Chapter 2 Exercises

Jeremy Wong

CH-419 – Protein mass spectrometry and proteomics

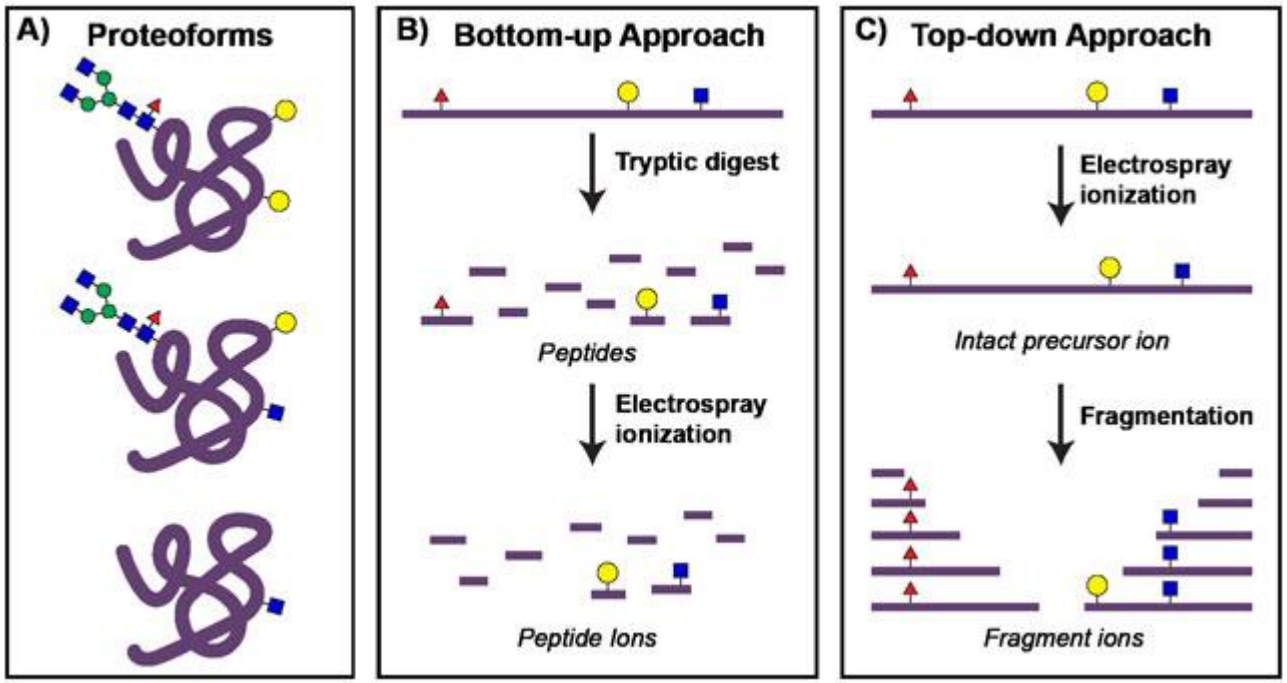
April 6, 2022

Question 1

Lecture slides: 5, 31

In what workflow(s) are peptides analyzed?
□ Top-down □ Shotgun □ Bottom-up □ Native MS

- Native MS: analysis of in-tact proteins
- Top down: analysis of in-tact proteins
- Shotgun and bottom up: peptides



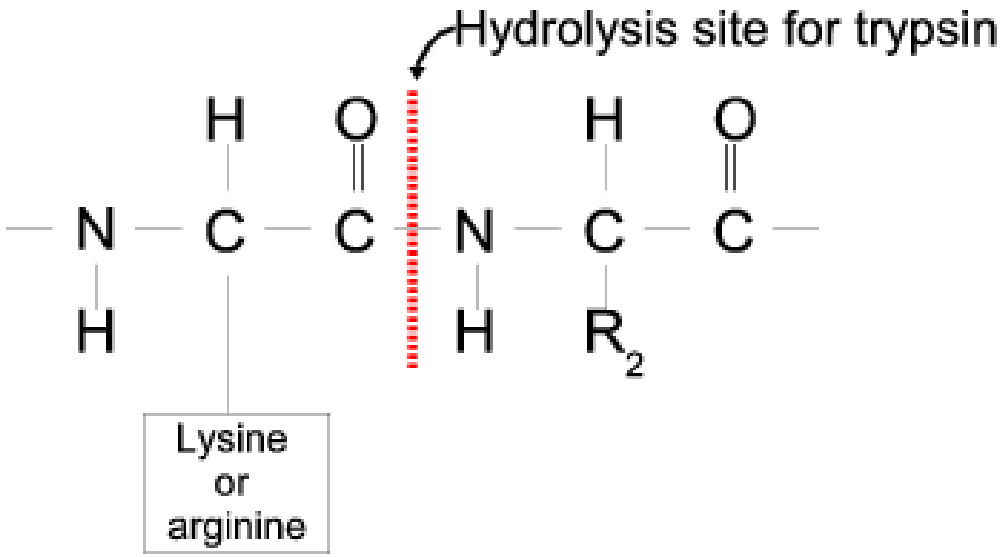
Question 2

What is the most common enzyme used for protein digestion in MS-based proteomics?

□ Lys-C □ Trypsin □ Chymotrypsin □ Pepsin

Question 2

Lecture slide(s): 39



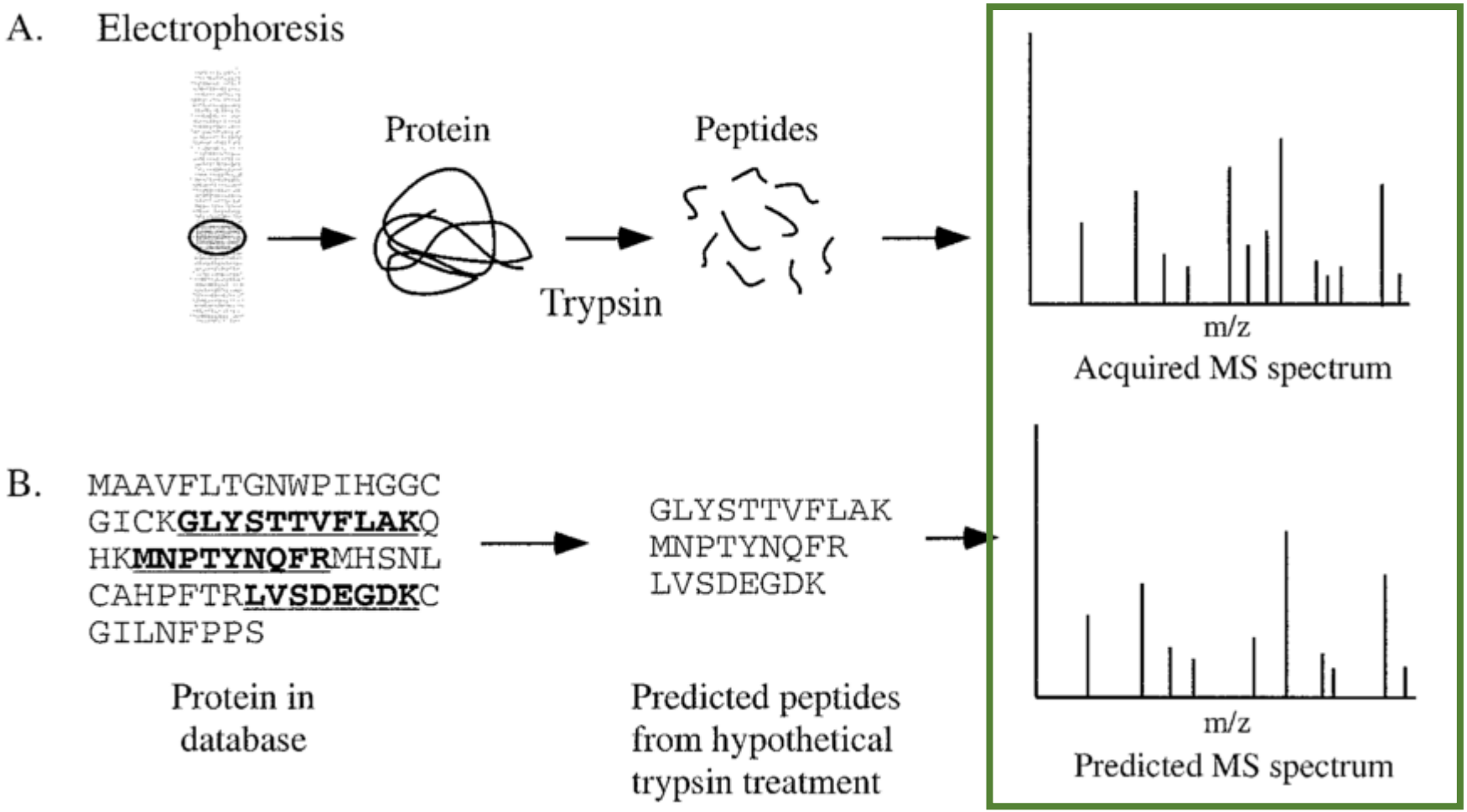
- Answer: Trypsin
- Trypsin cleaves at the C-terminal side of Lysine and Arginine residues
- Note: If a proline residue is the amino acid following Lys/ Arg, cleavage may be inhibited.

Question 3

Explain the principle of peptide mass fingerprinting (PMF)?

Peptide mass fingerprinting: MS^1 level

Compare these and match



Lecture slide(s): 8

Question 4

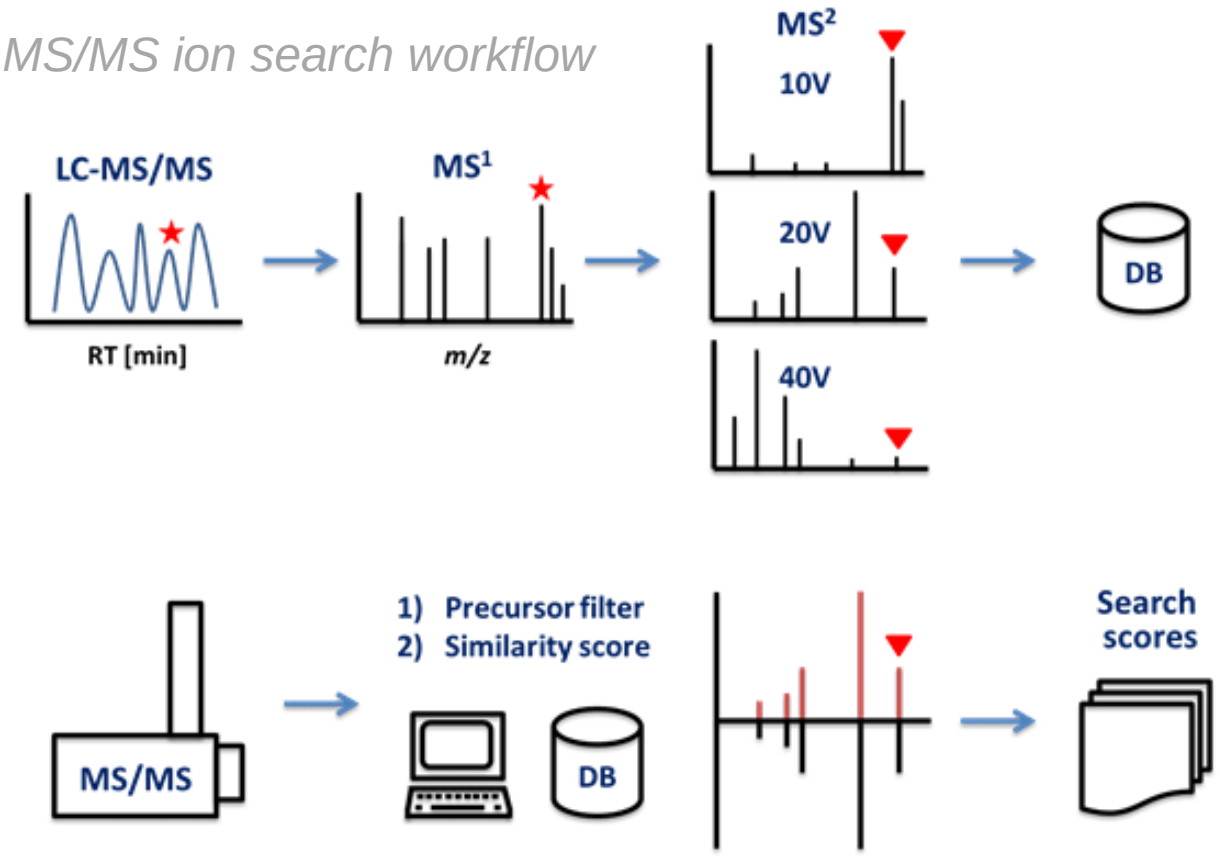
What is the difference between PMF and MS/MS ion search?

MS/MS Ion Search vs PMF

- MS/MS ion search: compares experimental spectra to theoretical spectra at the MS/MS level (fragmentation)

- PMF: compares experimental spectra to theoretical peptides generated upon tryptic digestion or any other type of digest at the MS^1 level.

MS/MS ion search workflow

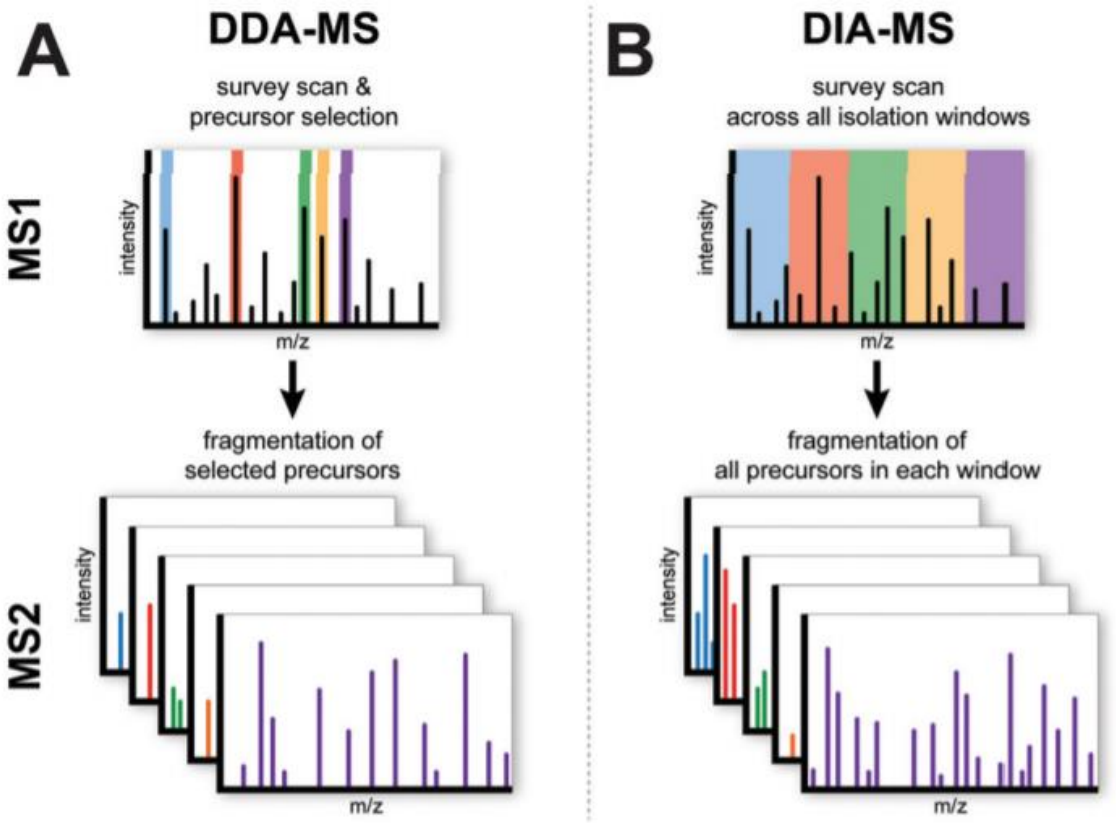


Question 5

For what does DDA stand for?

- □ Data-derived acquisition
- □ Data-dependent acquisition
- □ Data-independent acquisition
- □ None of these

DDA: Data dependent acquisition

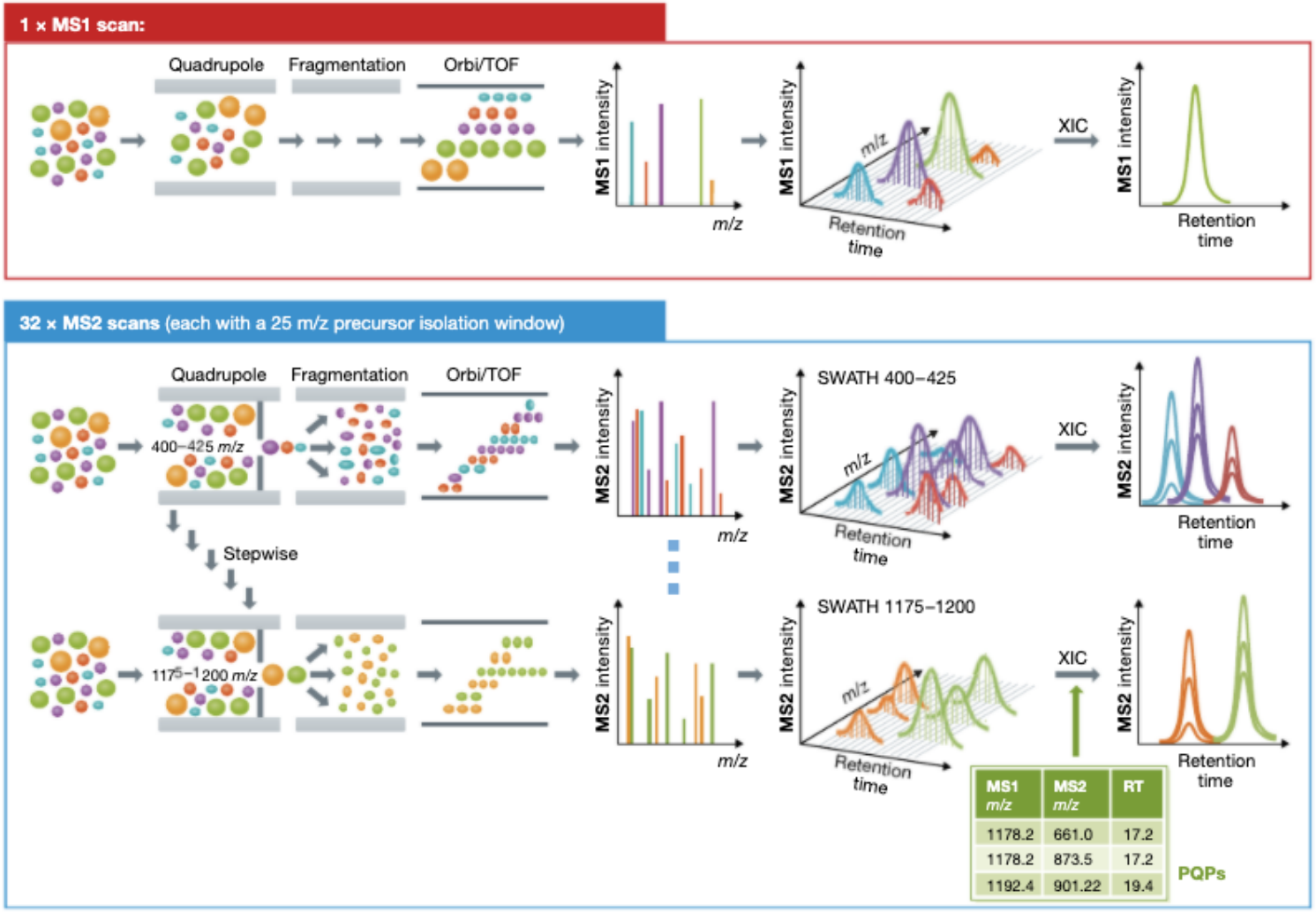


- Answer: data dependent acquisition
- In DDA: you fragment only selected precursors in a given scan window (usually the n most abundant peaks from MS^1)
- In DIA: you fragment all precursors within a given isolated scan window. Assign peaks to the appropriate parent ion. Much tougher in terms of data analysis

Krasny, L. et. al. *Mol. Omics* (2020)

Lecture slide(s): 14-18

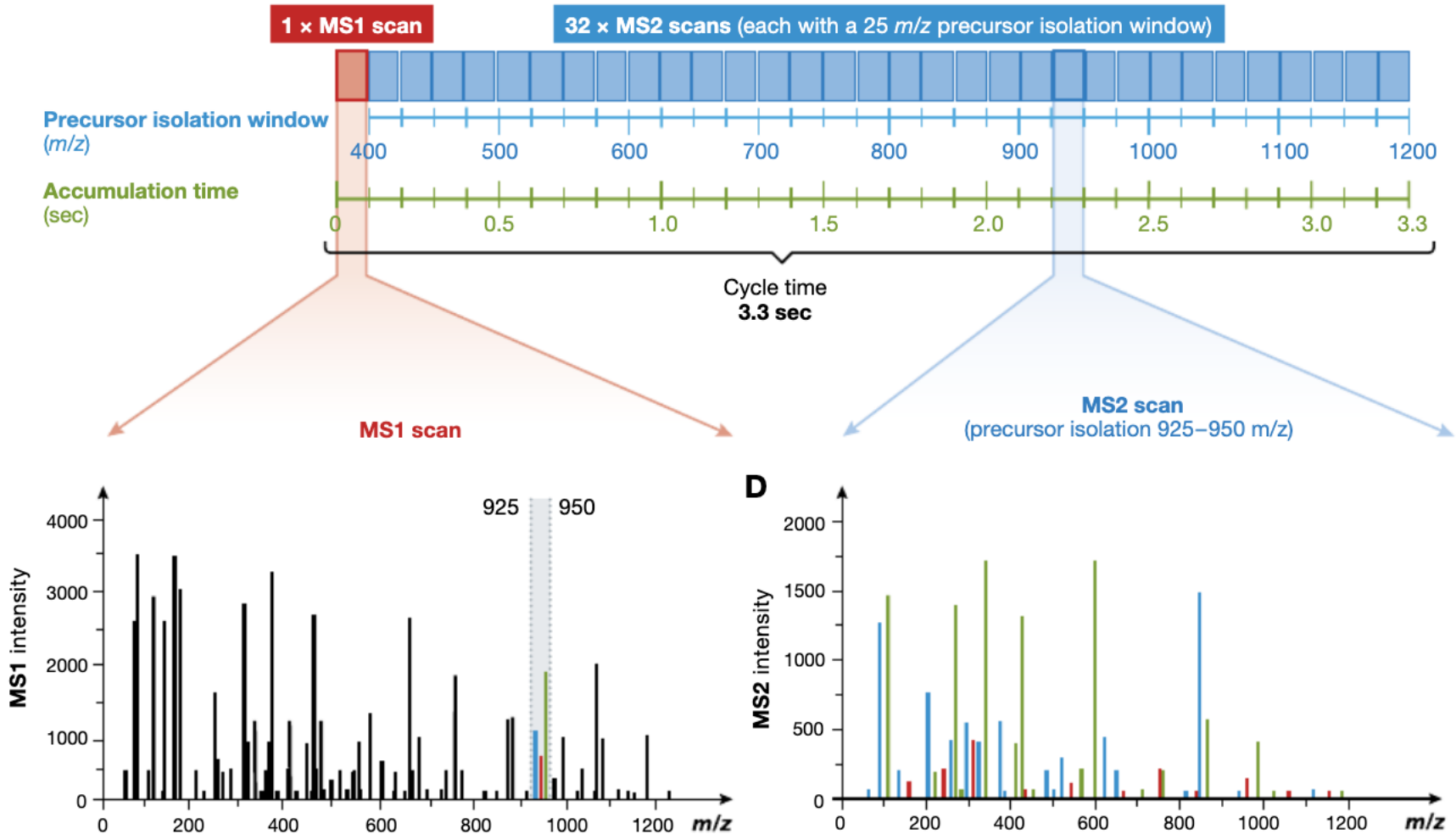
Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS)



Prof Ruedi Aebersold, ETHZ

Ludwig, C. et. al. *Mol. Sys. Biol.* (2018)

Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS)



Ludwig, C. et. al. *Mol. Sys. Biol.* (2018)

Question 6

Lecture slide(s): 14-18

In DDA, typically how many tandem mass spectra are recorded after an MS survey scan?
□ One □ Ten □ One hundred □ It depends

- Correct answers: 10/ it depends
- Previous slide - In DDA: you fragment only selected precursors in a given scan window **(usually the n most abundant peaks from MS_1)**
- **Typical number is 10, but again, depends on situation.**

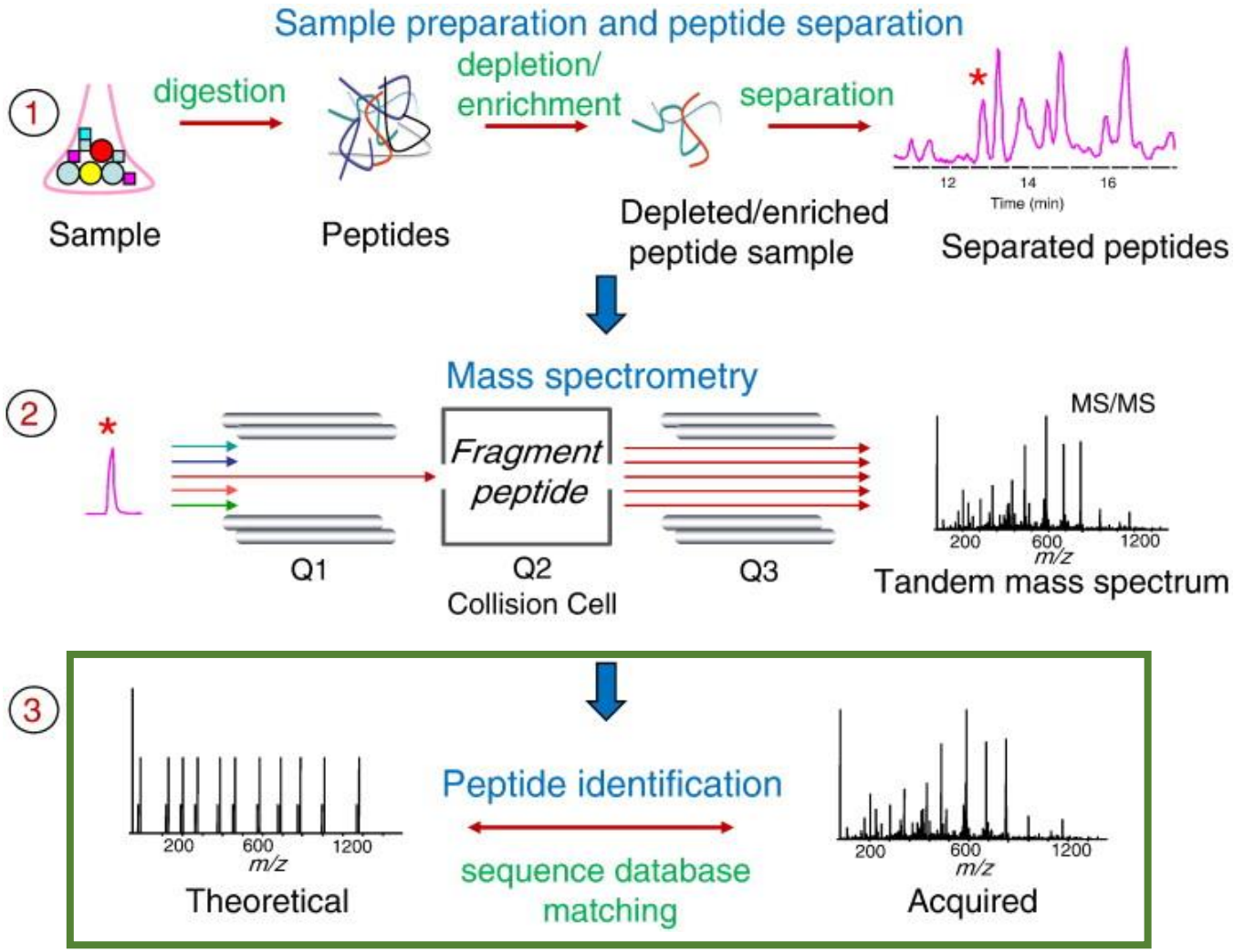
Question 7

In shotgun proteomics, how proteins are finally identified?

□ By comparing experimental tandem mass spectra with theoretical in silico generated tandem mass spectra
□ By using DNA sequence databases
□ By using mass information only
□ By de novo annotating every tandem mass spectrum

Shotgun proteomics workflow

Lecture slide(s): 5



- Answer: By comparing experimental tandem mass spectra with theoretical in silico generated tandem mass spectra

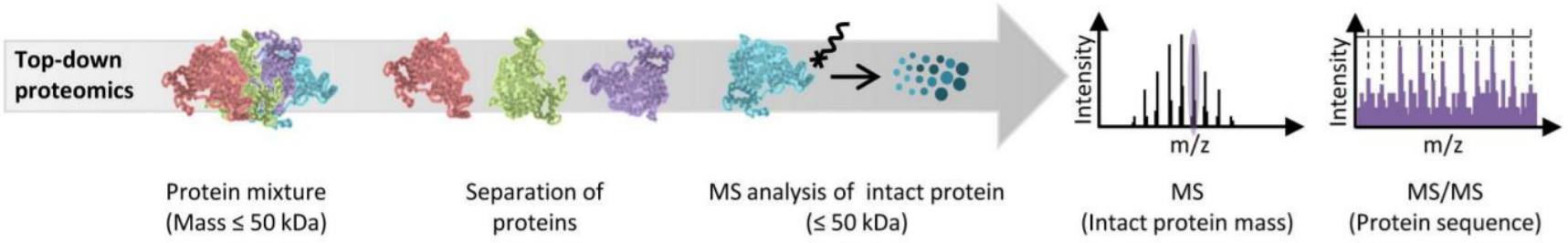
Question 8

Why top-down is relevant?

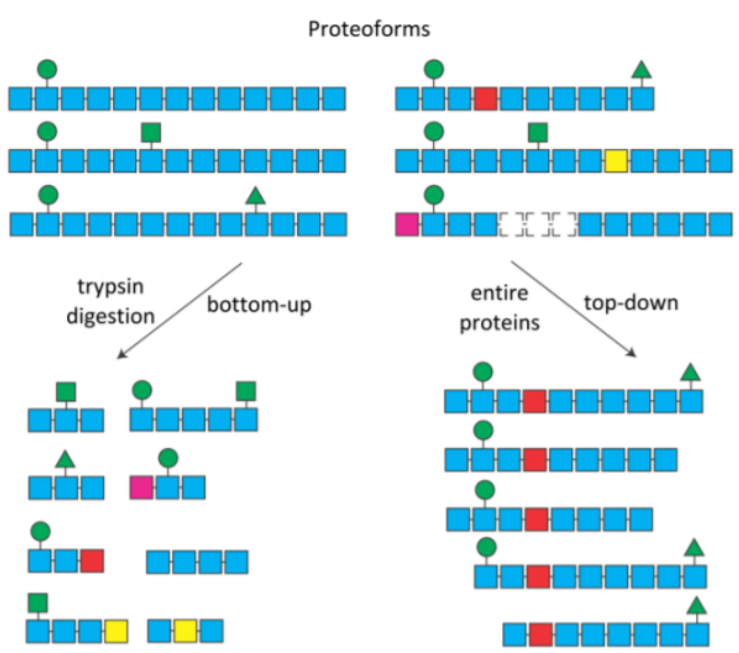
- □ Every proteomic lab is using top-down workflows
- □ Proteoforms can be characterized
- □ Protein sequence coverage is comprehensive
- □ Entire proteins are easier to separate

Top down proteomics

Lecture slide(s): 21-22



J. Proteome Res., 2013, 12 (3), pp 1067–1077



http://proteomique.ipbs.fr/front-page/top-down-proteomics/

- Intact protein analysis means 100% sequence coverage of a given protein
- Analysis of intact protein also provides information regarding proteoforms and post-translational modifications of this protein
- Answer: B and C

Question 9

Lecture slide(s): 21-22

In what application(s) is top-down particularly employed?

□ Characterization of biosimilars
□ Study of posttranslational modifications
□ Study of sequence variants
□ Protein identification

Answers (refer to previous slide):

□ Characterization of biosimilars
□ Study of posttranslational modifications
□ Study of sequence variants

- For protein identification, we do bottom-up.

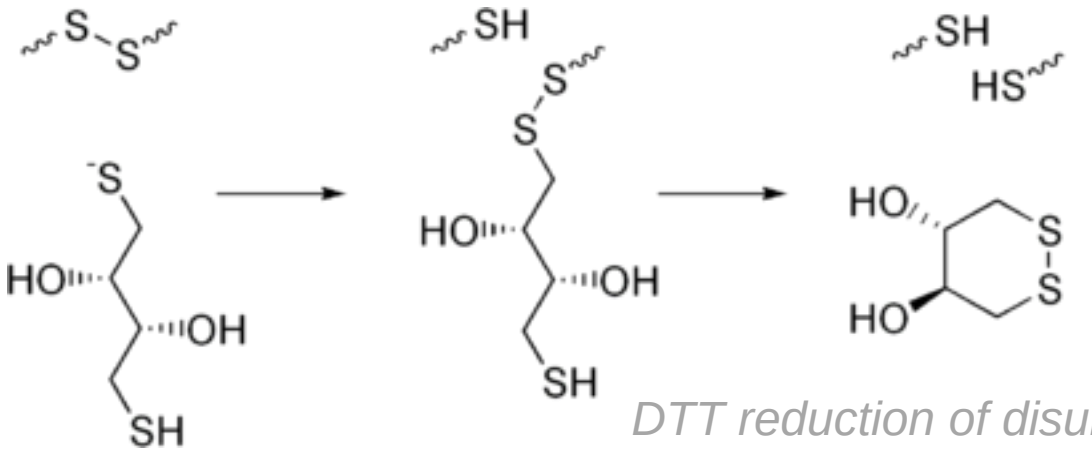
Question 10

For what step(s) of sample preparation are DTT or TCEP used?

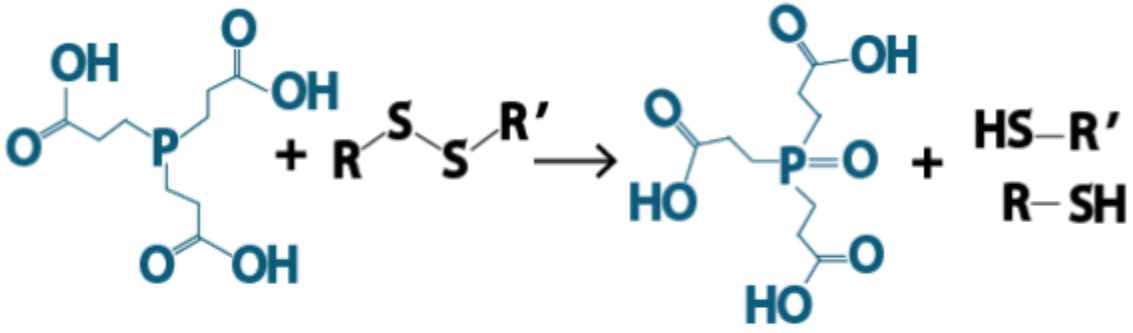
□ Protein digestion □ Cell lysis □ Disulfide bridge Reduction □ Sample fractionation

DTT and TCEP are used for disulfide bridge reduction

Lecture slide(s): 38



DTT reduction of disulfide bond



TCEP Compound containing disulfide **Oxidized Phosphine** Sulfhydril

TCEP reduction of disulfide bond

Question 11

Lecture slide(s): 35

What reagent(s) can be used for cell lysis?
□ Several detergent □ Iodoacetamide □ Sodium dodecyl Sulfate □ Acetone

- Answer: Several detergents or Sodium dodecyl sulfate (SDS)
- Typically, we use detergents to lyse cells. SDS is also a detergent

Question 12

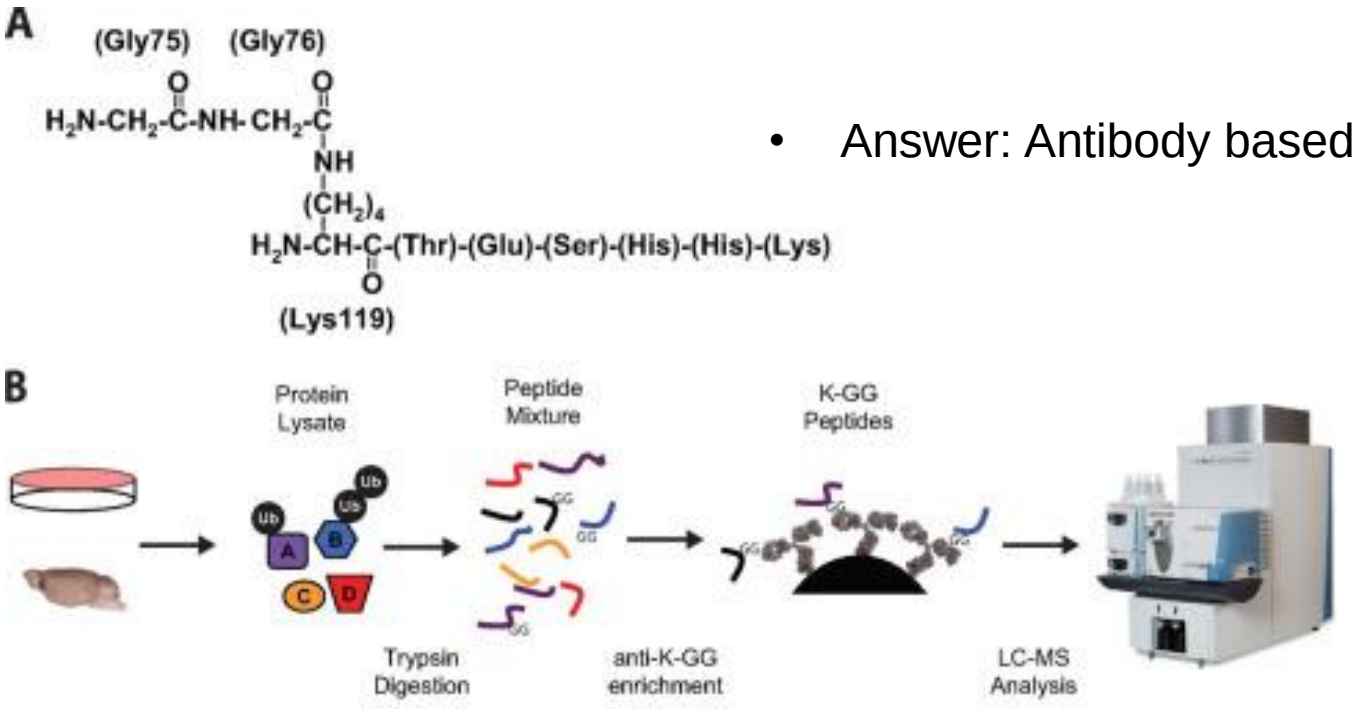
What method(s) is(are) used for protein/peptide enrichment?

□ Antibody-based capture
□ Depletion of abundant proteins
□ Lysis
□ Alkylation

Affinity based peptide enrichment for targeted analysis

Lecture slide(s): 37

- The use of an affinity probe which targets and binds to peptides of interest
- One common example is the use of antibodies which binds to K-GG peptides (a specific signature produced upon tryptic digest of ubiquitinated protein to enrich the lowly abundant ubiquitinated proteins prior to analyzing ubiquitination sites by proteomics



- Answer: Antibody based capture

Question 13

Where does trypsin hydrolyze the peptide bonds?

□ The carboxyl terminal side of arginine and glycine amino acid residues

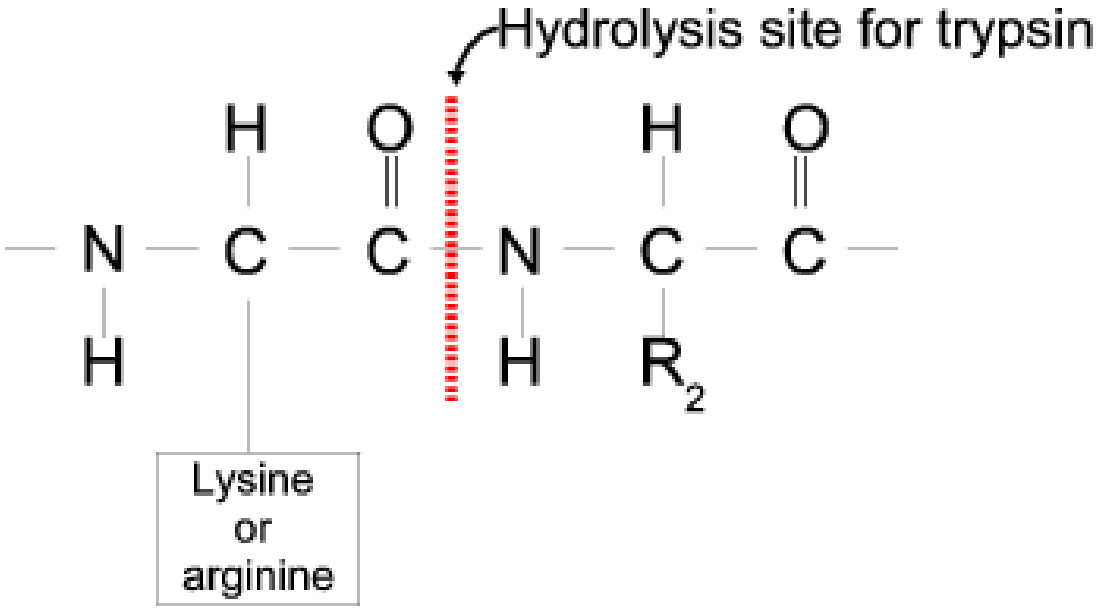
□ The carboxyl terminal side of arginine and lysine amino acid residues

□ The carboxyl terminal side of arginine and tryptophan amino acid residues

□ The carboxyl terminal side of cysteine and methionine amino acid residues

Trypsin's site of peptide bond hydrolysis

Lecture slide(s): 39



- Answer: The C-terminal side of Lysine and Arginine residues
- Note: If a proline residue is the amino acid following Lys/ Arg, cleavage may be inhibited.

Question 14

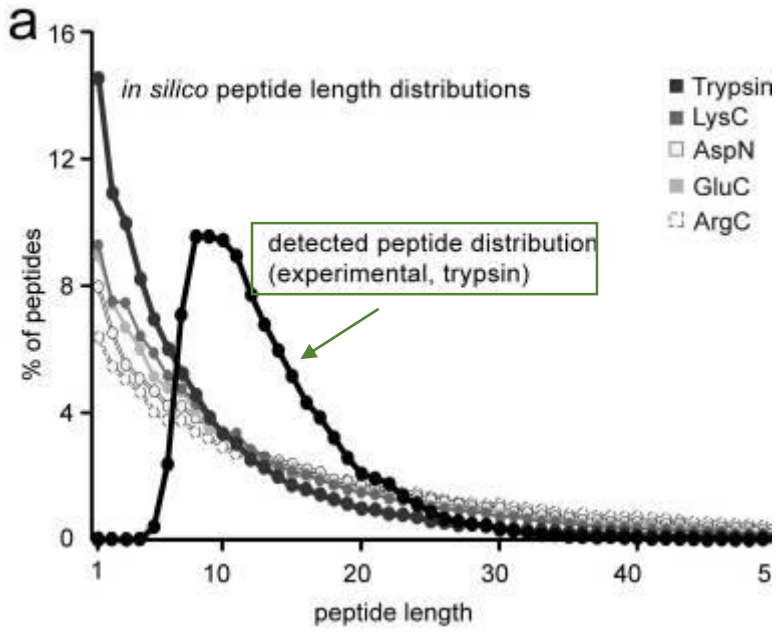
Why using different enzymes for protein digestion?

- □ Increase protein coverage
- □ Generate proteotypic peptides
- □ Trypsin is unspecific and additional enzymes are needed
- □ Digestion efficiency is not enough

Peptide analysis by CID

- **Proteotypic peptides**: peptides that uniquely identify each protein and are consistently observed when a sample mixture is interrogated by a (tandem) mass spectrometer (Mallick et al. 2007).

- In protein mass spectrometry, the most common way of identifying sequences is to do MS/MS on tryptic peptides using collision induced dissociation (CID)



Swaney, D. L. *et al. J. Proteome Res* (2010)

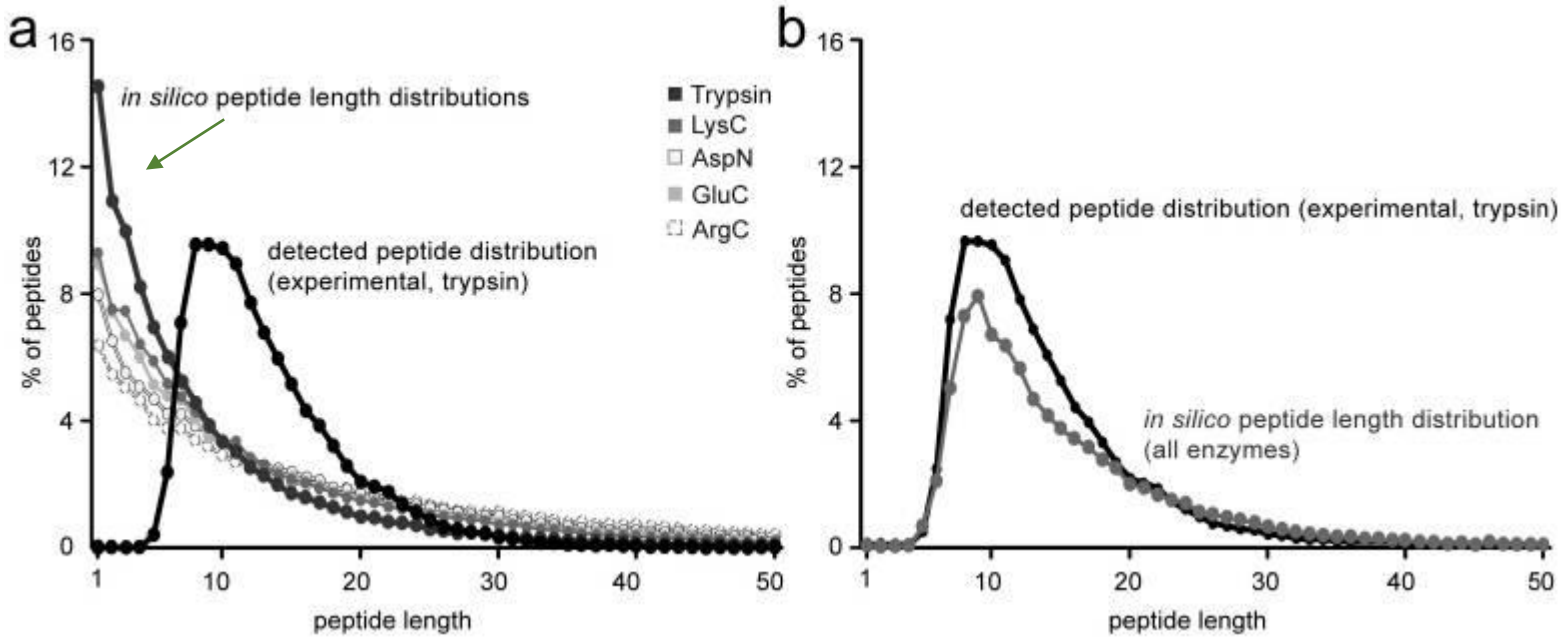
- The size of the peptide formed upon digest plays a role in whether or not they can be sequenced using MS/MS based techniques

- Note that 97% of all peptides identified with MS/MS based techniques fall within a range of 7–35 residues.

Digest with trypsin often form very small peptides

Lecture slide(s): 40

- Unfortunately, tryptic digest often lead to the formation of very small peptides (<6 residues), which cannot be analyzed using mass spec based techniques
- Therefore, in order to observe these peptides, you need to make sure that these specific amino acids in these small peptides are in the correct length
- To do so, you can digest your protein with multiple protease enzymes

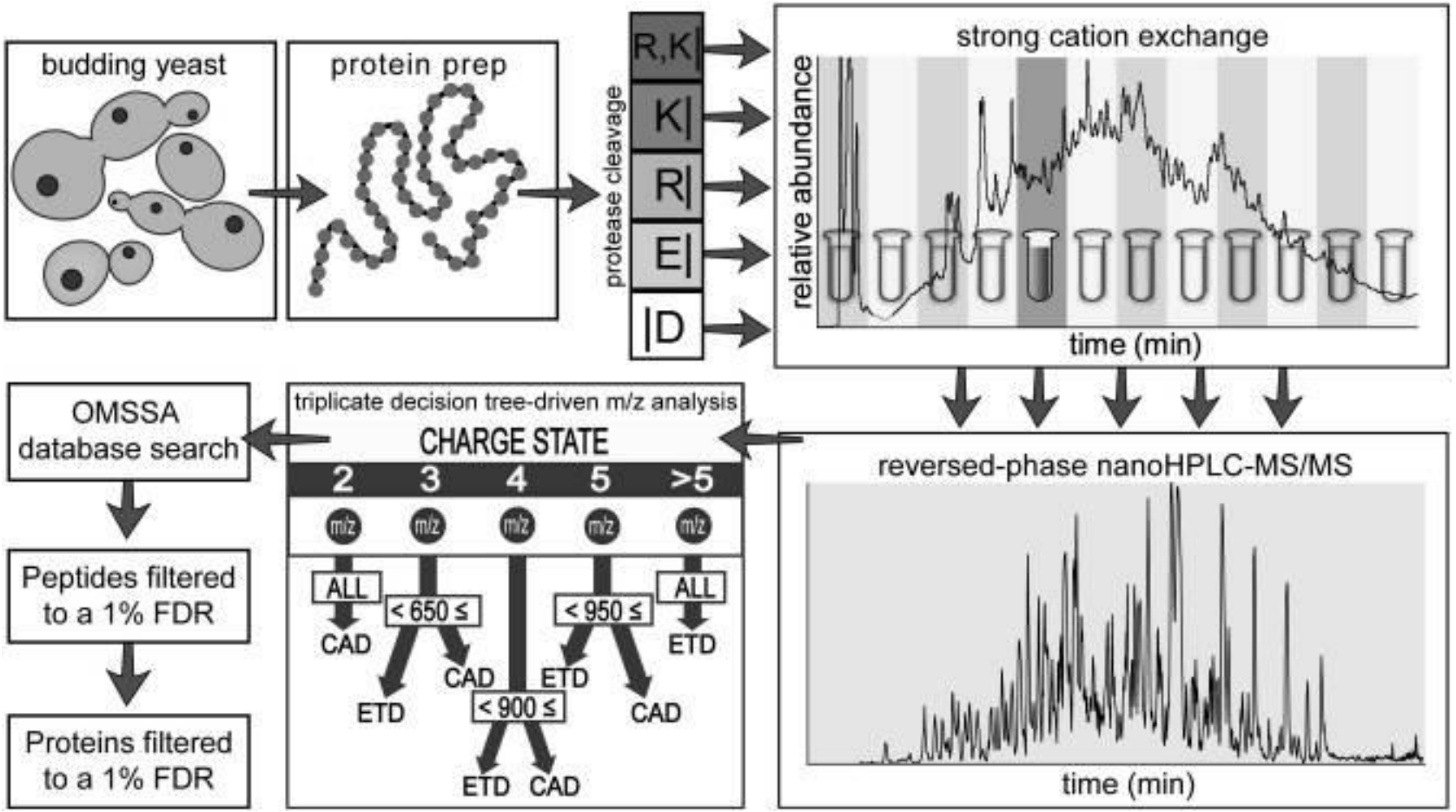


Swaney, D. L. *et al. J. Proteome Res* (2010)

This does not mean add multiple proteases to one tube, but rather separately

Lecture slide(s): 40

- Below is an example workflow in yeast



Swaney, D. L. *et al. J. Proteome Res* (2010)

What is the true beauty of using multiple proteases?

Lecture slide(s): 40

- If you consider the proteome at the amino acid level, we want to cover as many of these AAs as possible in a proteolytic peptide, 7-35 residues in length
- By doing multiple digests with different proteases, you increase the chance of detecting each of these AA in a peptide of the correct length for MS analysis
- This allows you to detect more residues using MS, aka increase sequence coverage

Protease	Trypsin	ArgC	AspN	GluC	LysC	All
Unique peptides	27822	12,452	21,654	17,968	20,619	92,095
CAD	15466	3,518	9,267	7,331	7,807	38,175
ETD	12356	8,934	12,387	10,637	12,812	53,920
Total scans	538,175	540,674	514,607	507,278	524,764	2,625,498
Proteins	3,313	2,708	3,183	2,813	3,030	3,908
Percent of ORFs	56.3	46.0	54.1	47.8	51.5	66.4
Non-redundant amino acids	346,510	191,686	287,188	235,851	304,984	742,312
Non-redundant amino acid proteome coverage (percent)	11.9	6.6	9.8	8.1	10.5	25.5
Average protein sequence coverage (percent)	24.5	18.6	21.5	20.9	24.3	43.4

Swaney, D. L. *et al. J. Proteome Res* (2010)

Question 15

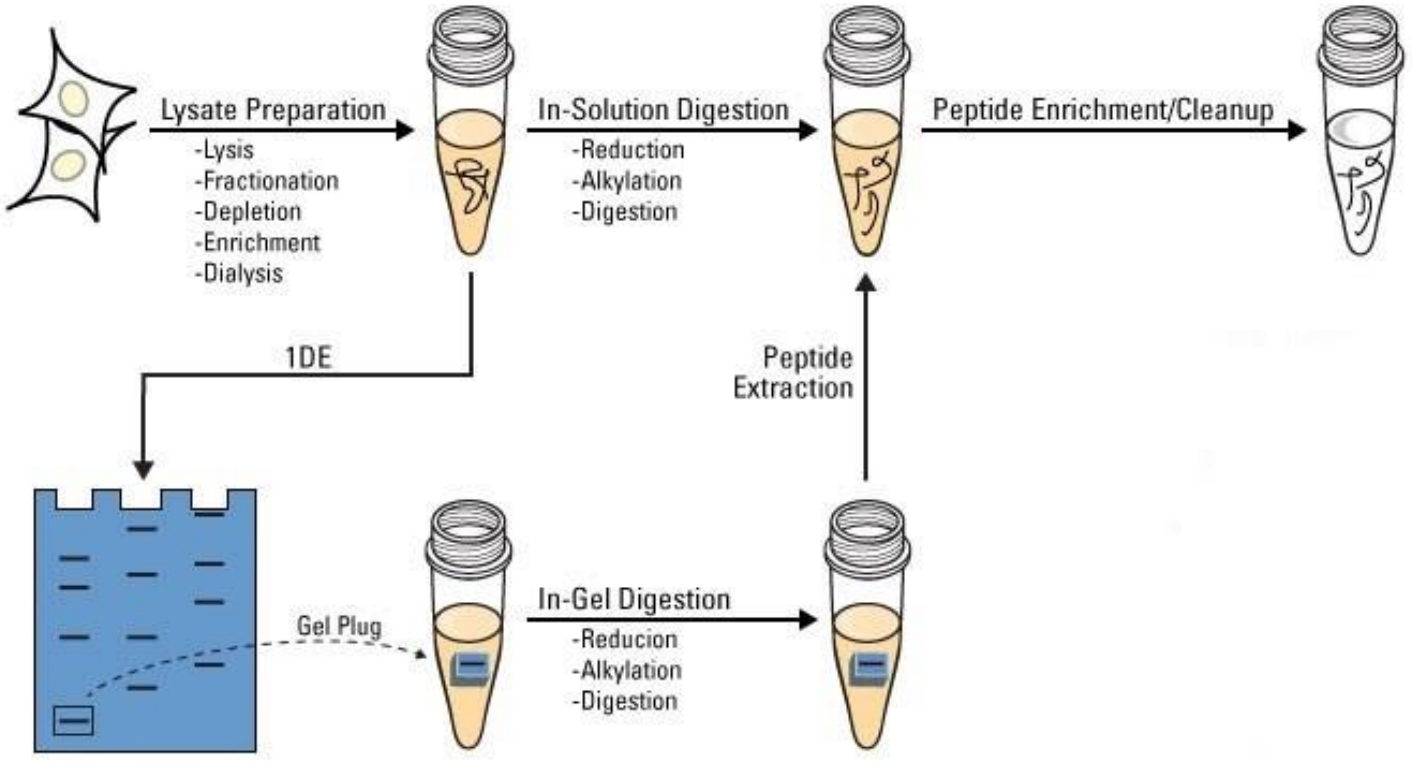
You have received a gel piece in a tube. You want to identify the protein(s) present in the sample. What workflow(s) would you recommend?

□ A top-down approach
□ In-gel digestion
□ A enrichment of the protein of interest
□ Use of chymotrypsin for protein digestion

In-gel digestion

- Whenever you receive a piece of gel, the go-to workflow will always be in gel digestion

- Reduction and alkylation steps are done to reduce disulfide bonds and prevent disulfide bridges from reforming. Increases protein coverage in proteins containing many disulfides



Question 16

Lecture slide(s): 34-39

You have received human blood plasma to analyze. You want to identify the maximal number of protein present in the sample. What workflow(s) would you recommend?

□ A top-down approach
□ A shotgun proteomic workflow
□ Not using mass spectrometry-based workflow
□ Depletion of abundant proteins

- The plasma is a highly complex sample mixture, we cannot utilize the top-down approach

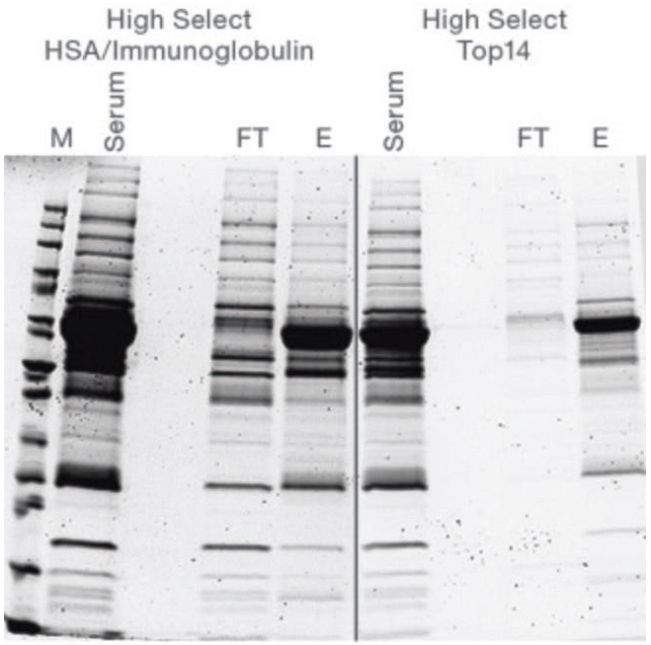
- This is a mass spectrometry course, so we think mass spectrometry is one of the most powerful techniques for chemical analysis, we will not say mass spec would not work for plasma samples ;)

Abundant protein depletion is very important for the analysis of low-abundant proteins

- Remember, there are super abundant proteins such as albumin in our blood. If we don't remove them, we will basically never see our low abundant proteins of interest via MS

- Usually depletion of such protein is done by addition of antibodies which bind to the abundant proteins, and we collect the flow through for shotgun proteomics analysis

Lecture slide(s): 37



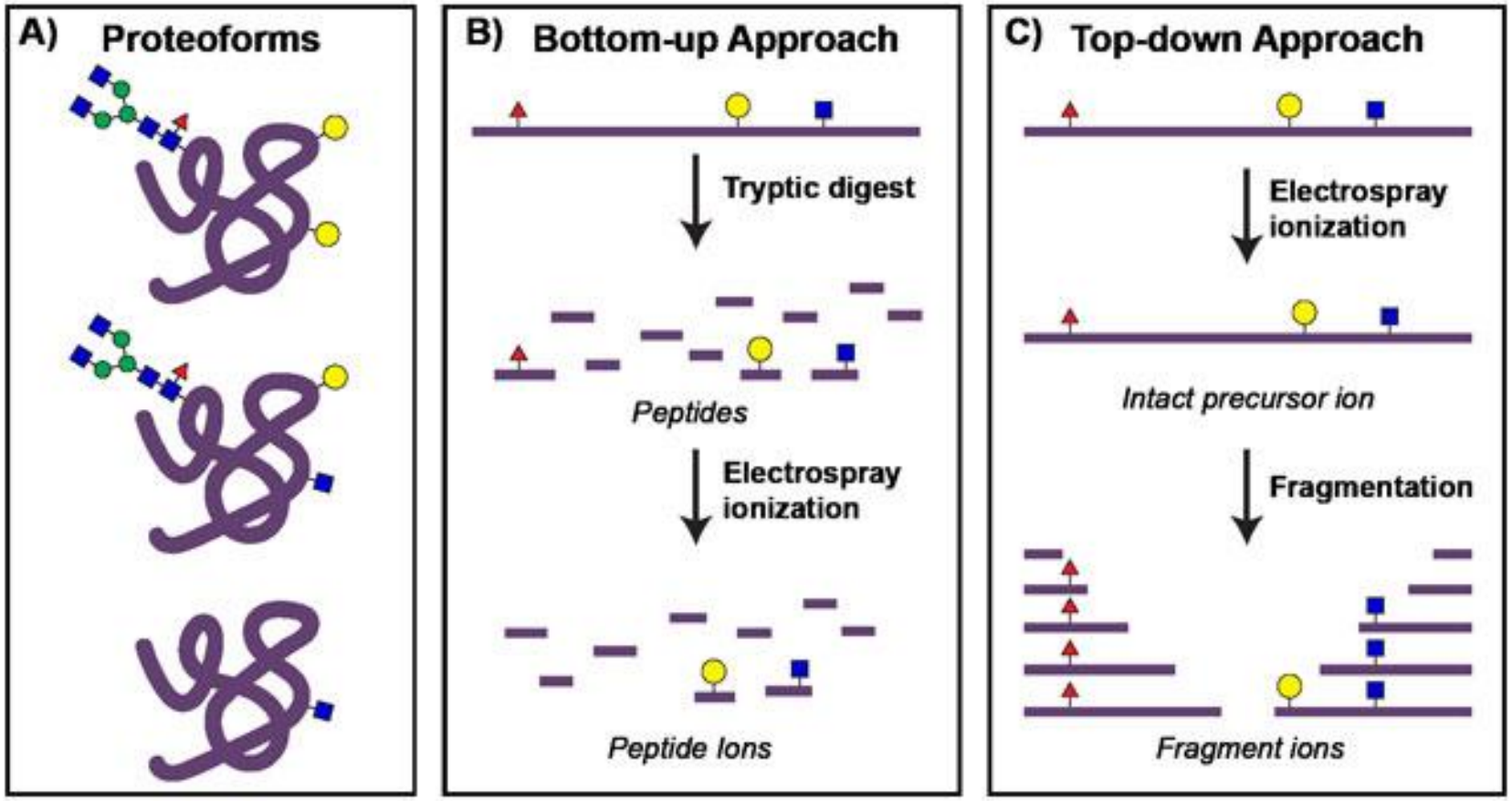
Question 17

You have received a purified protein to analyze. You want to characterize it. What workflow(s) would you recommend?

□ A top-down approach
□ A bottom-up approach
□ Use of different enzymes for protein digestion
□ An initial separation with 1D-gel electrophoresis

- The sample is purified, therefore we can already eliminate separation via electrophoresis
- If you want, you can still do bottom up as you can digest the protein
- If you want to increase sequence coverage, you can use different enzymes to digest this protein

For samples with complex modifications, you can choose to do bottom-up or top-down



Lecture slide(s): 5, 19-20