CMSC423

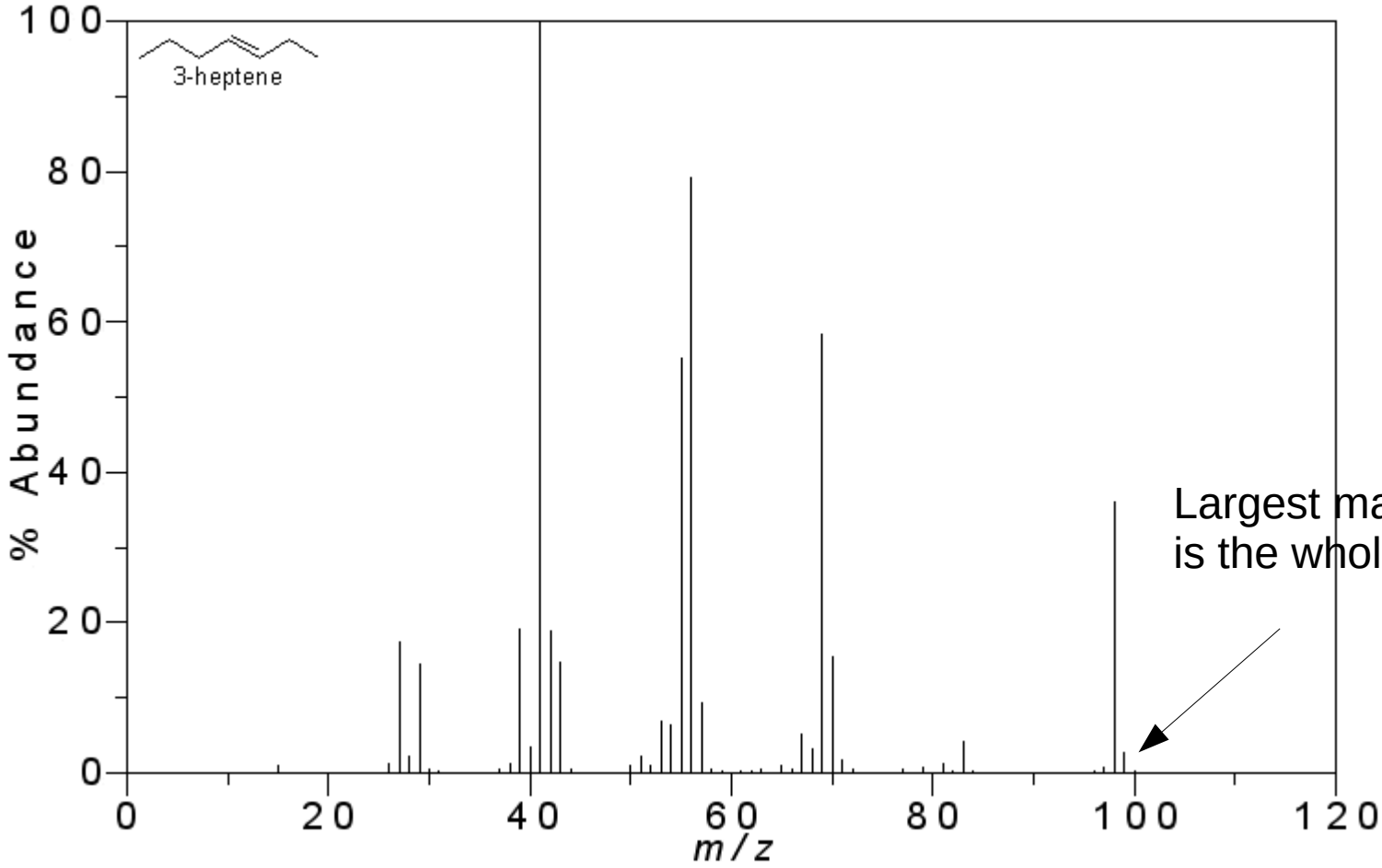
Chapter 4 – Proteomics/mass-spectrometry
Dealing with perfect spectra

General approach

- Guess a peptide
- See how well it matches spectrum
- Come up with new guess
- Repeat

- Sound familiar?

Some quick insights



How many peptide words have a given mass?

How many combinations of coins and bills make a same $ amount?, e.g., $0.5

brute force – not so simple

1. 1 + 1 + 1 + 1 + ... + 1 = $0.50
2 . 1 + 1 + 1 + 1 + ... + 5 = $0.50
3. 1 + 1 + 1 + 1 + ... + 5 + 5 = $0.50
...
11. 5 + 5 + 5 + 5 + ... + 5 = $0.50
12. 5 + 5 + 5 + 5 + ... + 10 = $0.50
...
16. 10 + 10 + 10 + 10 + 10 = $0.50
17. 1 + 1 + 1 + 1 + 1 + 5 + ... + 5 + 10 = $0.50
...

MANY!

Stop and Think!

Can you compute the number of ways
to make change from a certain $ amount?

Dynamic programming can help

The order of letters matters

- Weight(ELVISLIVES) = Weight(EVVIISLLES)

- Stop and Think!

  Can the whole spectrum (not just total mass) help?

Yes!

- ELV – ISLIVES
- EVV – IISLLES

No peptide EVV in first string, nor ELV in second.

Algorithm 1

- Assume experimental spectrum is perfect
- Generate all peptides of length 1
- Discard the ones not found in spectrum
- Extend the remaining ones by one amino acid
- Discard the ones incompatible with spectrum
- Repeat...
- ...Until one peptide has exact same spectrum as experimental one

Matching spectra

Peptide: GASP 57-71-87-97

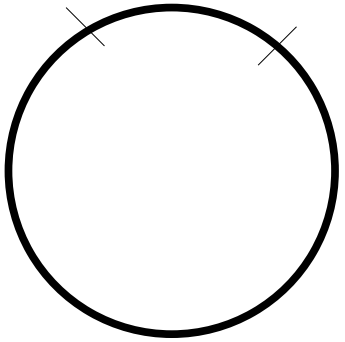
Experimental:	G	A	S	P	GA	PG	AS	SP	GAS	PGA	SPG	ASP	**GASP**
	57	71	87	97	128	154	158	184	215	225	241	255	**312**

Partial peptides (bold if matching):

GAS:	G	A	S	GA	AS	GAS	Consistent with spectrum
	57	**71**	**87**	**128**	**158**	**215**	
APS:	A	S	P	AP	PS	APS	Inconsistent with spectrum
	71	**87**	**97**	168	**184**	**255**	

Linear or circular spectrum?

Exactly 2 cuts



ELVISISALIVE

E LVISISALIVE
L VISISALIVEE
...
ELVI SISALIVE
LVIS ISALIVEE
...
SALI VEELVISI
...

Split all circular rotations of the string in 2 pieces
or
isolate 1 letter, 2 letters, 3 letters, etc. from circularized string

What if we only recovered part of spectrum (partial solution)?

Linear spectrum

Circular spectrum

ELVISISALIVE

E LVISISALIVE
L VISISALIVEE
...
ELVI SISALIVE
LVIS ISALIVEE
...
SALI VEELVISI
...

In partial string, rotated strings are not possible

ELVISISAL

E LVISISAL
~~L VISISALIVEE~~
...
ELVI SISAL
~~LVIS ISALIVEE~~
...
~~SALI VEELVISI~~
...

Linear spectrum - pieces of linear peptide that could be part of circular spectrum

ELVISISALIVE

E LVISISALIVE

E **L** VISISALIVE

EL **V** ISISALIVE

EL **VI** SISALIVE

Summary

- Create table of peptides of increasing length
- Check each peptide's LINEAR spectrum against experimental spectrum (check for containment)
- Discard peptides with masses not in experimental spectrum
- Stop when one peptide has CIRCULAR spectrum matching experimental spectrum

Next: dealing with imperfect spectra