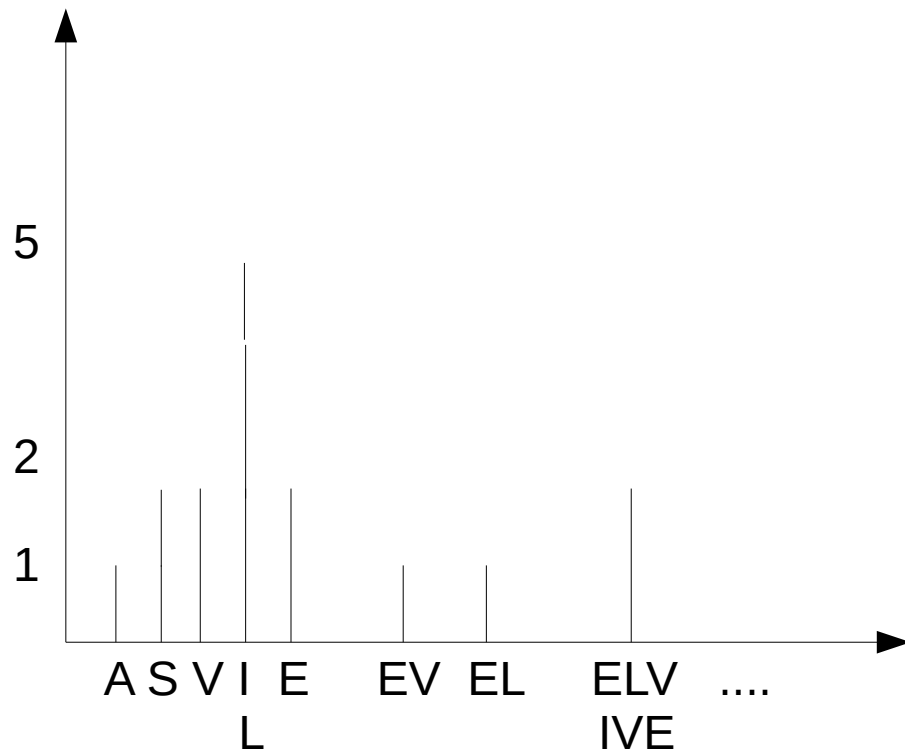
ELVIS ISALIVE

ELVISI SALIVE

ELVISIS ALIVE

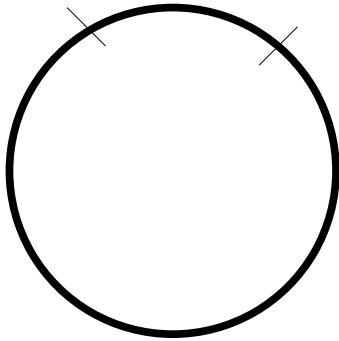
...

Peptide spectrum



Cyclic spectrum?

Exactly 2 cuts



ELVISISALIVE

E LVISISALIVE

L VISISALIVEE

...

ELVI SISALIVE

LVIS ISALIVEE

...

SALI VEELVISI

...

Whole peptide + all possible breaks into 2 pieces

What is the runtime for creating the cyclic spectrum of a peptide of size length k ?

Our goal

- Given a real spectrum
- Find peptide that generated it

- The other way around is easy:
 - break peptide in every each way
 - calculate weights
 - compute predicted spectrum

Next: the algorithm