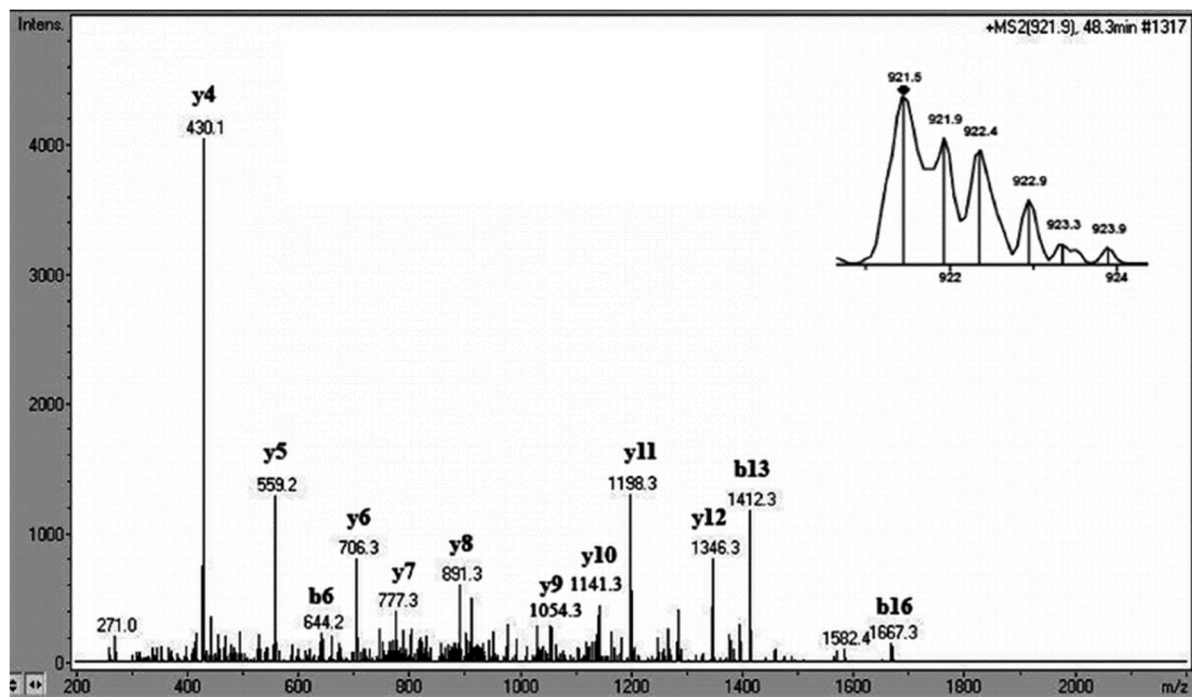


TD module production de données omics – protéomique.

Thibaut Léger – 08/03/2021

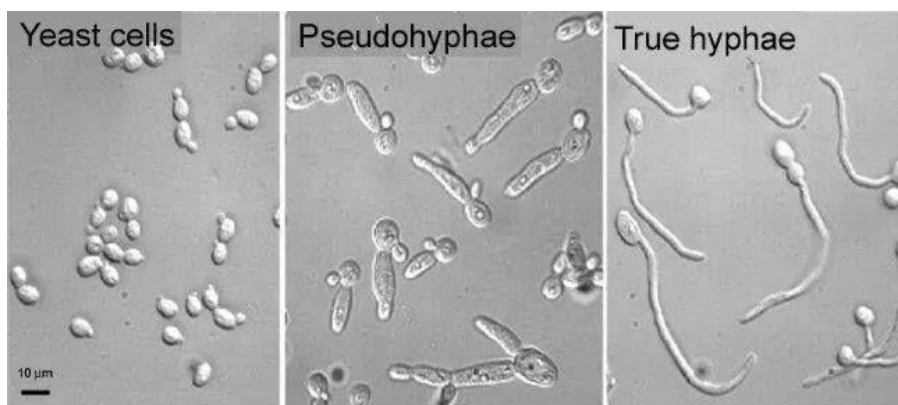
Un run LC-MS/MS a été réalisé à partir d'une digestion par la trypsine d'un lysat total d'un organisme (à déterminer lors du TD). On s'intéresse particulièrement à une protéine. Le spectre MS2 d'un de ses peptides est donné ci-dessous.

- 1) Trouver la séquence du peptide à partir du spectre MS2 (déconvolué) acquis en LC-MS/MS. (masses des acides aminés en annexe ci-dessous).



- 2) Retrouver l'identité de la molécule et l'espèce concernée avec Blastp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>)
- 3) Créer un fichier mgf et retrouver l'identité de la molécule avec Mascot (Matrix Science, http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS).
Rq : Autres solutions : Maxquant et Perseus (<https://maxquant.net/maxquant/>)
- 4) Est-ce que le profil d'abondance des isotopologues correspond au profil théorique obtenu par MIDAs (<https://www.ncbi.nlm.nih.gov/CBBresearch/Yu/midas/index.html>)?

- 5) Trouver des informations sur la protéine et sa séquence de la protéine sur Uniprot (<https://www.uniprot.org/>). Quelle est la localisation intracellulaire de cette protéine ?
 - 6) Réaliser une digestion théorique de la protéine à partir de la séquence en fasta trouvée dans Uniprot avec PeptideMass (https://web.expasy.org/peptide_mass/)
 - 7) Ma protéine est-elle facilement analysable ? -> GMPDB (<https://gpmdb.thegpm.org/>)
 - 8) Est-ce que ma protéine possède des partenaires protéiques connus ? -> String (<https://string-db.org/>), Biogrid (<https://thebiogrid.org/>)
 - 9) Des analyses de protéomique quantitative ont été réalisées avec une autre levure, *Candida albicans*. Ces analyses consistent à comparer les protéomes des formes hyphes (pathogène) et levure (dissémination). L'export de cette analyse de quantification relative est donné en pièce jointe. Il s'agira ici d'identifier les protéines différentielles entre les deux conditions F1 et F2 du fichier et de réaliser des analyses d'enrichissement en termes de Gene Ontology à l'aide de la fonction GO term Finder du site Candida Genome Database (CGD, <http://www.candidagenome.org/cgi-bin/GO/goTermFinder>).
- Ces analyses de termes GO devraient vous permettre de retrouver parmi les échantillons F1 et F2, celui qui correspond à la forme hyphes et celui qui correspond à la forme levure.**
- 10) Pour d'autres espèces, il est possible d'utiliser par exemple SGD (*Saccharomyces cerevisiae*), Panther (<http://www.pantherdb.org/>), Cytoscape avec ses plugins BINGO, CLUEGO ou le logiciel Gene Set Enrichment Analysis (GSEA) pour de nombreuses espèces. Ouvrir Cytoscape 3.7.1 et installer le plugin BINGO. Utiliser les accessions trouvées dans CGD (point 9) et interroger BINGO (utilisation de la database de Gene Ontology de *Candida albicans*)

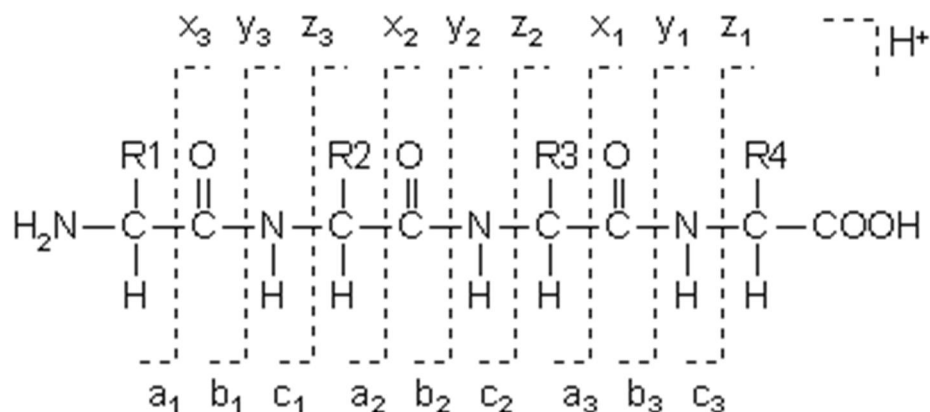


ANNEXES :

A) Masses des acides aminés :

| Name | 3-letter code | 1-letter code | Residue Mass | Immonium ion | Related ions | Composition |
|---------------|---------------|---------------|--------------|--------------|------------------------|---|
| Alanine | Ala | A | 71.03711 | 44 | | C ₃ H ₅ NO |
| Arginine | Arg | R | 156.10111 | 129 | 59,70,73,87,100,112 | C ₆ H ₁₂ N ₄ O |
| Asparagine | Asn | N | 114.04293 | 87 | 70 | C ₄ H ₆ N ₂ O ₂ |
| Aspartic Acid | Asp | D | 115.02694 | 88 | 70 | C ₄ H ₅ NO ₃ |
| Cysteine | Cys | C | 103.00919 | 76 | | C ₃ H ₅ NOS |
| Glutamic Acid | Glu | E | 129.04259 | 102 | | C ₅ H ₇ NO ₃ |
| Glutamine | Gln | Q | 128.05858 | 101 | 56,84,129 | C ₅ H ₉ N ₂ O ₂ |
| Glycine | Gly | G | 57.02146 | 30 | | C ₂ H ₃ NO |
| Histidine | His | H | 137.05891 | 110 | 82,121,123,138,166 | C ₆ H ₇ N ₃ O |
| Isoleucine | Ile | I | 113.08406 | 86 | 44,72 | C ₆ H ₁₁ NO |
| Leucine | Leu | L | 113.08406 | 86 | 44,72 | C ₆ H ₁₁ NO |
| Lysine | Lys | K | 128.09496 | 101 | 70,84,112,129 | C ₆ H ₁₂ N ₂ O |
| Methionine | Met | M | 131.04049 | 104 | 61 | C ₅ H ₉ NOS |
| Phenylalanine | Phe | F | 147.06841 | 120 | 91 | C ₉ H ₉ NO |
| Proline | Pro | P | 97.05276 | 70 | | C ₅ H ₇ NO |
| Serine | Ser | S | 87.03203 | 60 | | C ₃ H ₅ NO ₂ |
| Threonine | Thr | T | 101.04768 | 74 | | C ₄ H ₇ NO ₂ |
| Tryptophan | Trp | W | 186.07931 | 159 | 11,117,130,132,170,100 | C ₁₁ H ₁₀ N ₂ O |
| Tyrosine | Tyr | Y | 163.06333 | 136 | 91,107 | C ₉ H ₉ NO ₂ |
| Valine | Val | V | 99.06841 | 72 | 44,55,69 | C ₅ H ₉ NO |

B) Nomenclature de fragmentation des peptides :





C] Structure d'un fichier mgf :

BEGIN IONS

TITLE=scan=986 centroided 200 most intense (50 total peaks)

RTINSECONDS=297.916

PEPMASS=???? 12066.720502853394

CHARGE=1+

430.1 4000

... (à compléter)

END IONS