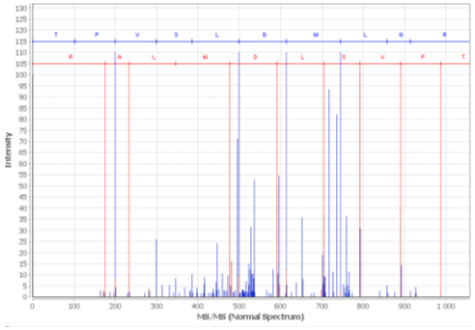
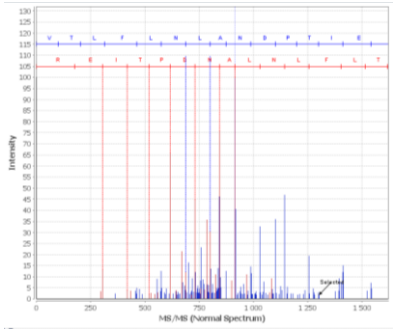
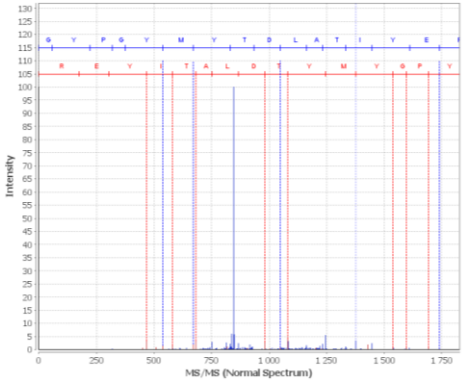
L'identification de protéines séquencées en mode MS/MS : des difficultés encore non résolues.

Dominique Tessier, groupe bioinformatique, plateforme BIBS
Unité BIA, INRA centre Angers-Nantes


Intensity
MS/MS (Normal Spectrum)

Le contexte


Intensity
MS/MS (Normal Spectrum)


Intensity
MS/MS (Normal Spectrum)

4 problématiques majeures en protéomique

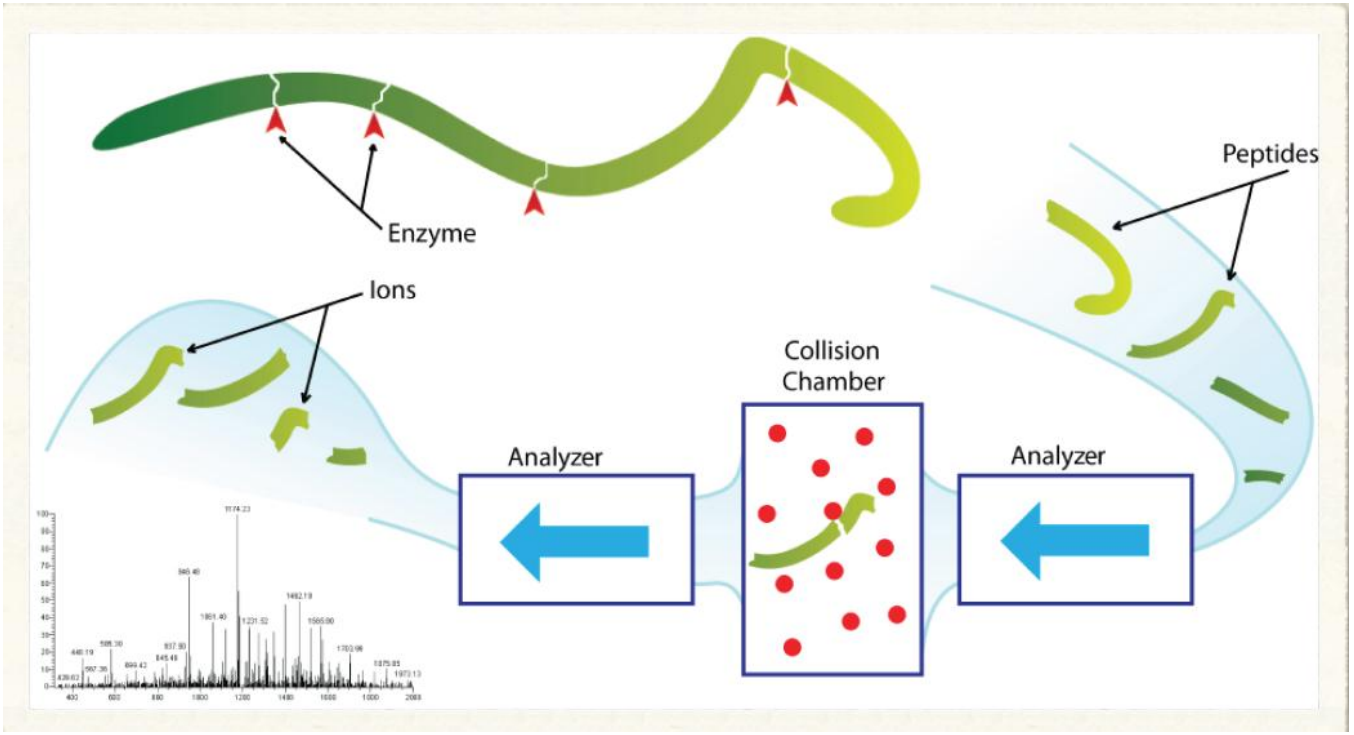
Identification de protéines

Caractérisation de protéines

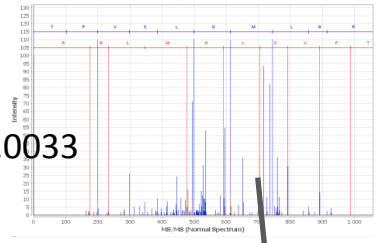
Quantification de protéines

Comparaison d'échantillons

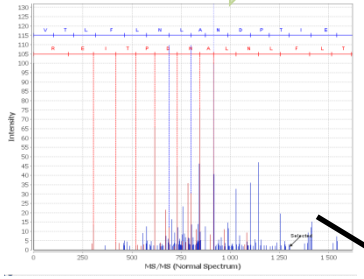
Analyse par spectrométrie de masse en mode MS/MS



Spectres → Peptides → Protéines

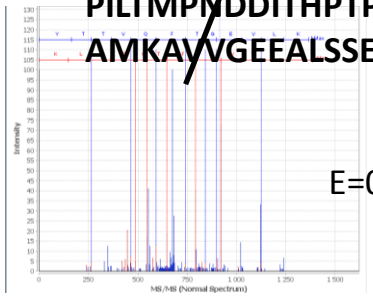


E=0.0033

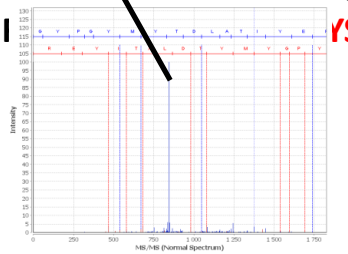


E=0.016

MGLVKDGADMEEGTLEIGMEYRTVSGVAGPLVILDKVKGPKYQEIVNIRLGDGTTRRGQVLEVDGEKAVVQVFEGTSGID
NKYTTVQFTGEVLKTPVSLDMLGRIFNGSGKPIDNGPPILPEAYLDISGSSINPSERTYPEEMIQTGISTIDVMNSIARGQKIP
LFSAAGLPHNEIAAQICRQAGLVKRLEKGKHAEGGGEDDNFAIVFAAMGVNMETAQFFKRDFEENGSMERVTLFLNLAN
DPTIERIITPRIALTTAEYLAYECGKHVLVILTDMSSYADALREVSAAREEVPGRRGYPGYMYTDLATIYERAGRIEGRTGSITQI
PILTMPNDDITHPTPDLTGYITEGQIYIDRQLHNRQIYPPINVLPSLSRLMKSAIGEGMTRRDHSDVSNQLYANYAIGKDVQ
AMKAVVGEEALSSEDLLYLEFLDKFERKFVAQGAYDTRNIFQSLDLAWTLLRII YSRDATH



E=0.00037



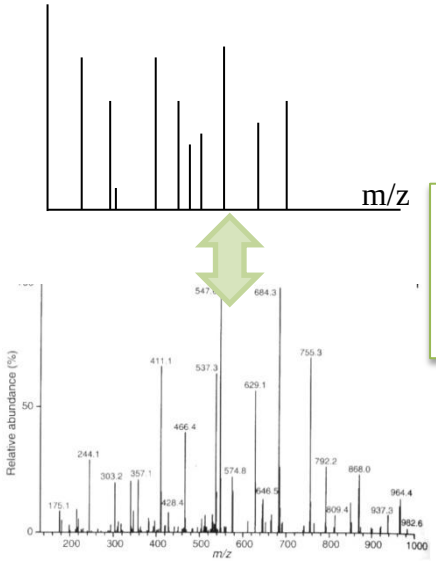
E=1.4^{E}-005

Interprétation : par comparaison…

Protein Database

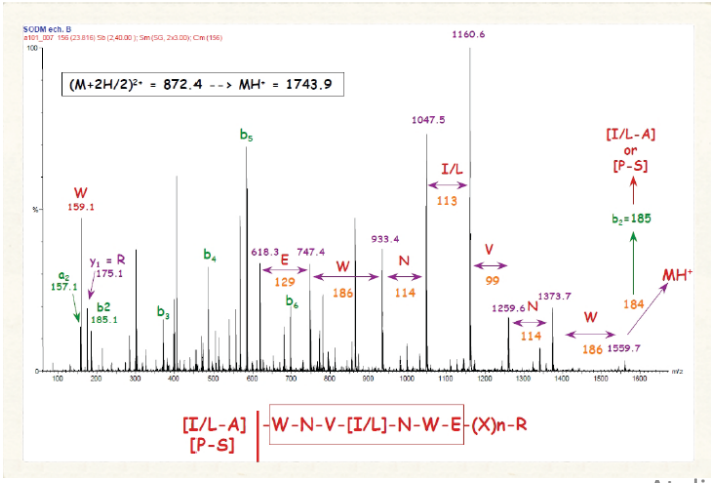
.....
NECFLSHK
DDSPDLPK
WVTFISLLLLFSSAYSR
.....

Digestion *in silico*

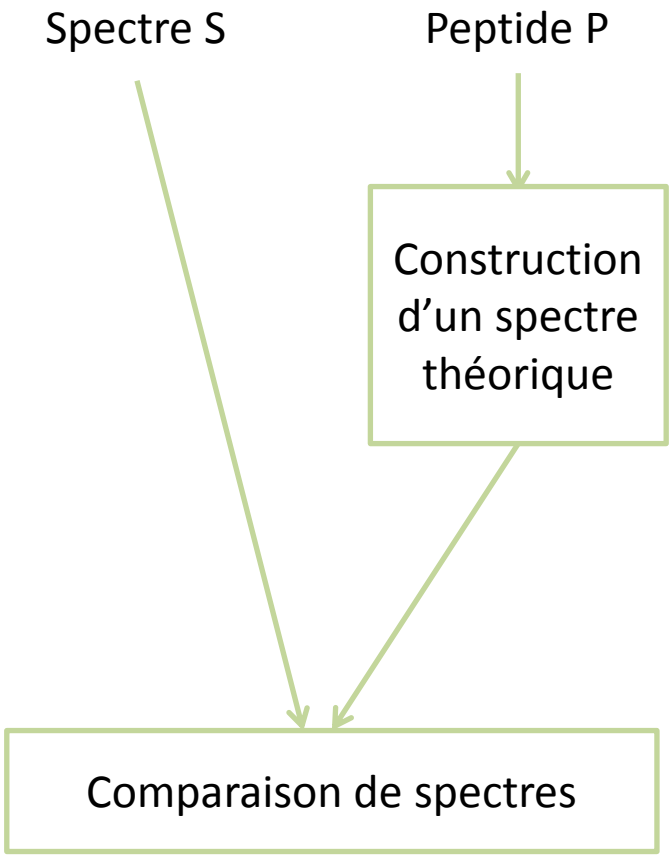


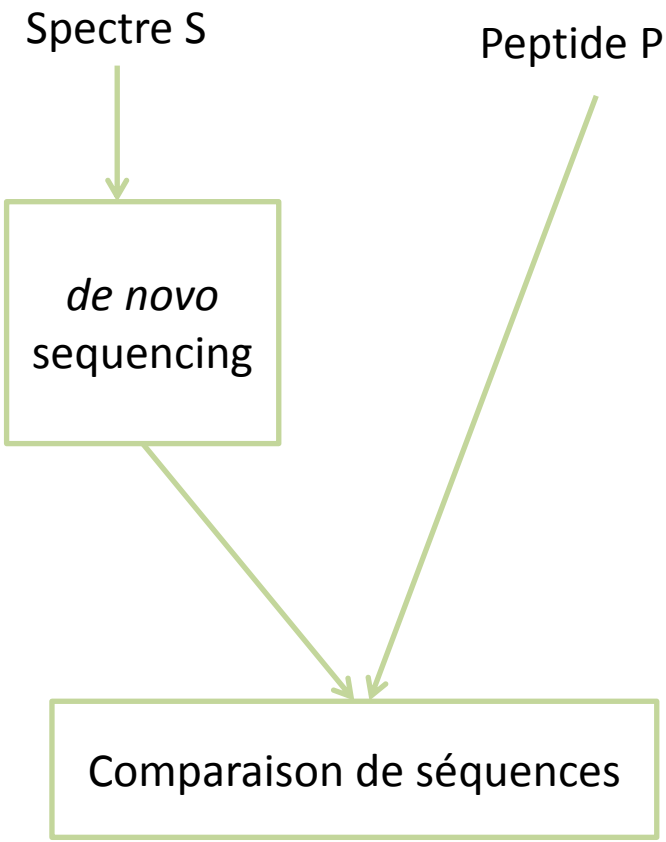
Comparaison entre les spectres théoriques et les spectres expérimentaux

..ou interprétation *de novo*



Interprétation des écarts de masse entre les pics


Spectre S
Peptide P
Construction d'un spectre théorique
Comparaison de spectres


Spectre S
Peptide P
de novo sequencing
Comparaison de séquences

Plan de la présentation

Le contexte

L'association spectre-peptide par approche comparative (Peptide-Spectrum Match)

L'association spectre-peptide par approche *de novo*

L'identification des protéines

La définition de « pipeline » d'analyse

Identifier et caractériser des protéines du point de vue de l'analyse bioinformatique

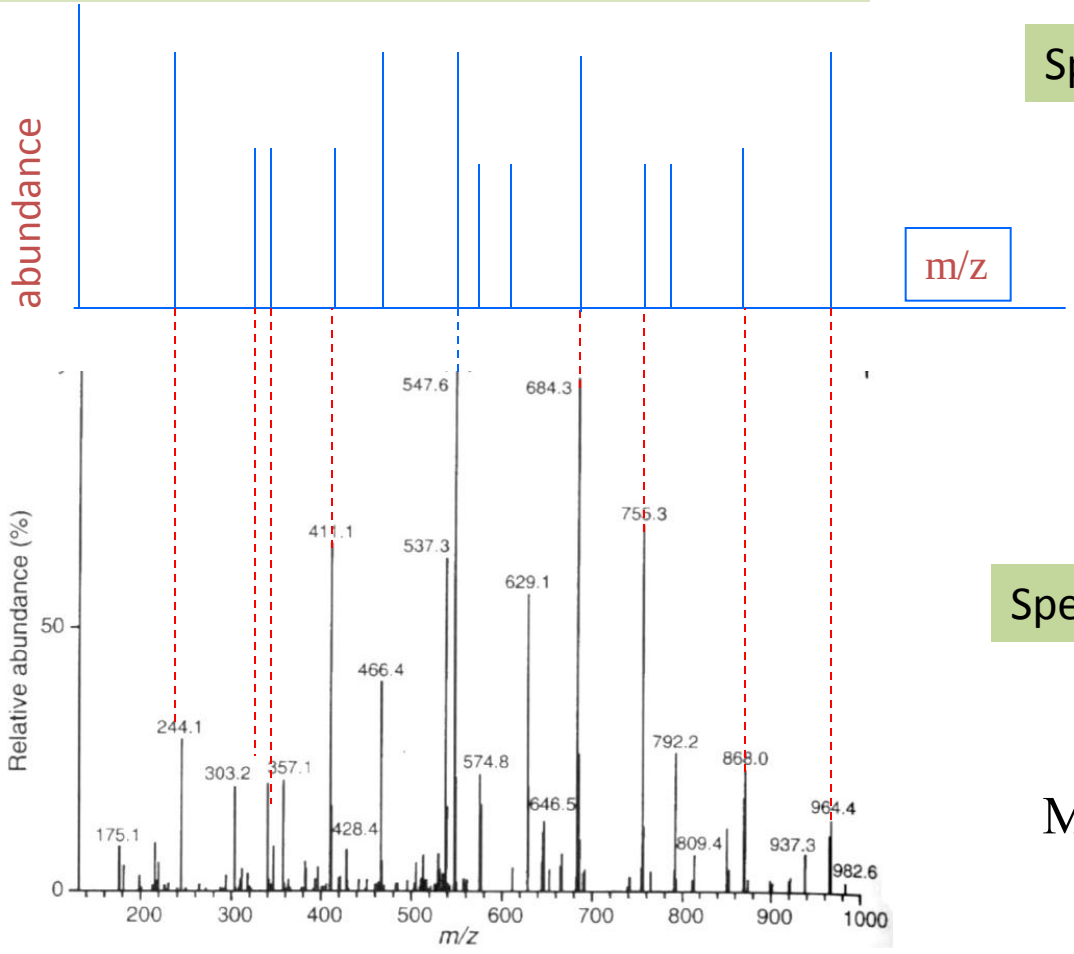
Sélectionner une base de données susceptible de contenir les protéines cherchées.

Comparer les propriétés des protéines cherchées avec les propriétés des protéines de la base de données – les masses des peptides, les propriétés physico-chimiques, l'espèce etc….-

Retenir les meilleures candidates

Calculer les probabilités pour évaluer les chances que les protéines prédites soient les protéines réelles

Comparaison de spectres: Shared peak count



Spectre théorique

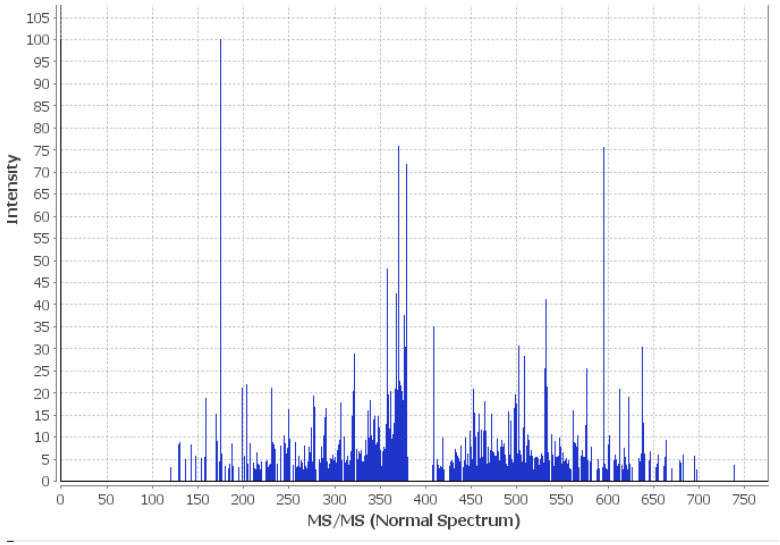
Une fonction de score évalue la similarité entre les 2 spectres

Spectre expérimental

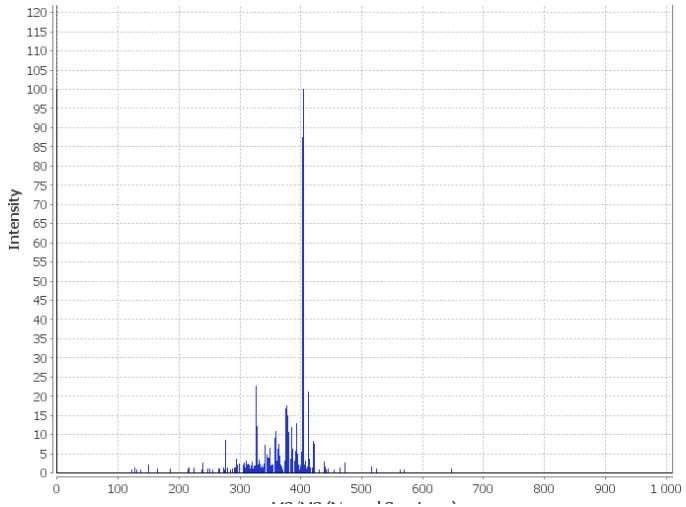
Mascot, X!Tandem, Sequest….

Une analyse en spectrométrie de masse génère des milliers de spectres

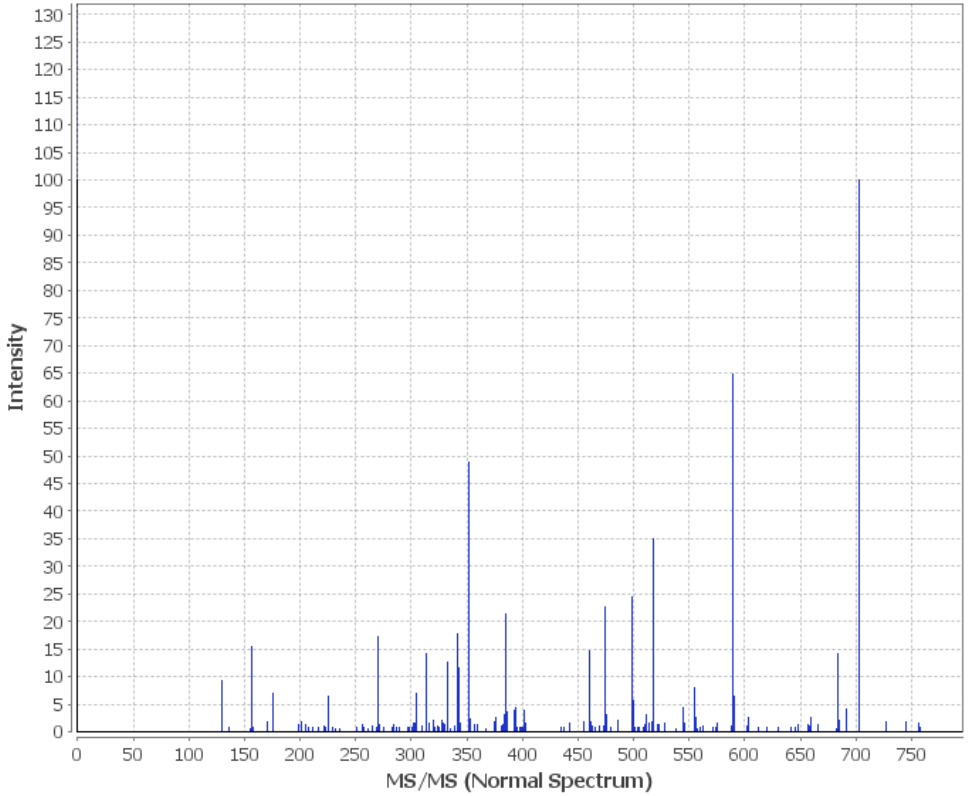
Des spectres très denses



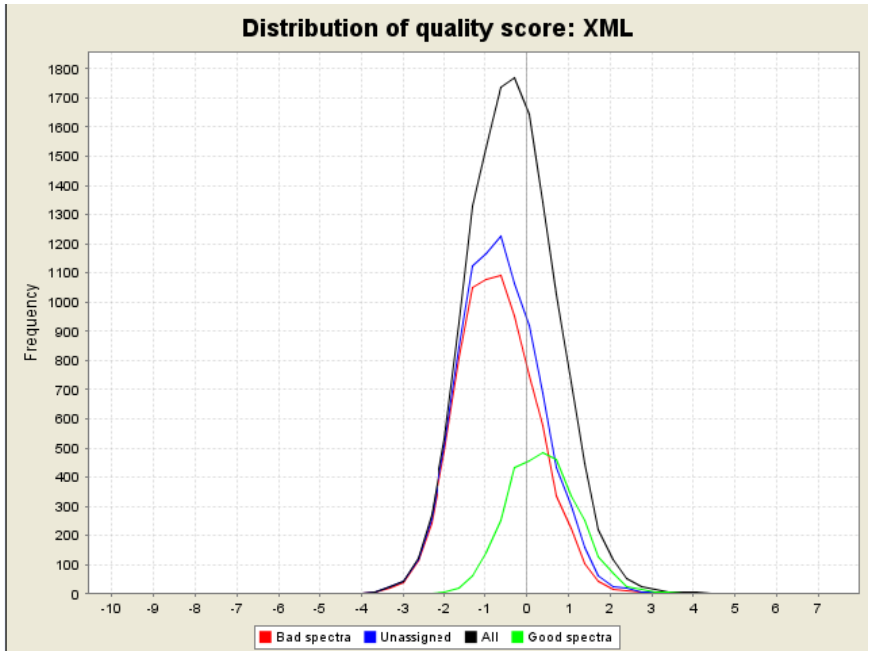
Des spectres peu informatifs



Mais aussi des spectres de bonne qualité:



De nombreux spectres de bonne qualité ne sont pas interprétés.



Qualscore : logiciel d'évaluation de la qualité des spectres

L'interprétation des spectres de masse pour l'identification et la caractérisation des protéines reste un problème ouvert

*Nesvizhskii, A.I., Roos, F.F., Grossmann, J., Vogelzang, M., Eddes, J.S., Gruissem, W., Baginsky, S. and Aebersold, R. (2006) Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides, Mol Cell Proteomics, **5**, 652-670.*

Les raisons du faible taux de reconnaissance de spectres

. Echantillon de mauvaise qualité
. Bruit
. Précision de masse insuffisante

. Des clivages qui ne correspondent pas aux coupures de l'enzyme
. Du polymorphisme de séquence
. Des modifications post-traductionnelles
. Superposition de peptides

On aimerait bien améliorer le taux d'interprétation !

Oui, mais……..

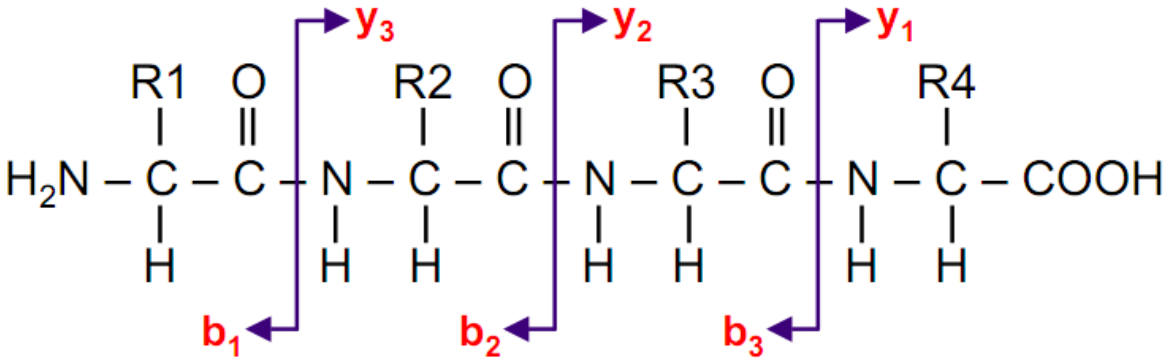
Quel est le coût réel des identifications erronées ?

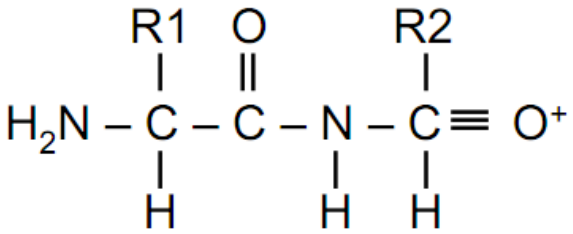
=> opportunités manquées
=> du temps perdu à suivre de fausses pistes

=> la pollution des BD avec des données fausses
=> manque de confiance dans les données de protéomique

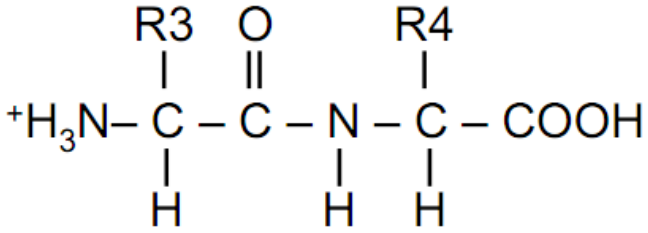
White, F.M. (2011) The potential cost of high- throughput proteomics, *Sci Signal*, **4**, pe8

Fragmentation idéale……

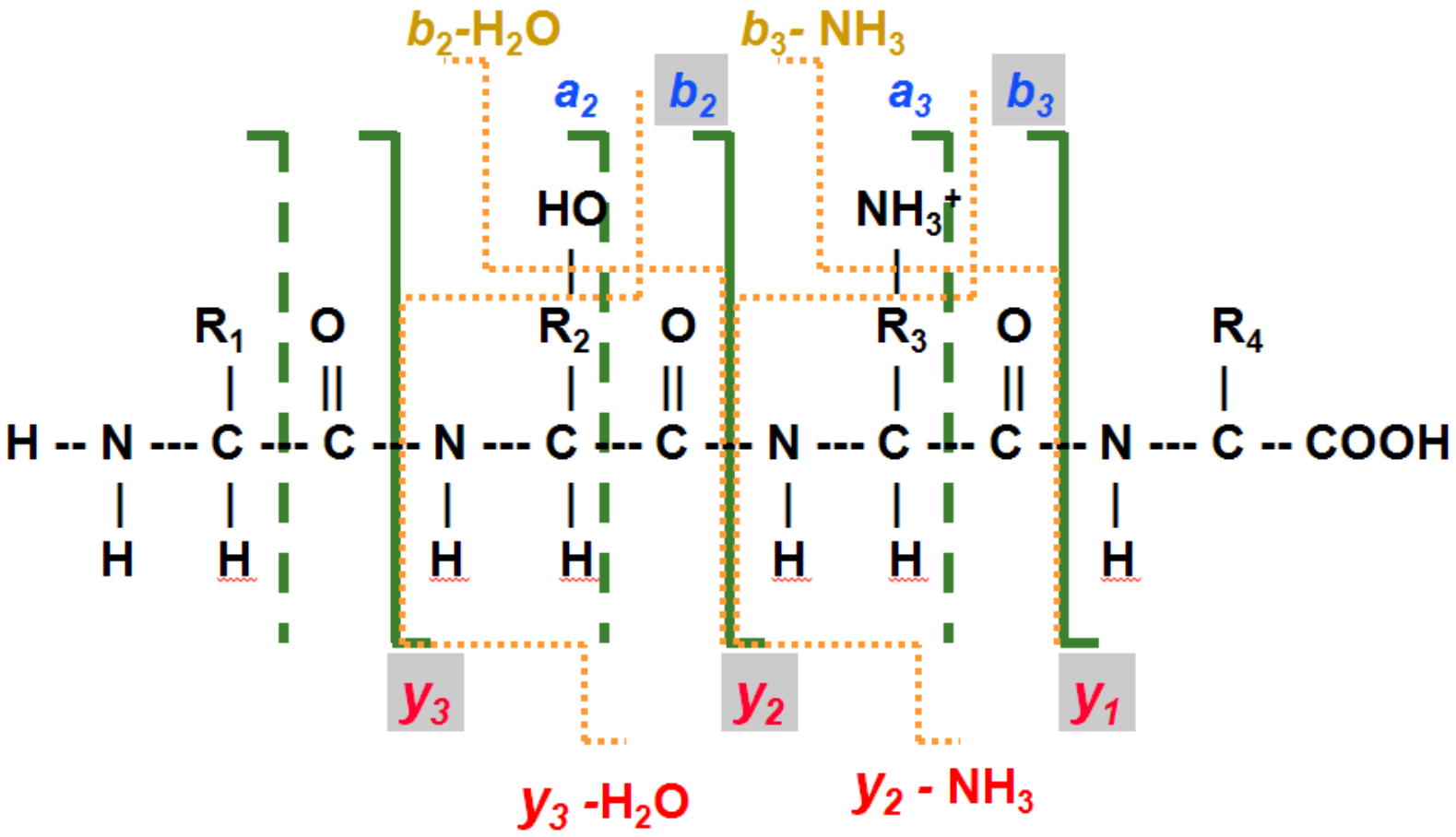




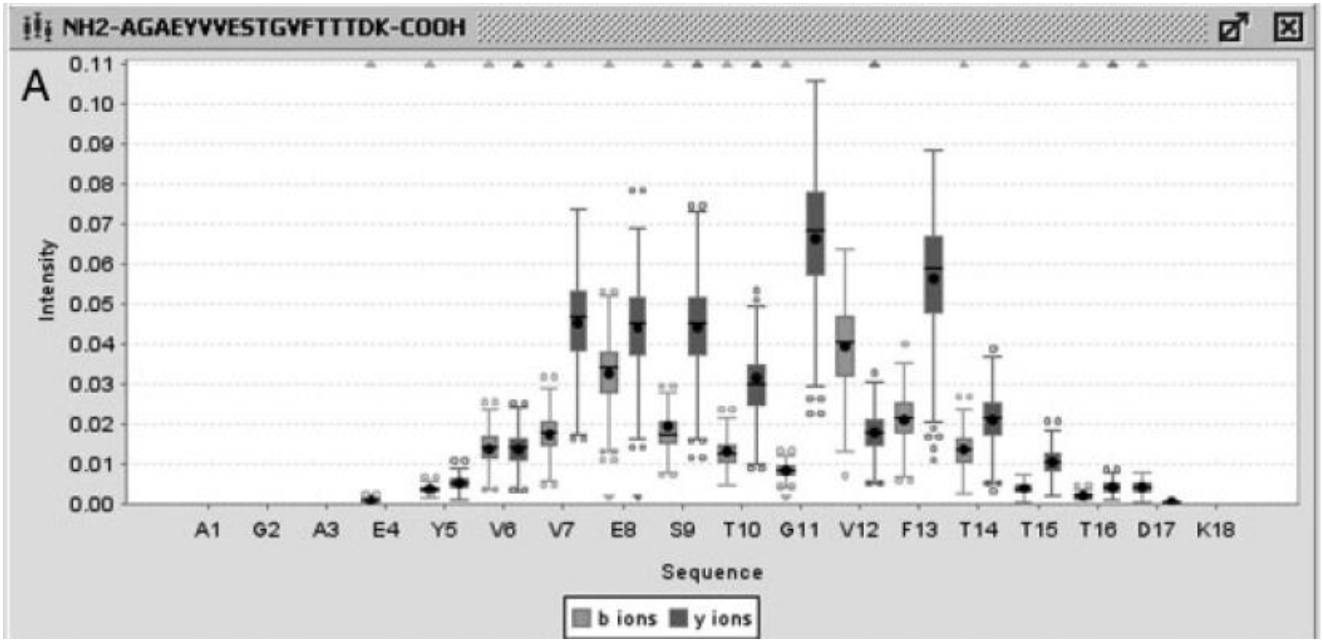
b2



y2

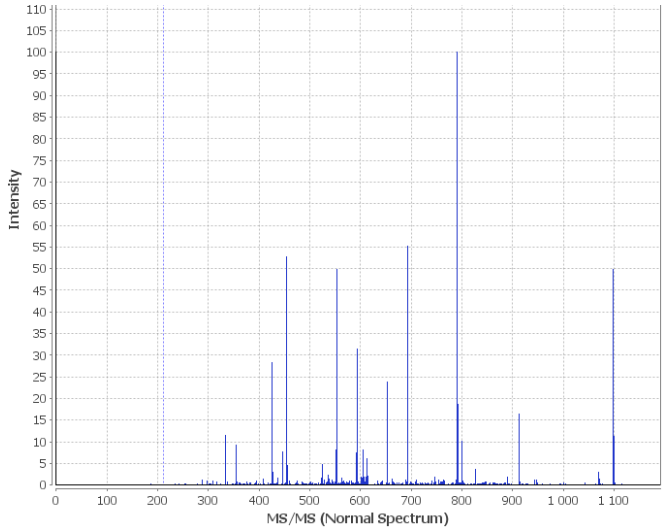

b_2-H_2O
b_3- NH_3
a_2
b_2
a_3
b_3
HO
NH_3^+
R_1
O
R_2
O
R_3
O
R_4
H -- N --- C -- C --- N --- C -- C --- N --- C -- C --- N --- C -- COOH
H
H
H
H
H
H
H
y_3
y_2
y_1
y_3 -H_2O
y_2 - NH_3

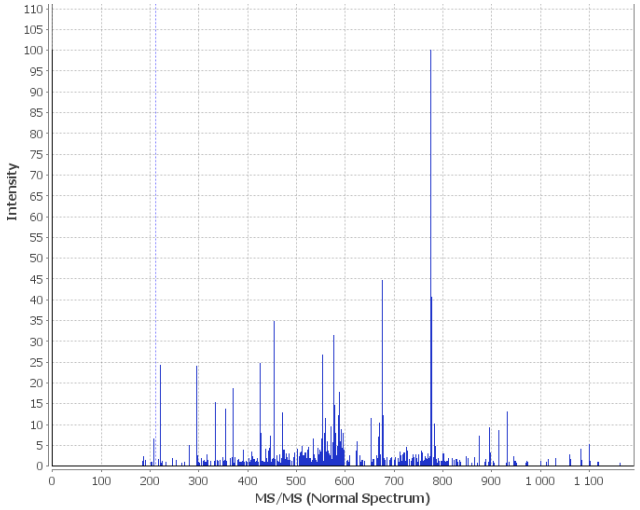
Analyse de la fragmentation de 856 identifications du peptide NH2-AGAEYVVESTGVFTTTDK-COOH – pics correspondants aux ions y et b (Orbitrap)

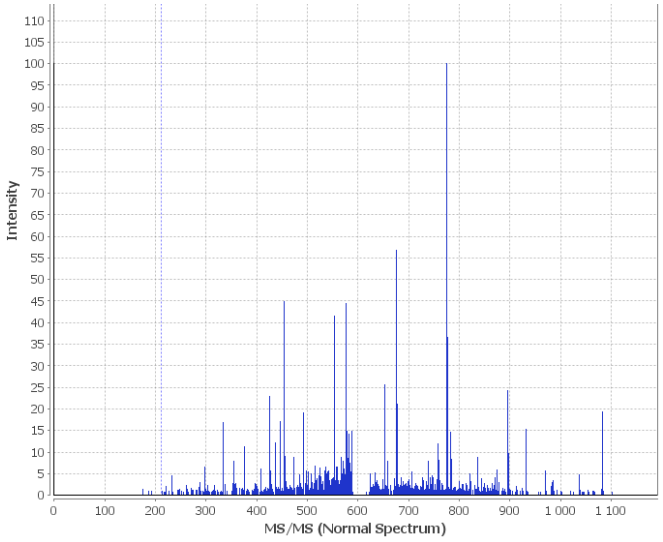


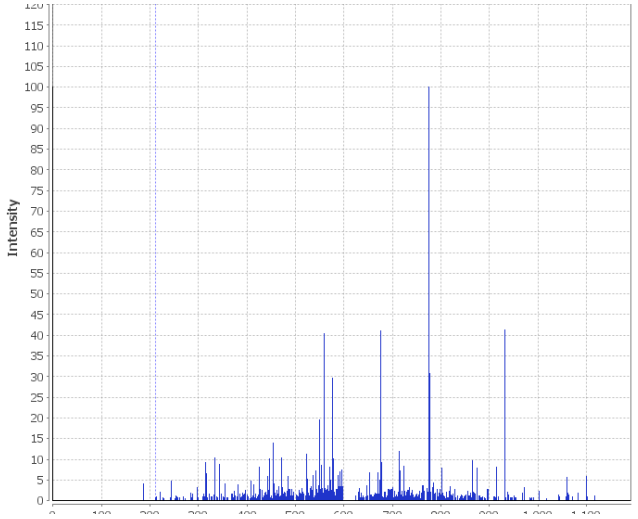
FragmentationAnalyser: An open source tool to analyze MS/MS fragmentation data. H. Barsnes et. Al.. Proteomics 2010

Une assez grande variabilité des données générées







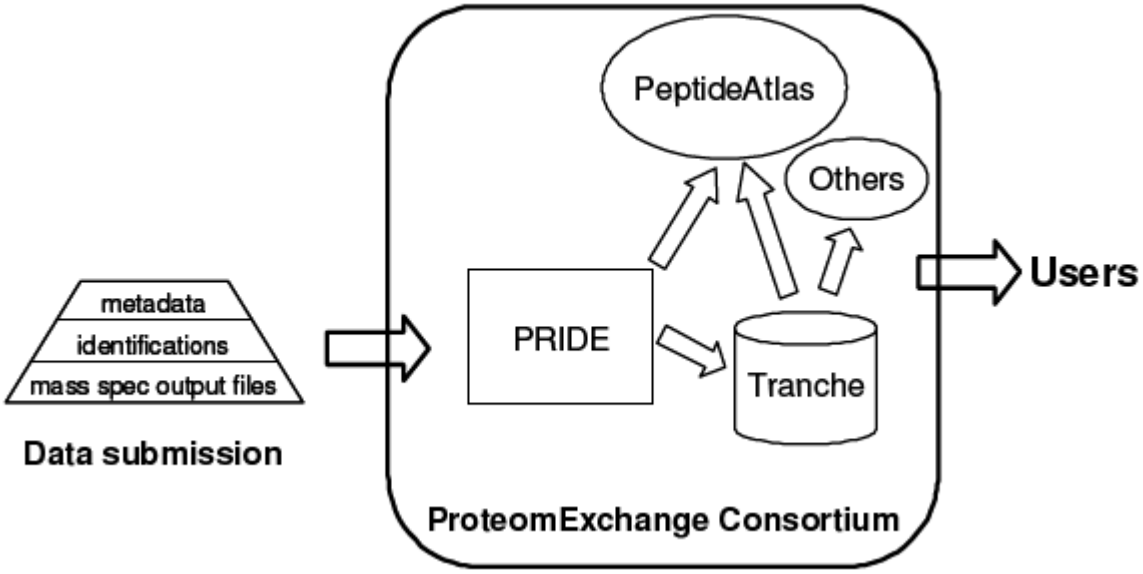


b-ion –H_2O
b-ion –NH_3
y-ion –H_2O
y-ion –NH_3
a
a–H_2O
a –NH_3

bruit, contaminants

4 spectres issus de l'analyse d'un même échantillon représentant de manière sûre le même peptide EGEGVVVMLWK (60% des pics les plus intenses sont affichés, pvalue > 0.99)

Collections non redondantes de spectres obtenus par LC MS/MS



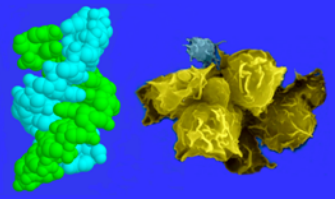
PRIDE Statistics – Novembre 2012

26 589	Experiments
11 380 511	Identified Proteins
64 110 228	Identified Peptides
5 528 099	Unique Peptides
337 882 643	Spectra

The Association of Biomolecular Resource Facilities

Research, Technology, Communication, Education

http://www.abrf.org



ABRF Annual meeting

ABRF 2012

March 17-20

Orlando, FL

The ABRF is an international society dedicated to advancing core and research biotechnology laboratories through research, technology, communication and education. Membership comprises over 700 scientists and administrators, representing over 300 research institutions, who volunteer their efforts on many Research Groups and Committees to provide a unique infrastructure for the benefit of core facilities and their industrial partners.

ABRF Research Groups

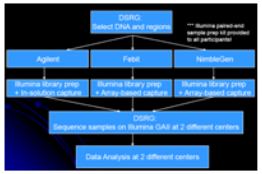
benchmark reagents, instruments, software and individual laboratory performance in response to the interest of the members

- ✓ Antibody Technology (ARG)
- ✓ DNA Sequencing (DSRG)
- ✓ Genomic Variation (GVRG)
- ✓ Glycoprotein (gPRG)
- ✓ Light Microscopy (LMRG)
- ✓ Metabolomics (MRG)
- ✓ MicroArray (MARG)
- ✓ Molecular Interactions (MIRG)
- ✓ Nucleic Acids (NARG)
- ✓ Protein Expression (PERG)
- ✓ Protein Sequencing (PSRG)
- ✓ Proteome Informatics (iPRG)
- ✓ Proteomics (PRG)
- ✓ Proteomics Standards (sPRG)

Standardization and Guidelines

Quantitative microscopy

Benchmark the applications of Next Gen sequencers



Use the ABRF infrastructure to launch your new Research Group, Committee, or Meeting

Core Administrators Network

To promote communication between core personnel and core administrators, and among core administrators. This is coordinated by a new Core Administrators Network – Coordinating Committee (CAN-CC).

ABRF Committee service

is an excellent means of active contribution by our members and a significant networking opportunity

- ✓ Executive Board
- ✓ ABRF Award
- ✓ Affiliates and Chapters
- ✓ Career Development
- ✓ Core Administrators Network - Coordinating Committee
- ✓ Corporate Advisory
- ✓ Corporate Relations
- ✓ Education
- ✓ Finance and Investments
- ✓ Membership
- ✓ Nominations
- ✓ Publications
- ✓ Survey
- ✓ Travel Award
- ✓ Web Site

ABRF Affiliates & Chapters

Infrastructure assistance to regional interest groups to have meetings that promote common interests. This is facilitated by a new Affiliates and Chapters Committee (ACC)

MWACD Midwest Association of Core Directors

C-CAMP

MCBIOS.ORG

Association of Shared Resource Directors

ASSOCIATION CANADIENNE DE CYTOMÉTRIE CANADIAN CYTOMETRY ASSOCIATION

NERLSCD

Sponsors

Advanced Analytical · Agilent Technologies · Fox Chase Cancer Center

illumina · Roche · Memorial Sloan-Kettering Cancer Center

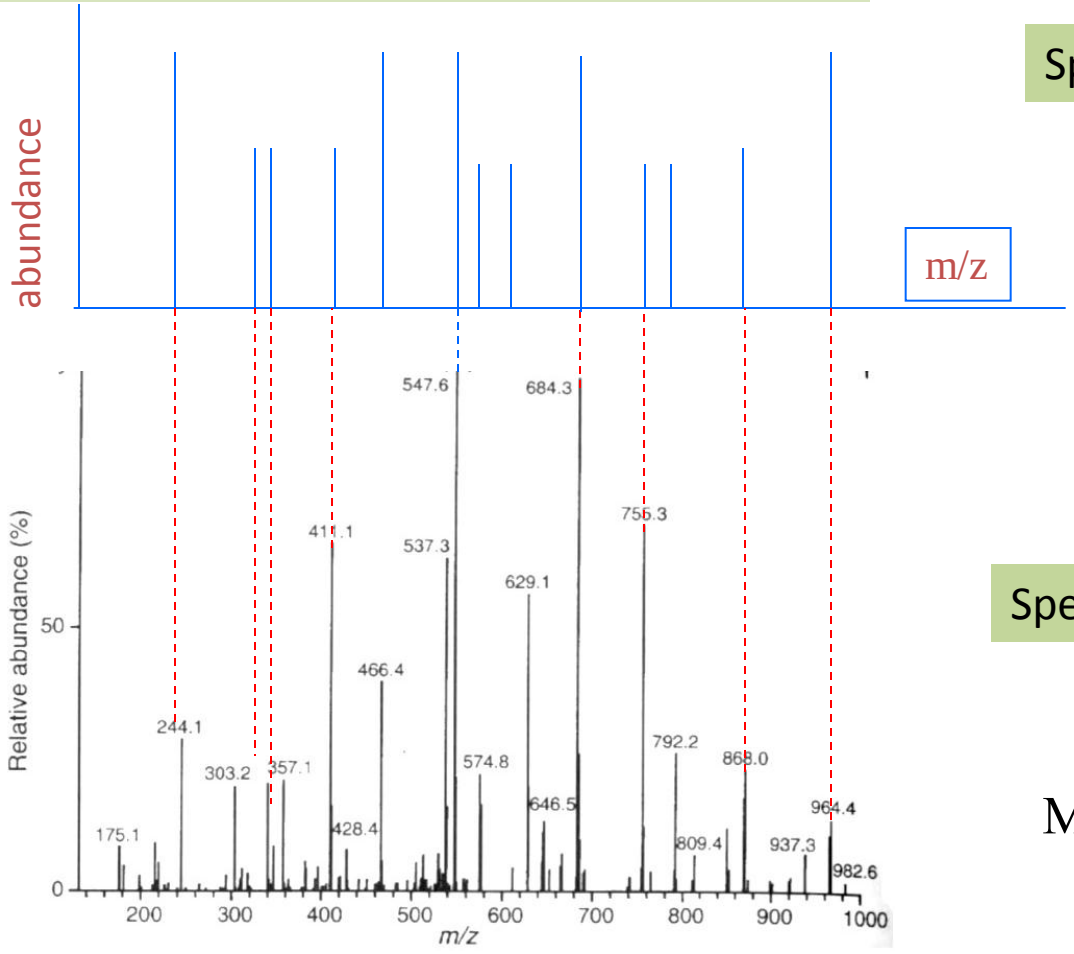
Ingenuity Systems · Caliper LifeSciences · UF ICBR

integenX · GE Healthcare · New York University

Leica Microsystems · Promega · University of Nebraska Medical Center

Waters · Vanderbilt University Medical Center · Delaware Biotechnology Institute

Comparaison de spectres: Shared peak count



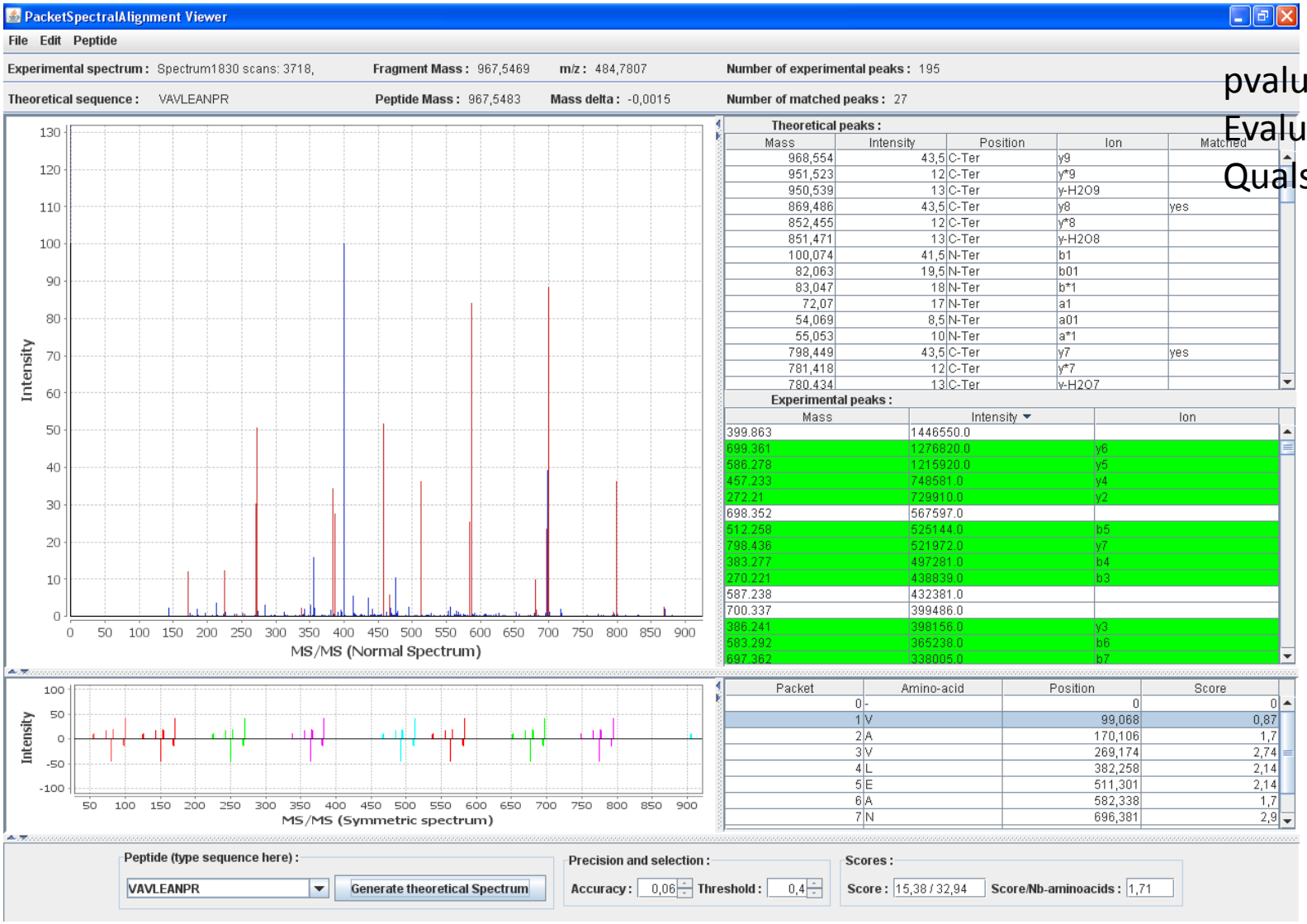
Spectre théorique

Une fonction de score évalue la similarité entre les 2 spectres

Spectre expérimental

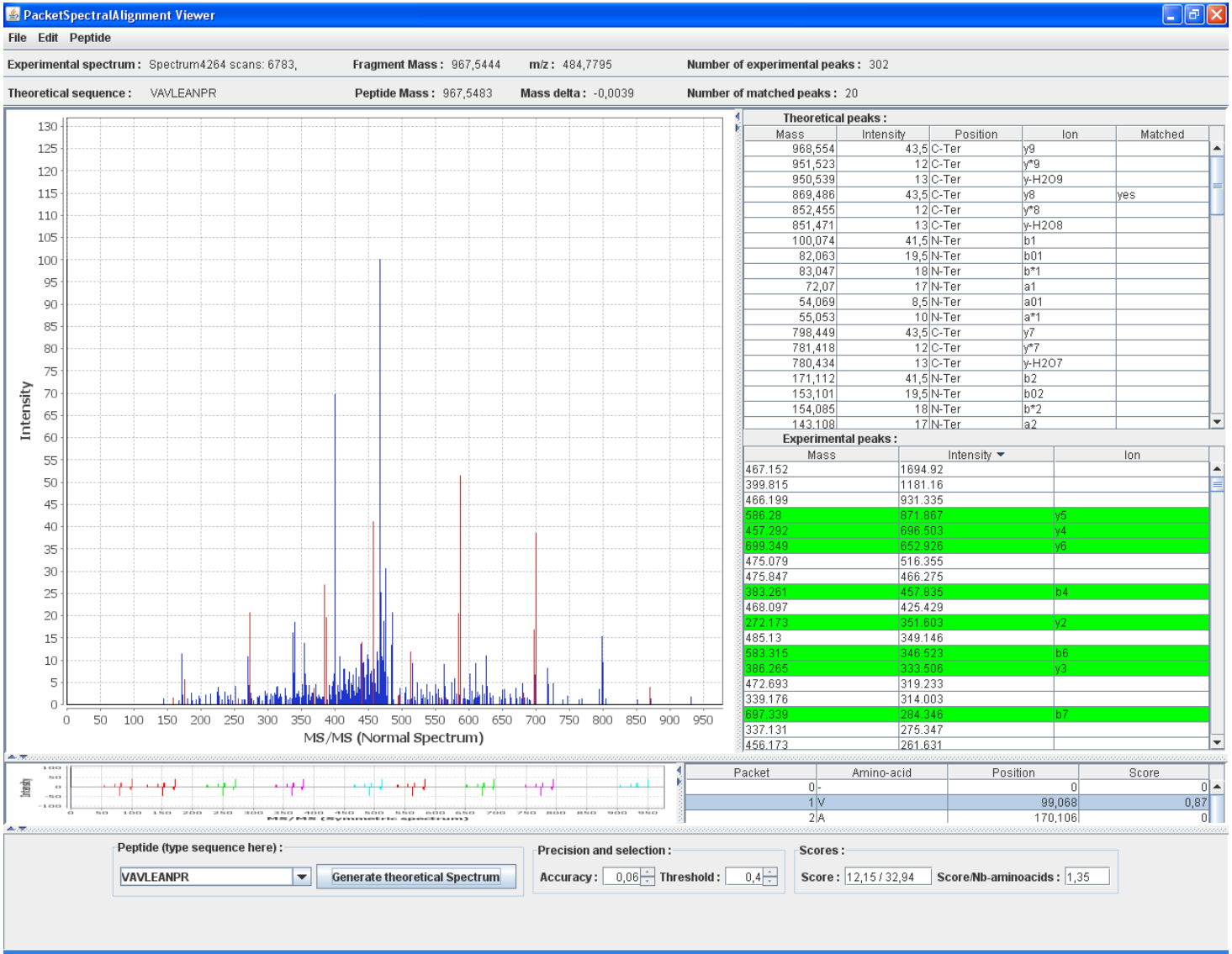
Mascot, X!Tandem, Sequest….

Exemple: un spectre de bonne qualité (qualscore > 3) bien interprété



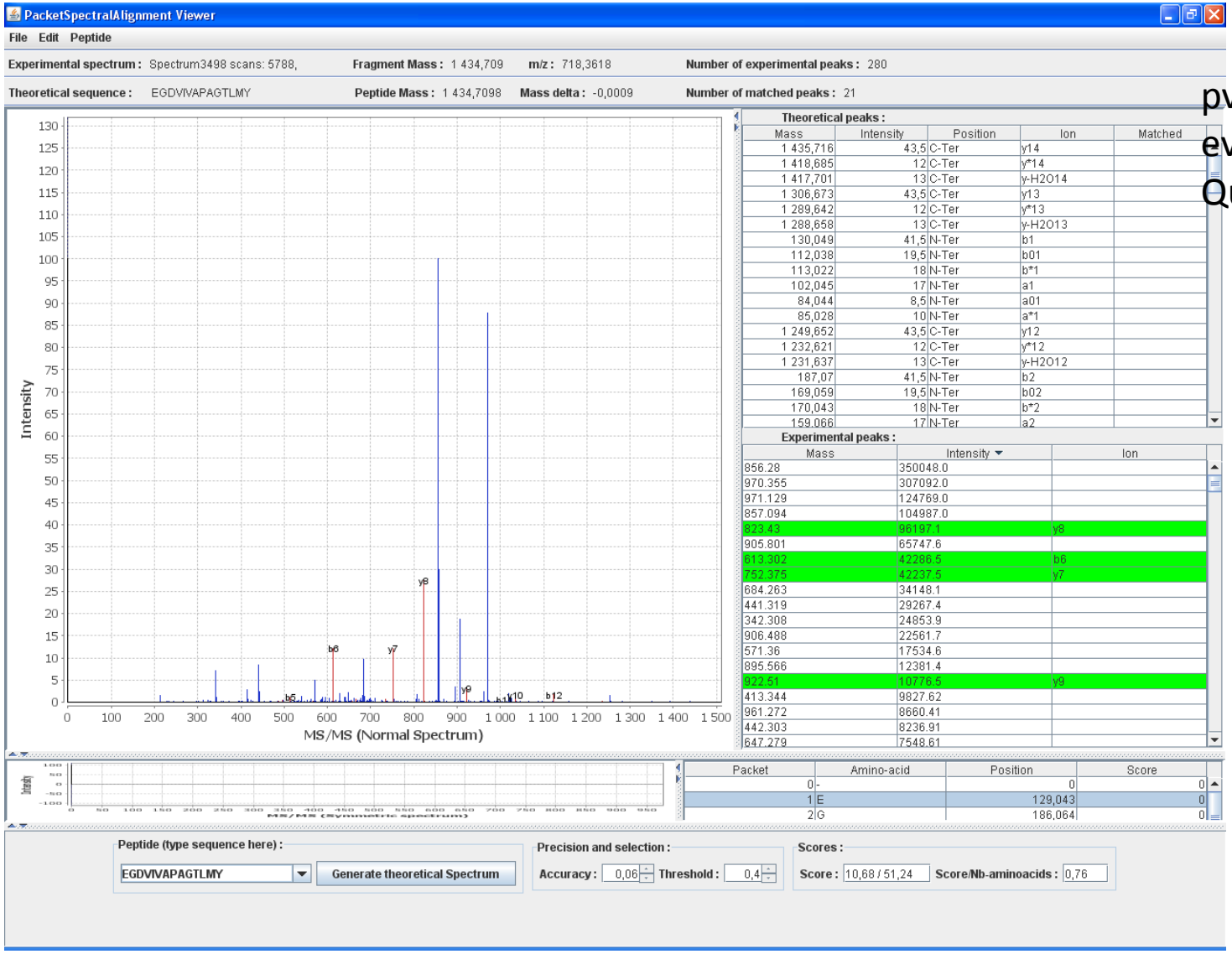
pvalue=0.9983
Evalue=0.047
Qualscore=3.55

Exemple : un spectre bien interprété pourtant de mauvaise qualité



pvalue=0.9924
evalue=0.091
Qualscore=-1.09

Un spectre de bonne qualité et une interprétation incertaine

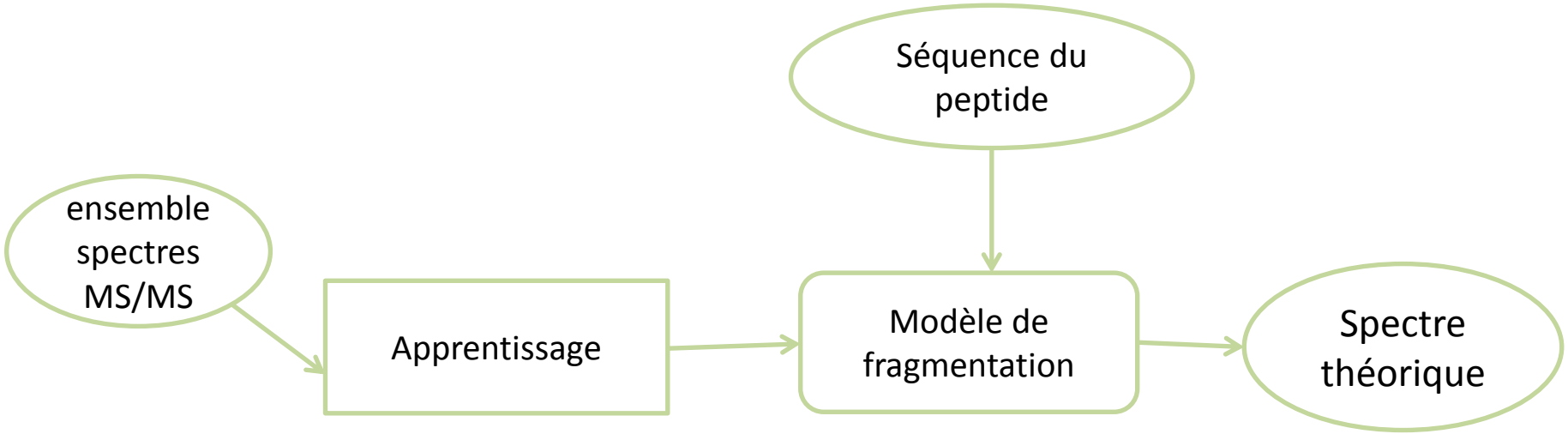


pvalue=0.2626
evalue=0.46
Qualscore=4.35

Les algorithmes de comparaison de spectres :

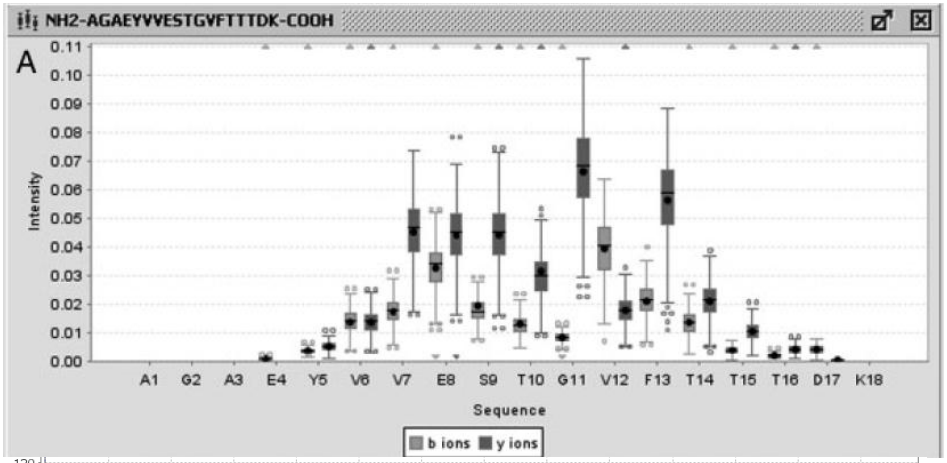
Quels sont les types d'ions considérés ?
Comment sont définies les intensités des spectres théoriques ? *
Quel est l'algorithme de comparaison utilisé ?
Quel est le calcul de score ? *
Comment les mutations , les modifications post-traductionnelles sont-elles prises en compte ? *
Comment sont validées les attributions ?

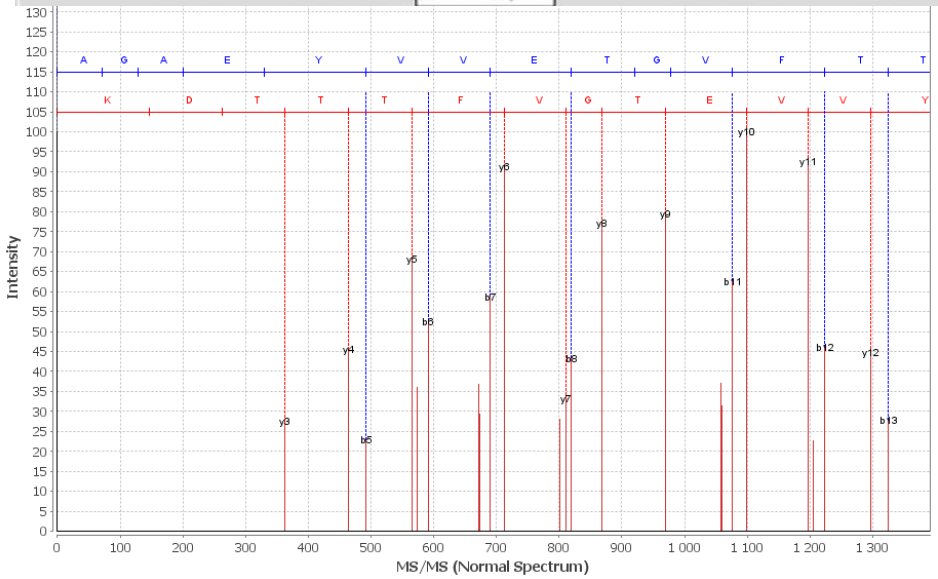
Mieux modéliser un spectre théorique = prédire le modèle de fragmentation



Exemple PepNovo+ : la prédiction ne porte pas sur l'intensité des pics, mais sur l'ordre de l'intensité des pics. Basé sur une méthode RankBoosting

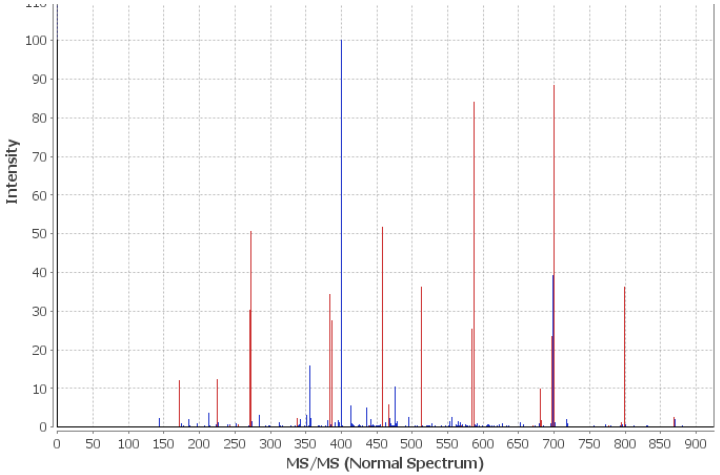
Predicting Intensity Ranks of Peptide Fragment Ions. Frank, A.M. J. Proteome Research, 8:2226-2240, 2009
Ranking-Based Scoring Models for Peptide-Spectrum Matches. Frank, A.M. J. Proteome Research, 8:2241-2252, 2009



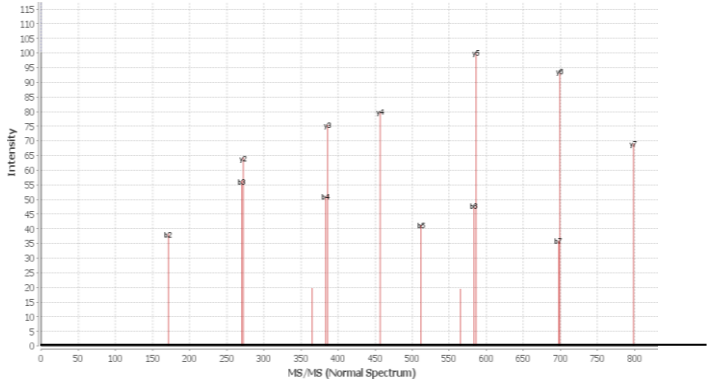


Peptide AGAEYVVETGVFTTTDK

Fragmentation prédite par Pepnovo+



Peptide VAVLEANPR



Fragmentation prédite par PepNovo+

Banques de spectres expérimentaux

⇒ Des scores plus discriminants qu'en comparaison avec des spectres théoriques

⇒ Un espace de recherche plus petit et plus précis

⇒ Identification plus facile de spectres sortant de l'ordinaire donc potentiellement porteur de modifications post-traductionnelles

⇒ L'intérêt de ces banques décroît lorsque les spectres sont de bonnes qualité

NIST : http://peptide.nist.gov/

PeptideAtlas : http://www.peptideatlas.org/speclib/

PRIDE : http://www.ebi.ac.uk/pride/

Les algorithmes de comparaison de spectres :

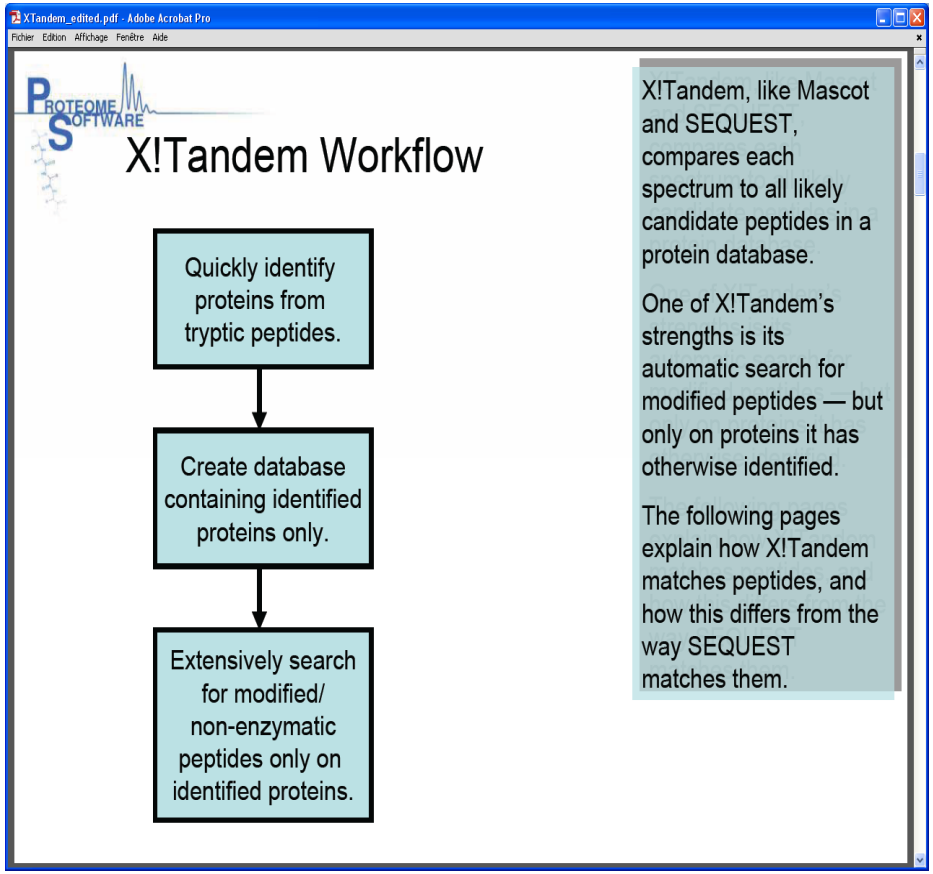
Quels sont les types d'ions considérés ?
Comment sont définies les intensités des spectres théoriques ? *
Quel est l'algorithme de comparaison utilisé ?
Comment le score est-il calculé ? *
Comment les mutations , les modifications post-traductionnelles sont-elles prises en compte ?
Comment sont validées les attributions ?

Le calcul de score : un exemple simple : X!Tandem

Craig, R. and Beavis, R.C. (2003) A method for reducing the time required to match protein sequences with tandem mass spectra, *Rapid Commun Mass Spectrom*, **17**, 2310-2316.

An explanation of the X!Tandem MS/MS spectra search program developed by Craig,R. and Beavis,R.C. (2003) *Rapid Commun. Mass Spectrom.*, **17**, 2310–2316.

Explication complète : www.proteomesoftware.com



Principaux avantages de X!Tandem

. Très rapide
. Prend en compte les peptides semi-tryptic
. Prend en compte le polymorphisme
. Score basé sur des calculs de probabilité

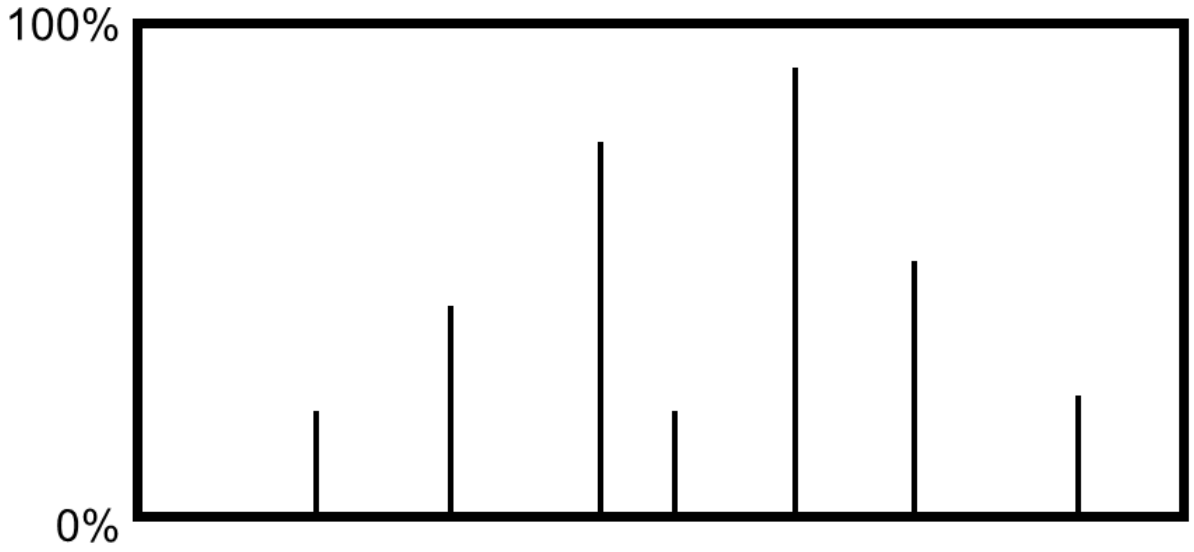
Hyperscore

$$HyperScore = \left(\sum_{i=0}^{n} I_i * P_i \right) * N_b! * N_y!$$

spectrum intensities

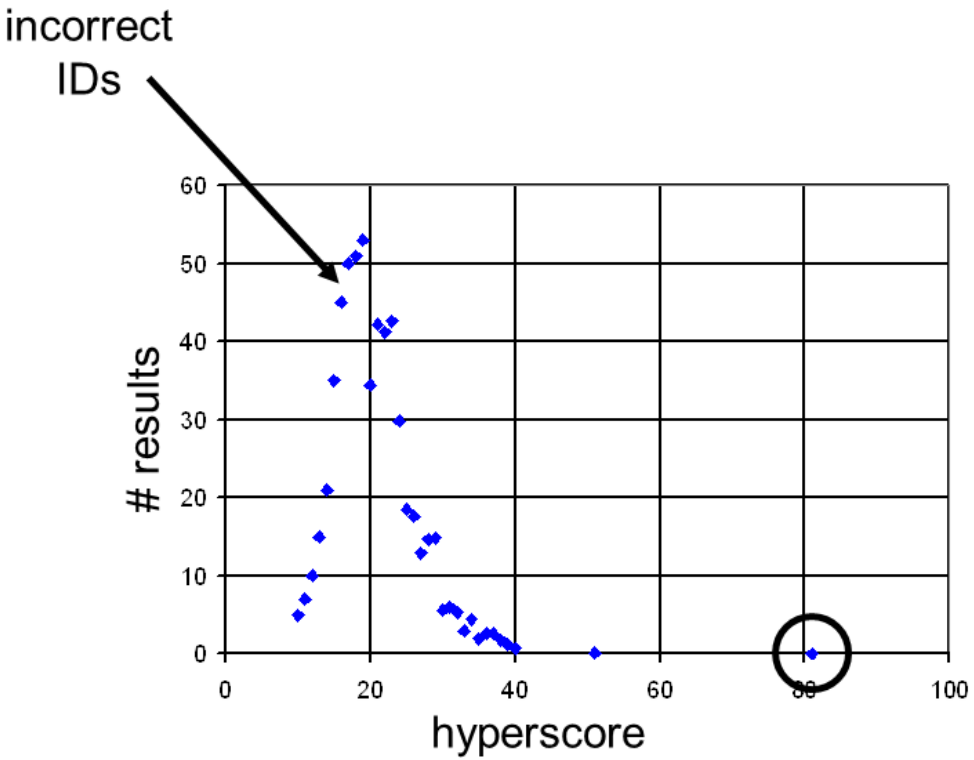
predicted? (1,0)

X!Tandem modifies the preliminary score by multiplying by N factorial for the number of b and y ions assigned. The use of factorials is based on the hypergeometric distribution.



similar peaks (y/b ions)

Histogram of hyperscores

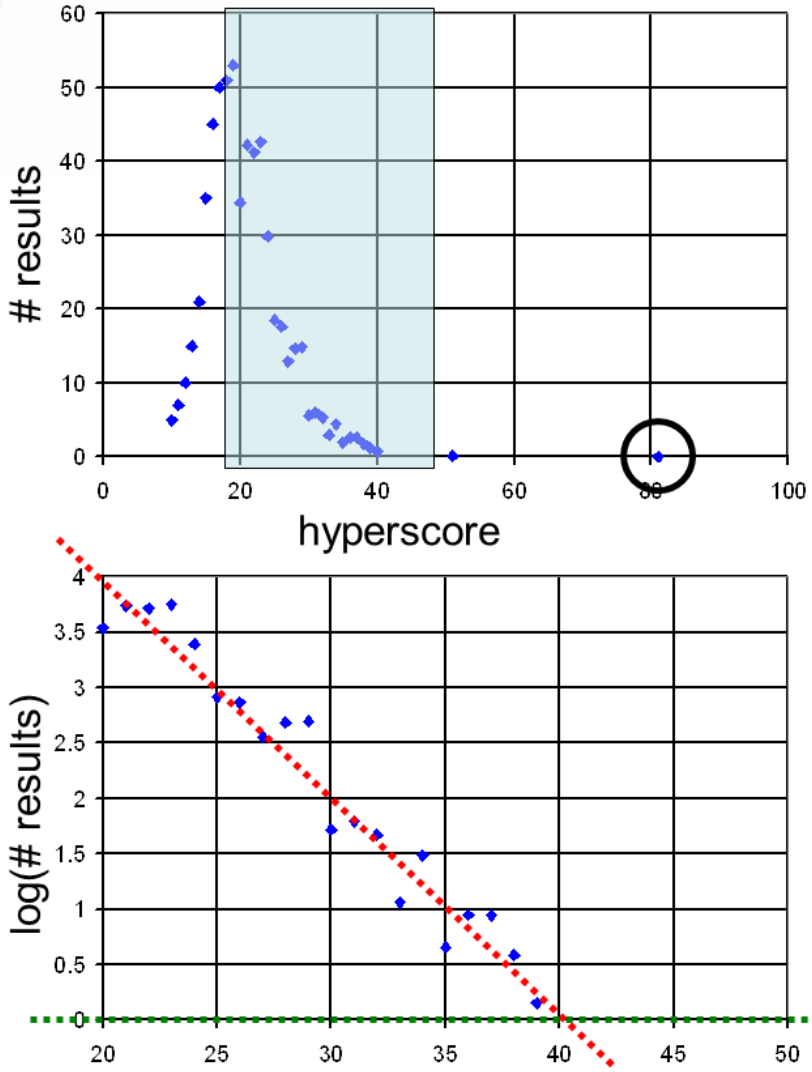


Next, X!Tandem makes a histogram of all the hyperscores for all the peptides in the database that might match this spectrum.

For example, in this figure, 52 peptides were found with a hyperscore of 19, and one peptide with a hyperscore of 83.

X!Tandem assumes that the peptide with the highest hyperscore is correct, and all others are incorrect.

Log histogram

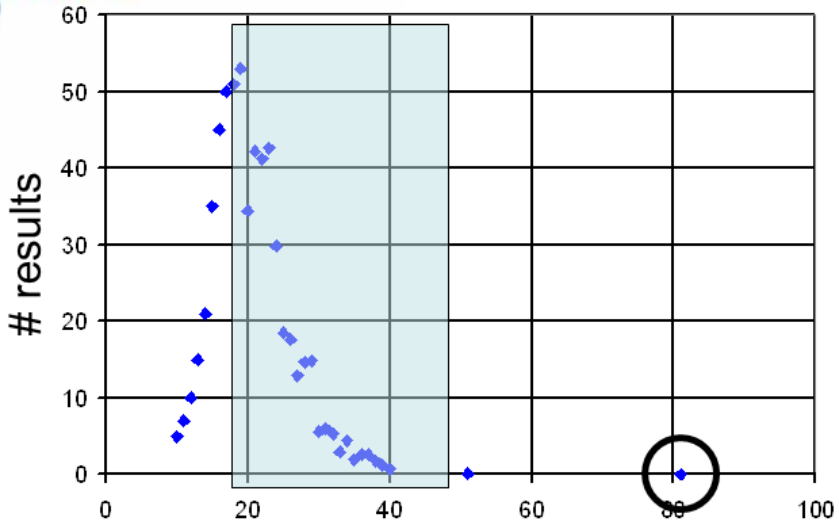


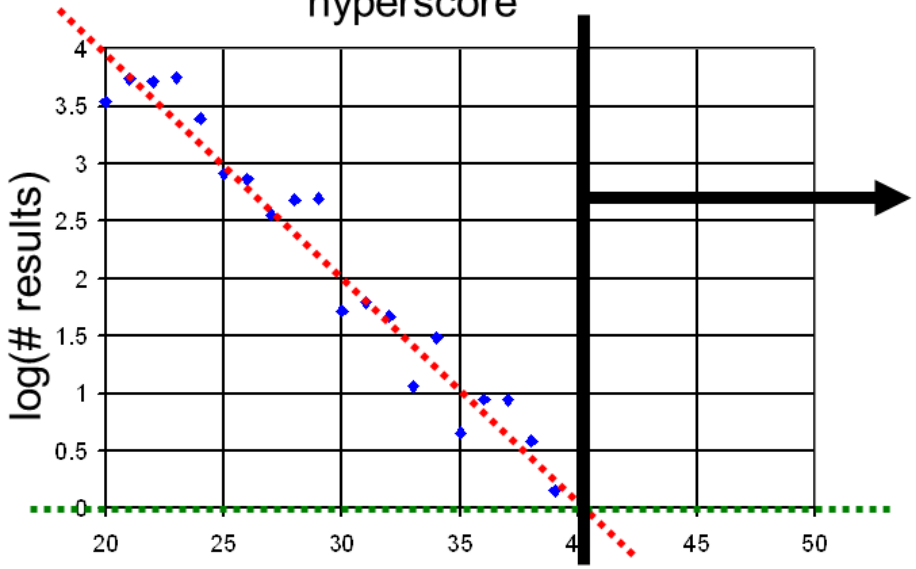
If the data on the right side of the histogram, (colored in upper figure) is taken and log-transformed, the data fall on a straight line.

A straight line is the expected result from a statistical argument that assumes the incorrect results are random.

Note: this histogram is calculated independently for each spectrum.

Significant scores





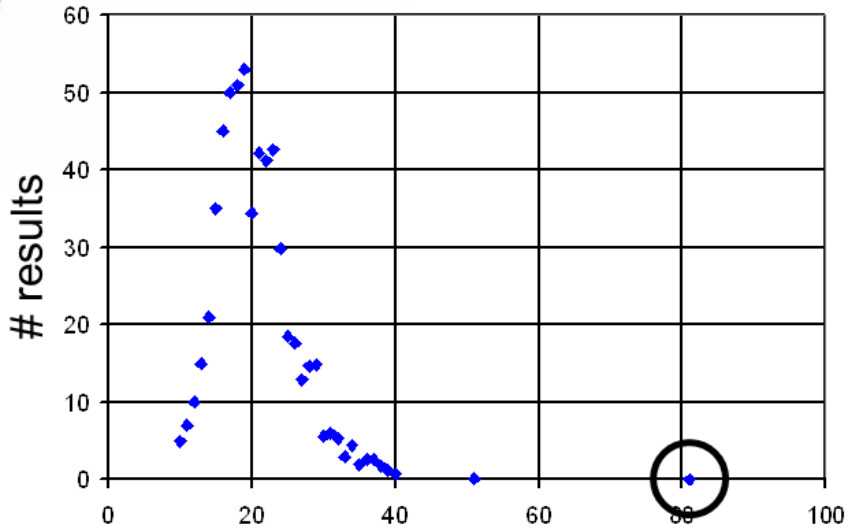
significant

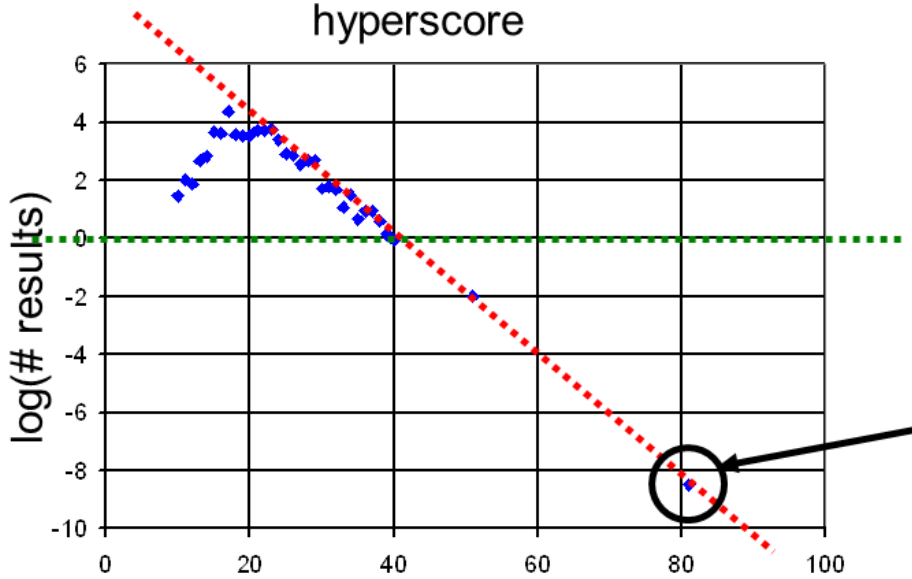
X!Tandem has already assumed that the top hyperscore is the only possible correct match.

This match is significant if it is greater than the point at which the straight line through the log data intersects the log(#results)=0 line.

Any hyperscores greater than this are unlikely to have arisen by chance.

E-value



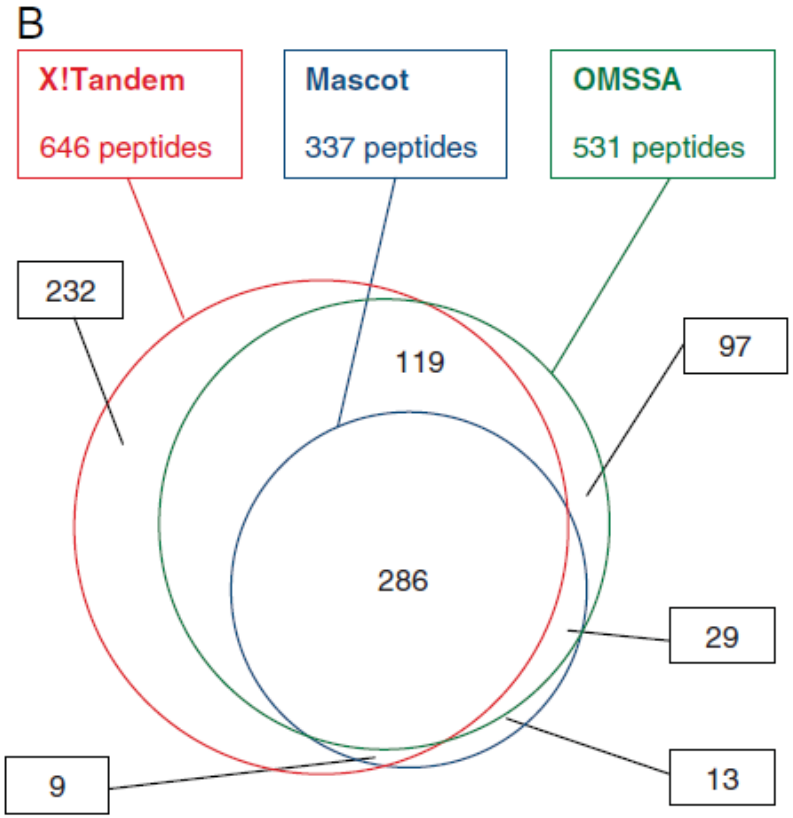


The E-value expresses just how unlikely a greater hyperscore is.

X!Tandem calculates the E-value by extrapolating the red line of the log histogram.

For the example shown, a hyperscore of 83 would occur by chance where the red line crosses 83. The log of this value — the E-value — is -8.2, as shown.

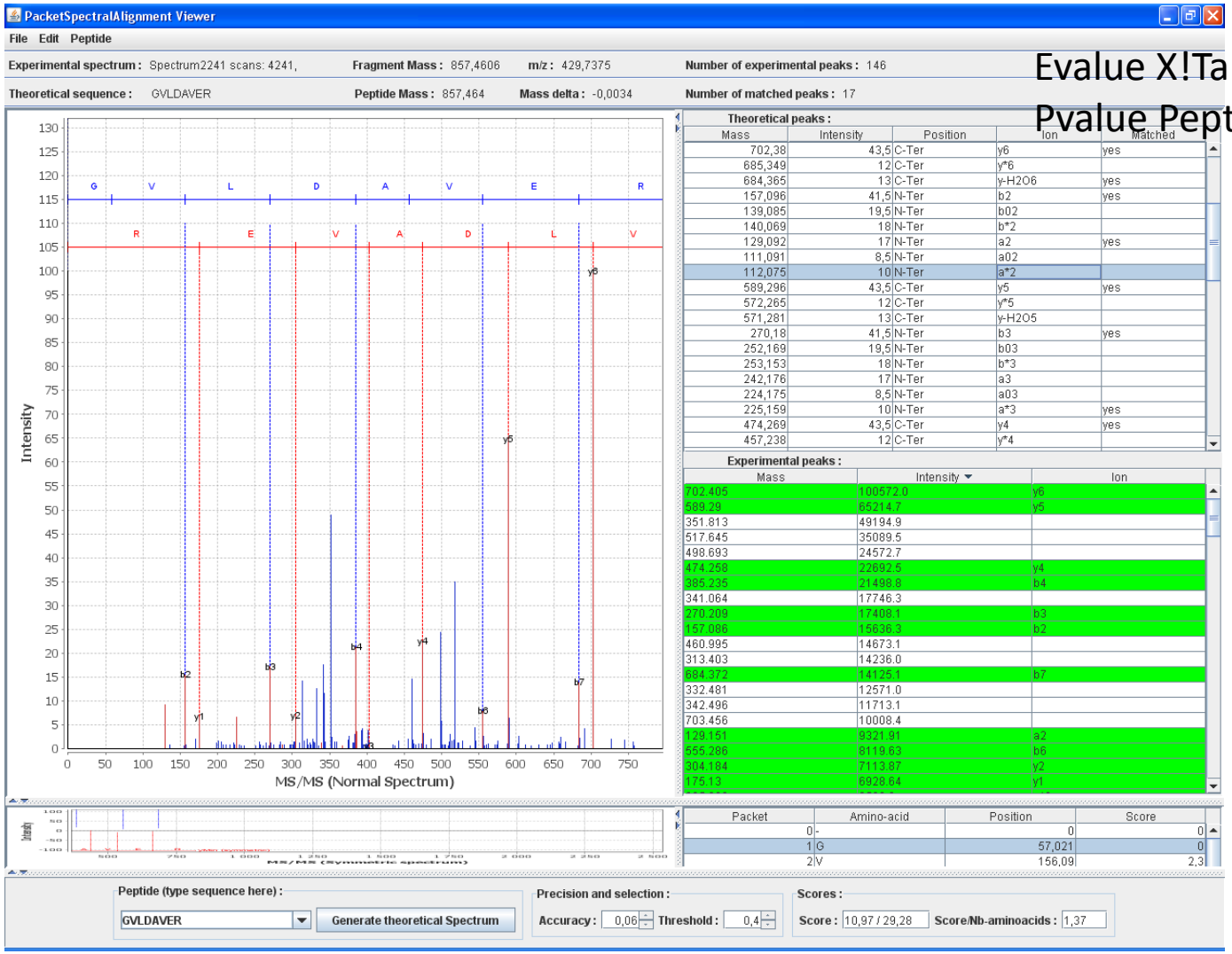
$E\text{-value} = e^{-8.2}$



Les différents moteurs de recherche sont en accord sur une majorité de peptides, mais chaque moteur permet d'identifier des peptides particuliers

Jeu de données UPS – standard, FDR=1%

Vaudel, M., Burkhart, J.M., Sickmann, A., Martens, L. and Zahedi, R.P. (2011) Peptide identification quality control, *Proteomics*, **11**, 2105-2114.



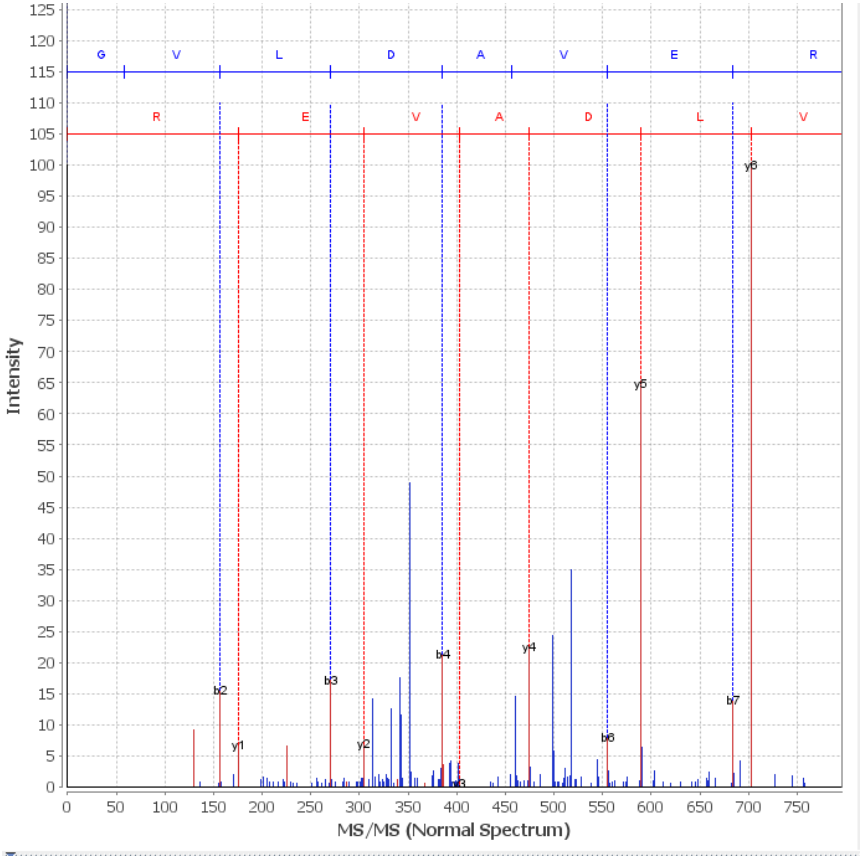
Evalue X!Tandem=4.6

Pvalue PeptideProphet =0.9387

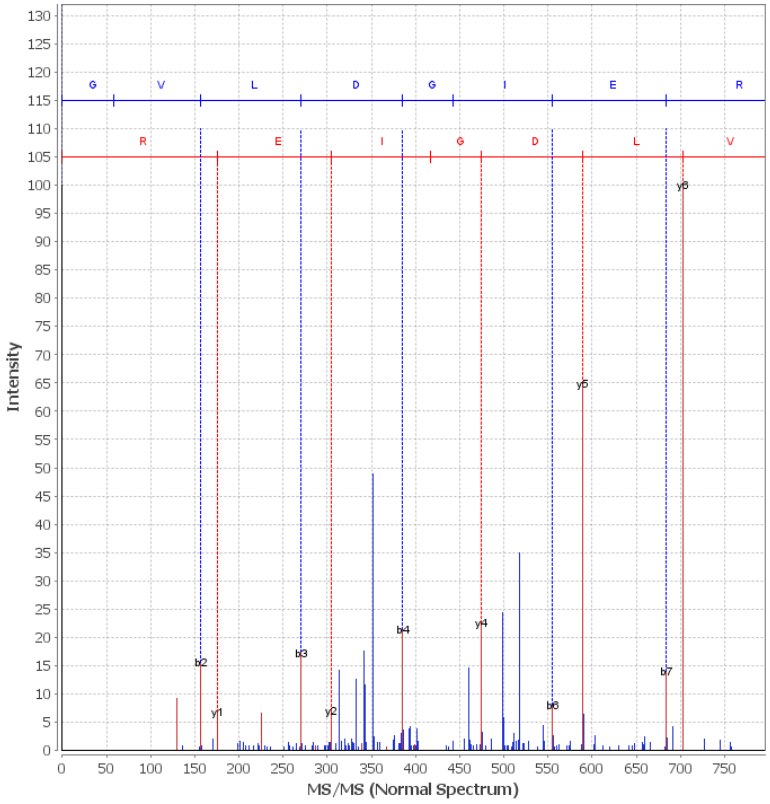
Un beau spectre, une evalue décevante ?

Séquence	Hyperscore	Delta masse	Expect
GVLDAVER	529	-0.0034	4.6
GVLDGIER	525	-0.0034	1
VGGGGGGGGER	488	-0.9371	2.6
RLATNQR	447	-0.023	7.2
RLAGYLR	408	-0.051	19
RLETLAR	401	-0.048	22
RLAAVKGK	400	-0.118	22

Interprétation GVLDAVER



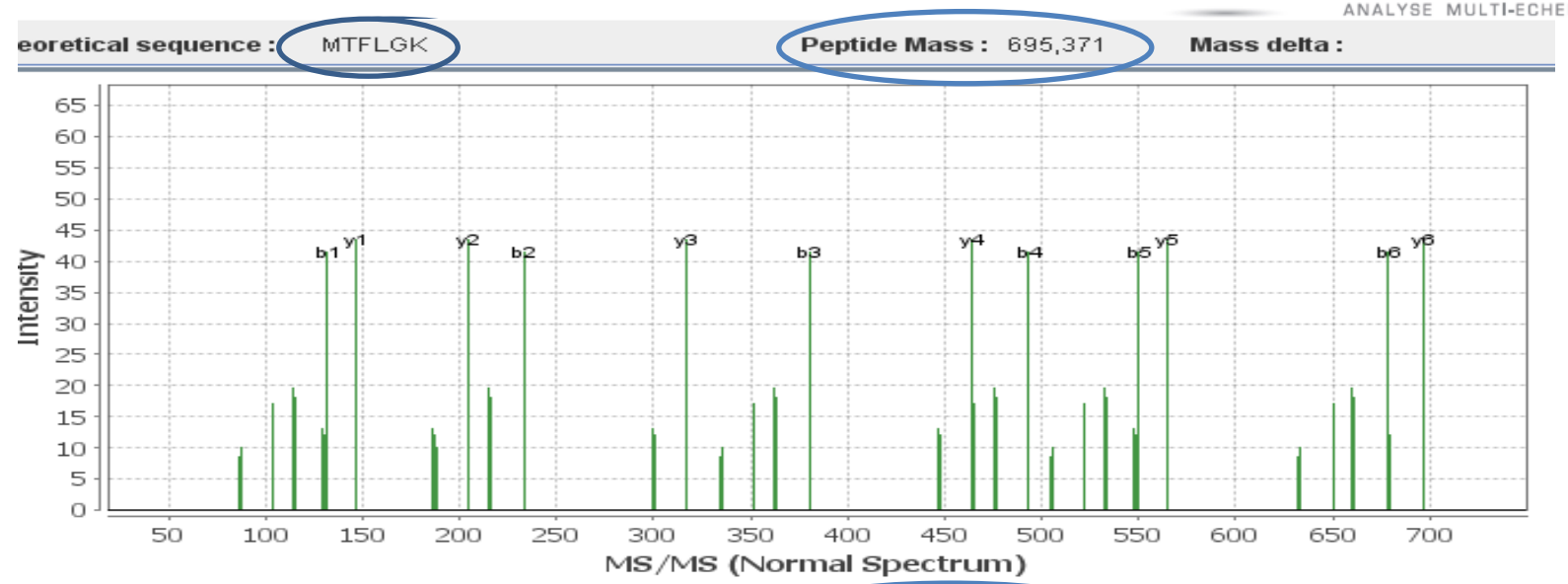
Interprétation GVLDGIER

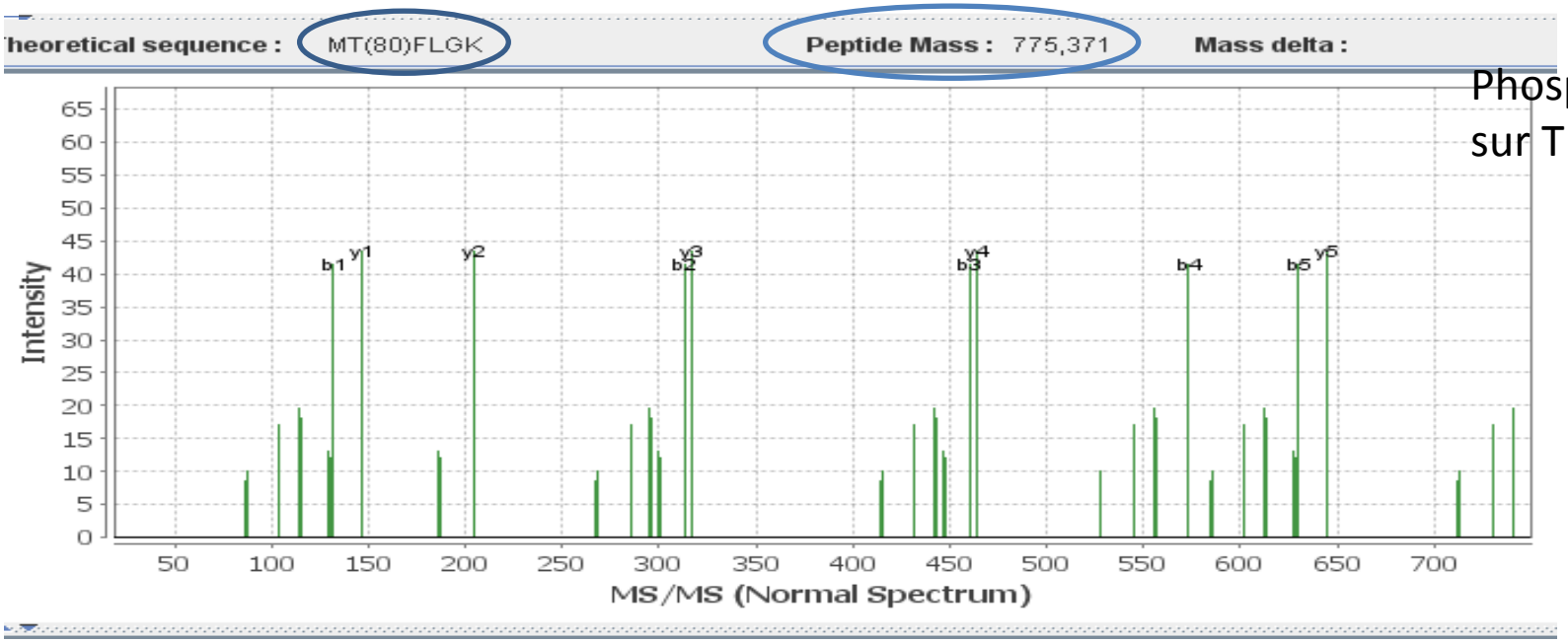


masse (AV) = masse(GI)

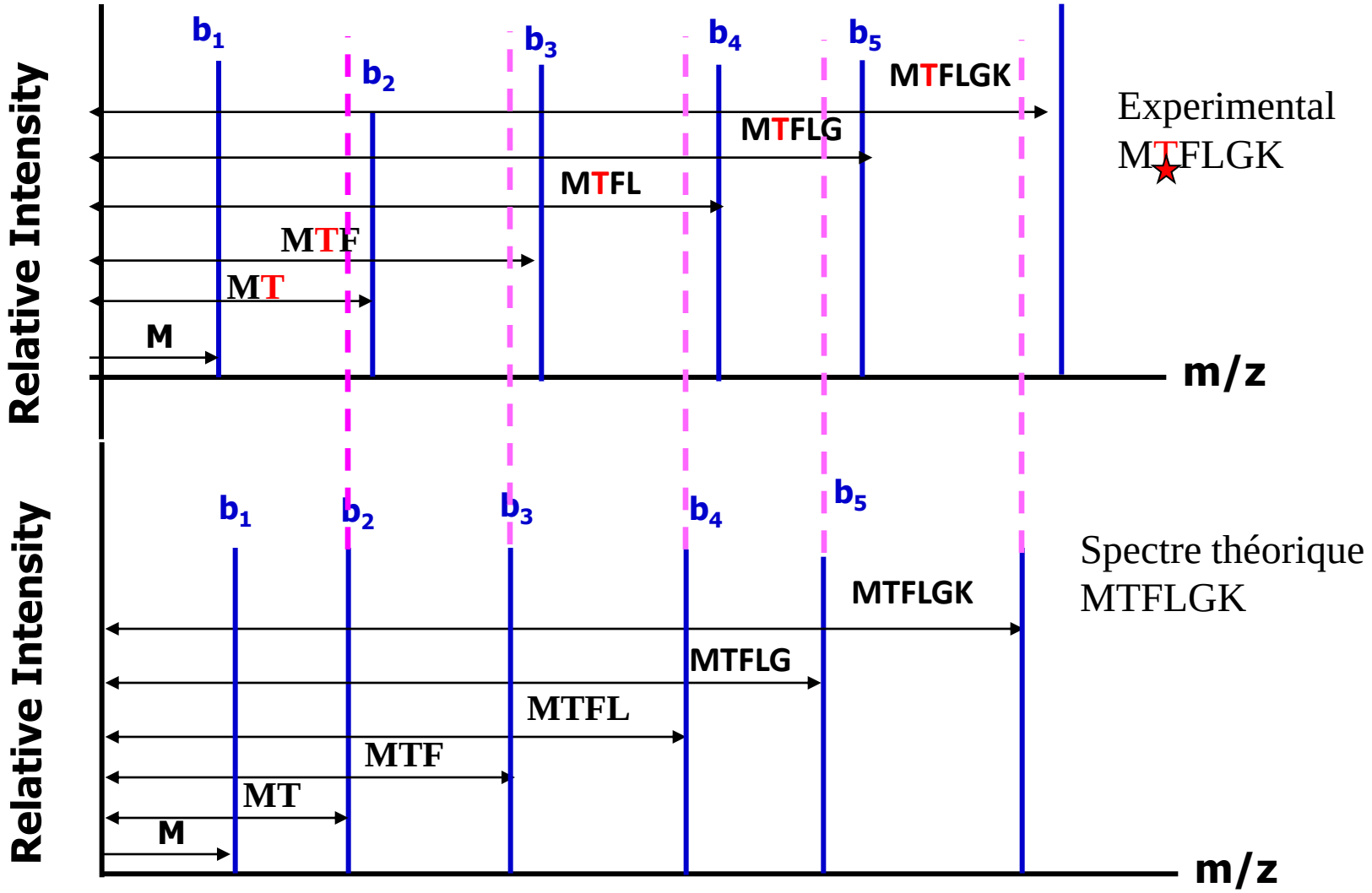
Les algorithmes de comparaison de spectres :

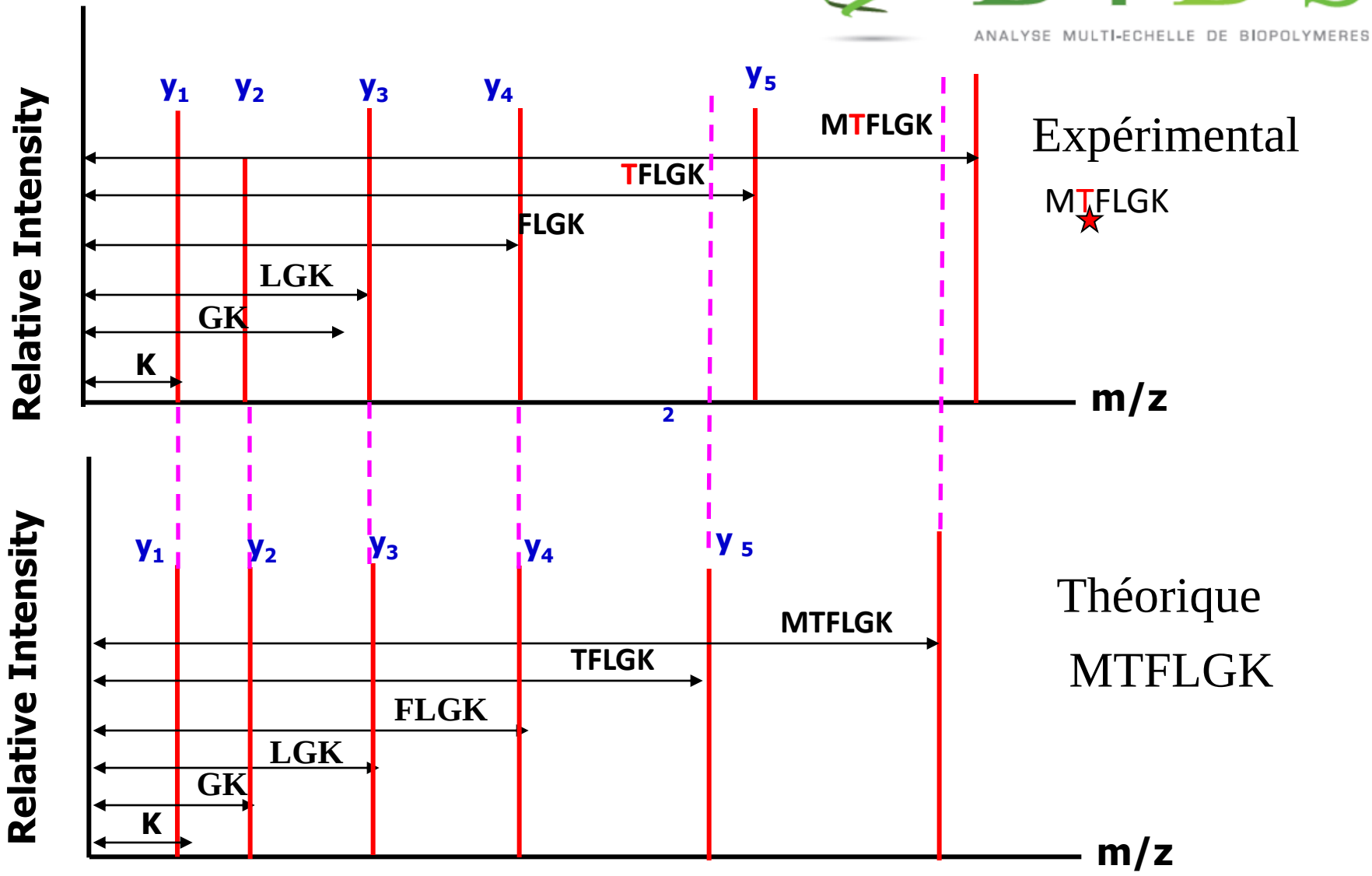
Quels sont les types d'ions considérés ?
Comment sont définies les intensités des spectres théoriques ? *
Quel est l'algorithme de comparaison utilisé ?
Quel est le calcul de score ? *
Comment les mutations , les modifications post-traductionnelles sont-elles prises en compte ?
Comment sont validées les attributions ?





Phosphorylation sur T (+80 Da)


Relative Intensity
b1
b2
b3
b4
b5
MTFLGK
MTFLG
MTFL
MTF
MT
M
m/z
Experimental
MTFLGK
Relative Intensity
b1
b2
b3
b4
b5
MTFLGK
MTFLG
MTFL
MTF
MT
M
m/z
Spectre théorique
MTFLGK


Relative Intensity
y1
y2
y3
y4
y5
MTFLGK
TFLGK
FLGK
LGK
GK
K
m/z
Expérimental
MTFLGK
Relative Intensity
y1
y2
y3
y4
y5
MTFLGK
TFLGK
FLGK
LGK
GK
K
m/z
Théorique
MTFLGK

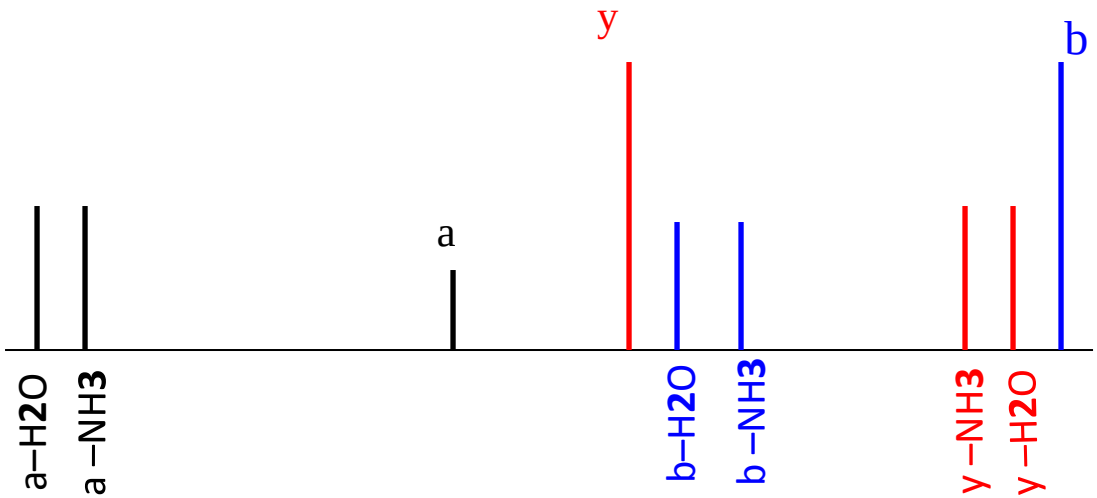
Idée originale de la thèse de Freddy Cliquet (2011): modéliser le spectre sous une forme différente à partir des relations qui lient les ions issus d'une même fragmentation

Relation entre les ions y et b: **masse(y) = Σ masse(aa) –masse(b) - 20**

masse (FLGK) = masse (MTFLGK) – masse (MT) -20

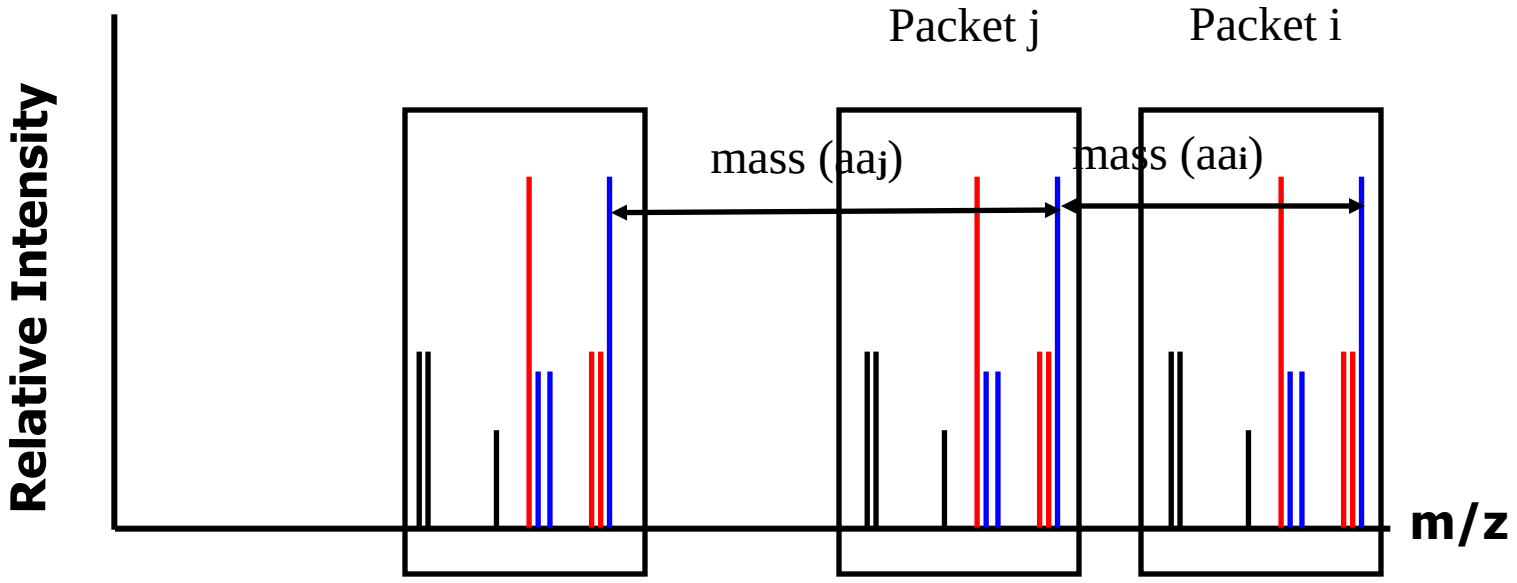
Nous pouvons regrouper les ions issus d'une même fragmentation en changeant le position des ions y

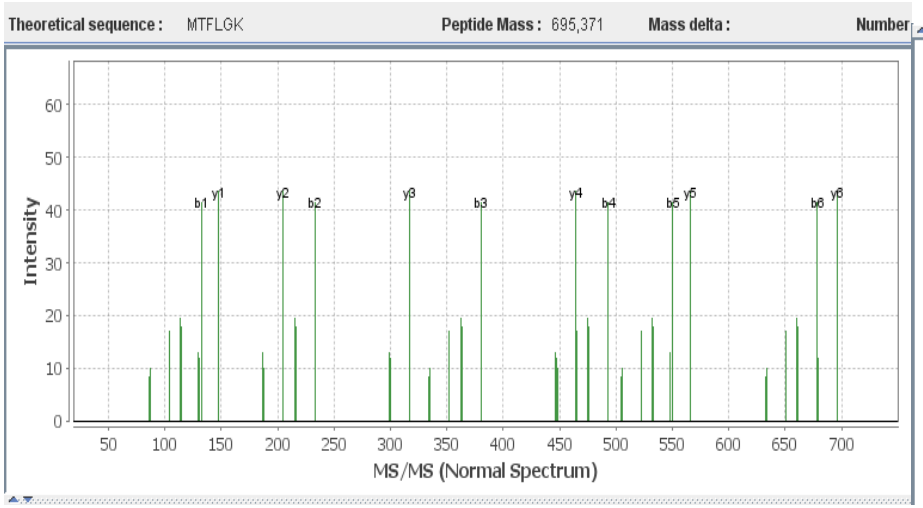
Position(y-ion) = Σ mass(aa) – mass(b-ion)

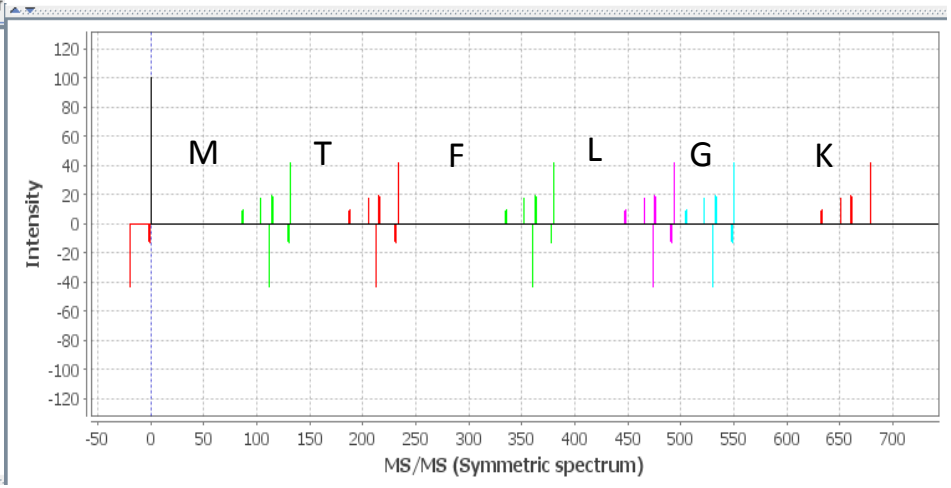


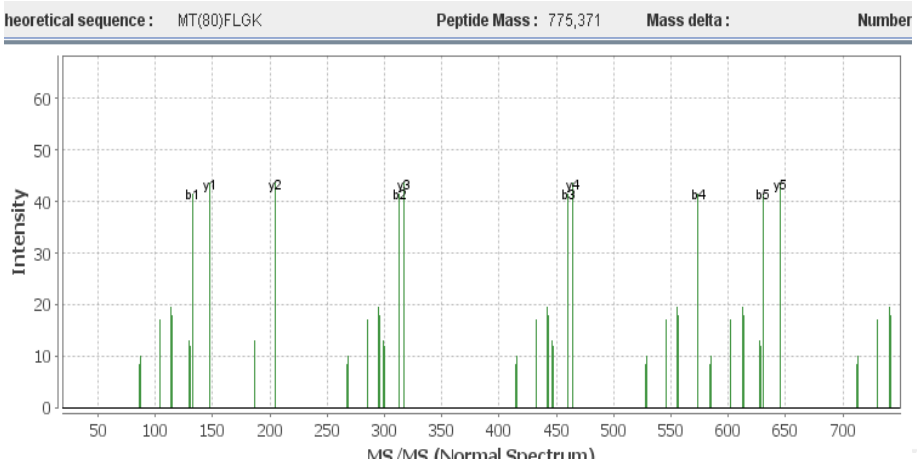
Les pics issus d'une même fragmentation sont regroupés dans un paquet

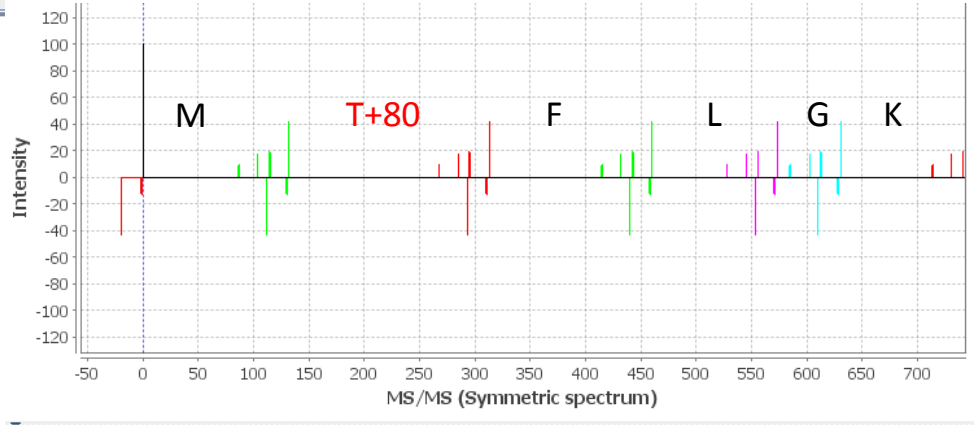
Un spectre théorique est alors représenté par une liste de paquets

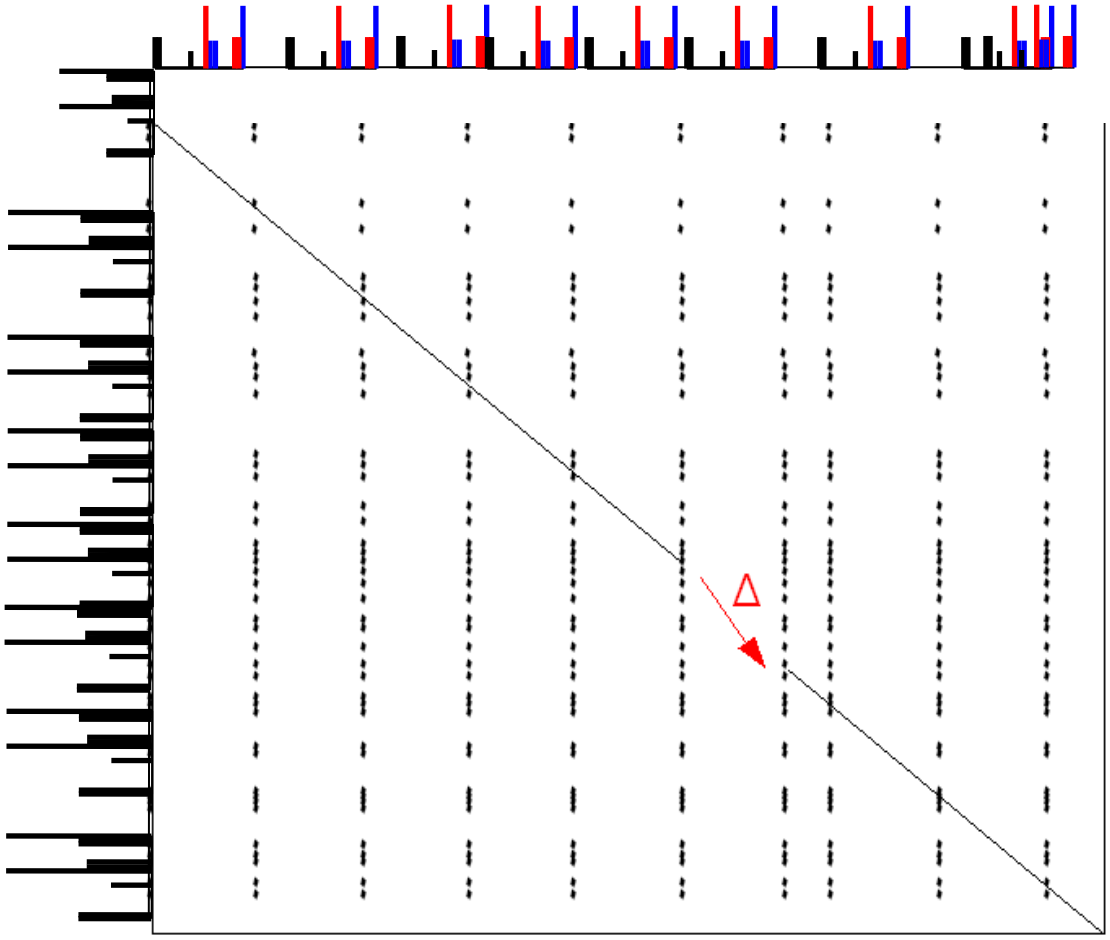



Theoretical sequence : MTFLGK
Peptide Mass : 695,371
Mass delta :
Number
Intensity
MS/MS (Normal Spectrum)


M
T
F
L
G
K
Intensity
MS/MS (Symmetric spectrum)


heoretical sequence : MT(80)FLGK
Peptide Mass : 775,371
Mass delta :
Number
Intensity
MS/MS (Normal Spectrum)


M
T+80
F
L
G
K
Intensity
MS/MS (Symmetric spectrum)



L'algorithme PSA (Packet Spectral Alignment) s'appuie alors sur la programmation dynamique pour aligner les paquets du spectre théorique sur le spectre expérimental comme l'algorithme SA développé par Pevzner et al.

[1] A. Pevzner, V. Dancik, and C. L. Tang. Mutation-tolerant protein identication by mass spectrometry. *J Comput Biol*, 7(6):777-787, 2000

La transformation du spectre théorique est simple !

La transformation du spectre expérimental est beaucoup plus complexe car on ne sait pas distinguer les différents types d'ions.
PSA génère donc un 'pic complémentaire' pour chaque pic, mais attention à ne pas générer de bruit (nombre de pics * 2!)

Les ions b du spectre théorique – resp. les ions y - ne seront alignés qu'avec les pics d'origine – resp. les pics complémentaires - du spectre expérimental.

Temps d'exécution longs :

AVG ~1 minute par spectrum sur un serveur linux et l'utilisation de 7 coeurs sur une banque d'environ ~500 000 peptides

L'algorithme ne peut être appliqué que sur

. Des spectres de bonne qualité

. Des spectres non-interprétés par des approches plus rapides

Cliquet, F., Fertin, G., Rusu, I. and Tessier, D. (2009) Comparison of Spectra in Unsequenced Species. . In (LNBI), L.N.i.B. (ed), *4th Brazilian Symposium on Bioinformatics (BSB 2009)*. Porto Alegre, Brazil.

Cliquet, F., Fertin, G., Rusu, I. and Tessier, D. (2010) Proper alignment of MS/MS spectra from unsequenced species. *Proc. 11th International Conference on Bioinformatics and Computational Biology (BIOCOMP 2010)*. CSREA Press, 766-772.

Plan de la présentation

Le contexte

L'association spectre-peptide par approche comparative (Peptide-Spectrum Match)

L'association spectre-peptide par approche *de novo*

L'identification des protéines

La définition de « pipeline » d'analyse

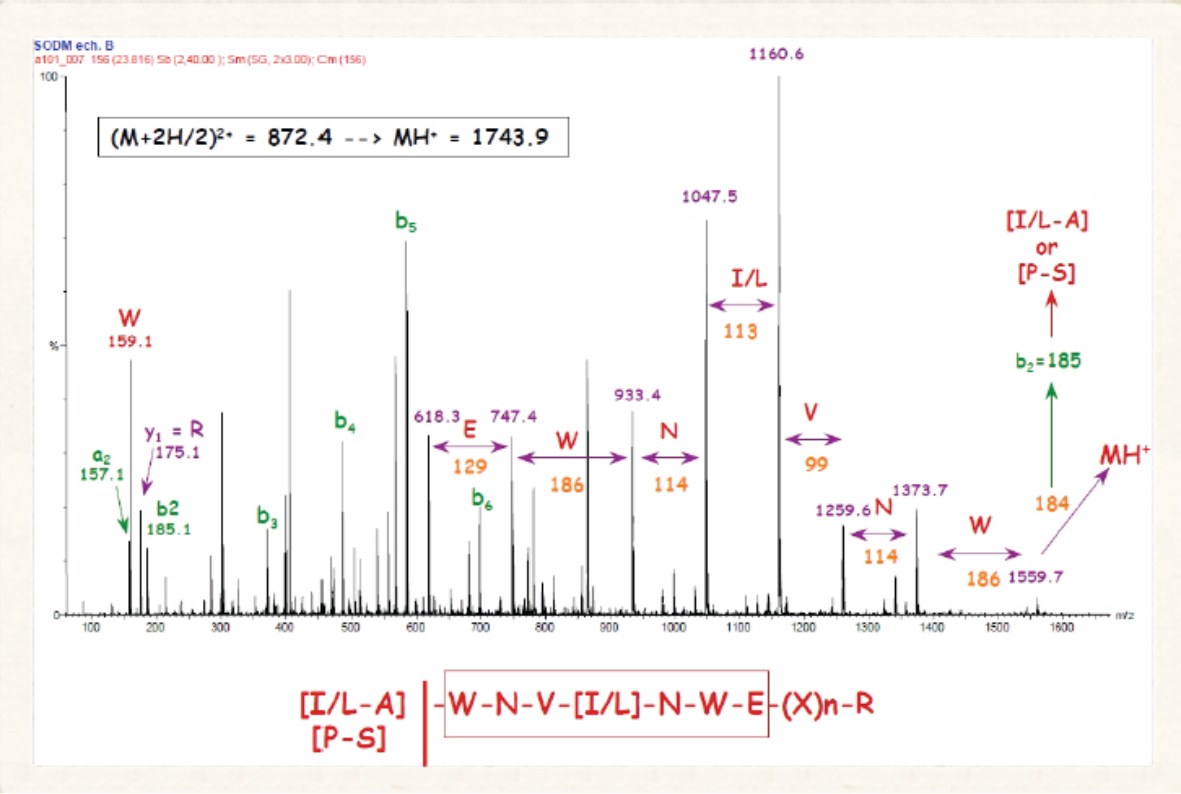
Plan de la présentation

Le contexte

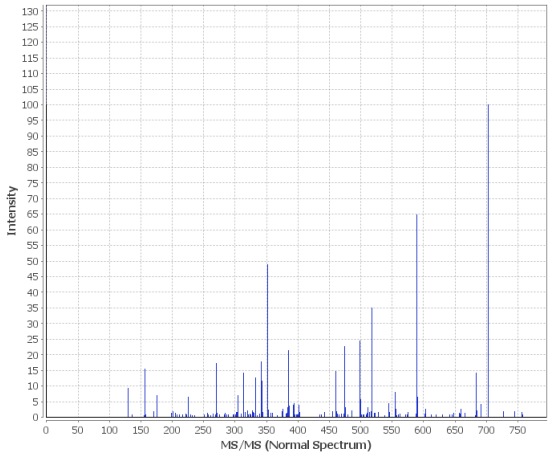
L'association spectre-peptide par approche comparative

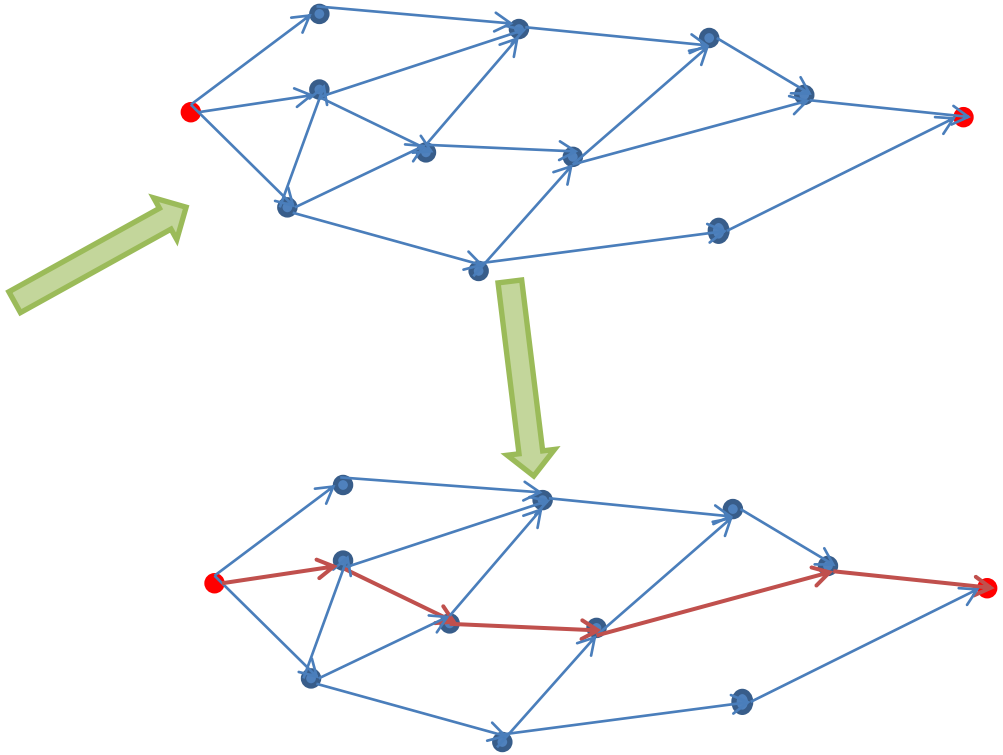
L'association spectre-peptide par approche *de novo*

L'identification de protéines

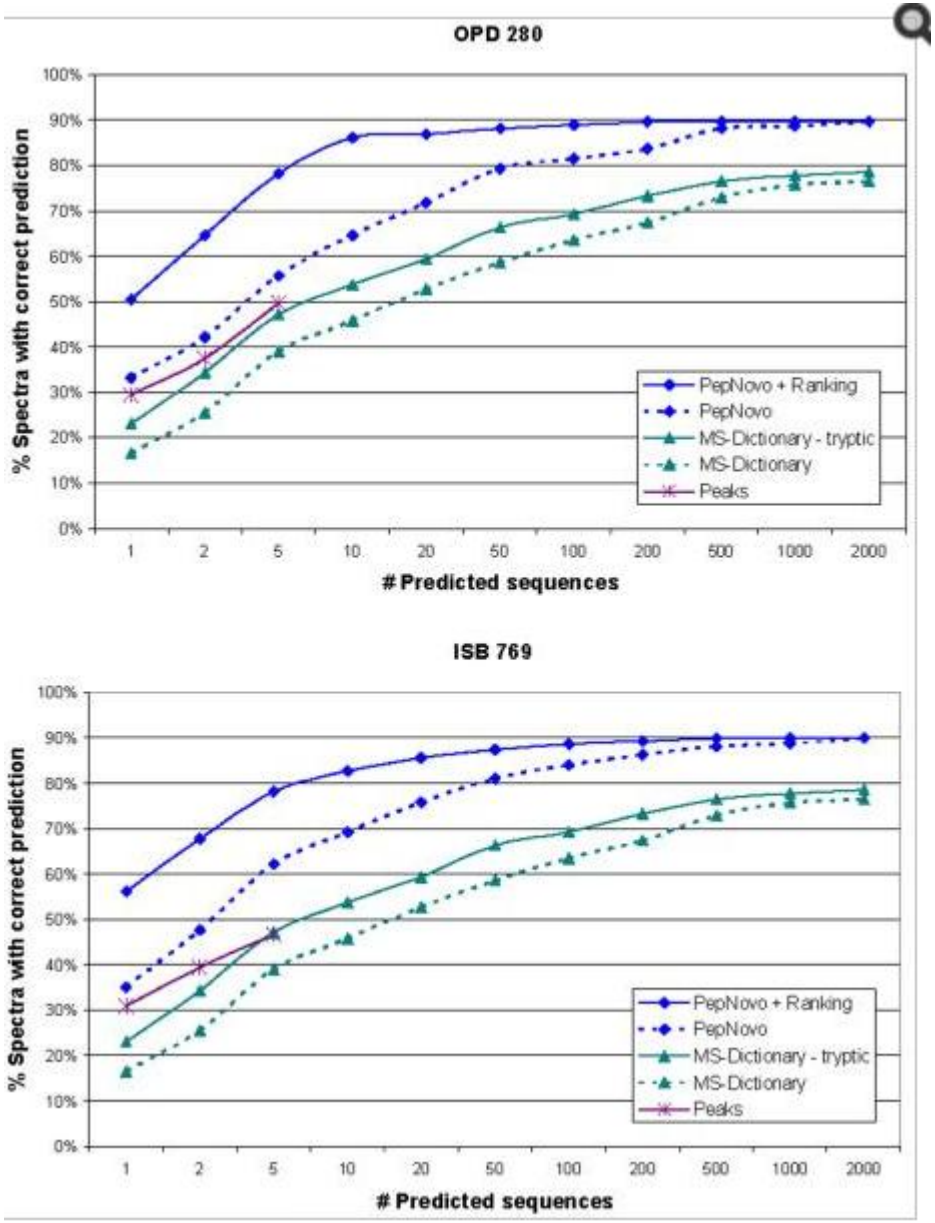

SODM ech. B
(M+2H/2)2+ = 872.4 --> MH+ = 1743.9
1160.6
1047.5
b5
I/L
113
[I/L-A]
or
[P-S]
W
159.1
b2=185
933.4
618.3
E
747.4
V
99
y1 = R
175.1
b4
W
186
N
114
MH+
a2
157.1
b2
185.1
b3
b6
1259.6
N
1373.7
184
W
186
1559.7
[I/L-A]
[P-S]
-W-N-V-[I/L]-N-W-E-(X)n-R

Utilisation de graphes de spectres (introduits par Bartels – 1990 -)





Recherche d’un chemin optimal



Le jeu d'évaluation correspond à une sélection de spectres : en moyenne, ils contiennent 47.7% des ions b et 51.2% des ions y

Ranking-Based Scoring Models for Peptide-Spectrum Matches. Frank, A.M. J. Proteome Research, 8:2241-2252, 2009

En résumé

Trouver une association entre un spectre et un peptide ? Facile !!!! Oui mais est-elle correcte ?

Erreurs:

des spectres de mauvaise qualité, mais pas toujours. Même des spectres de bonne qualité peuvent fournir des associations incorrectes si le peptide analysé n'est pas dans la banque de protéines.

Plan de la présentation

Le contexte

L'association spectre-peptide par approche comparative (Peptide-Spectrum Match)

L'association spectre-peptide par approche *de novo*

L'identification des protéines

La définition de « pipeline » d'analyse

Difficulté à assembler les peptides : Protein inference problem

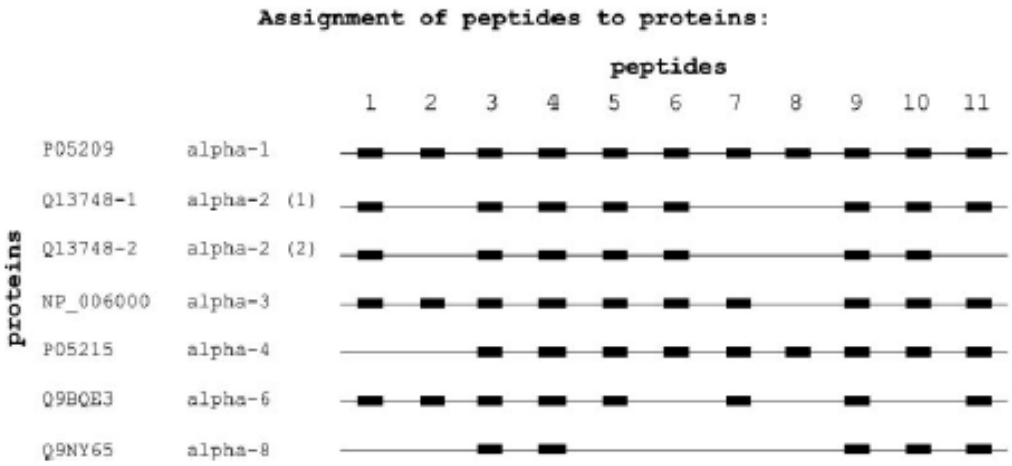
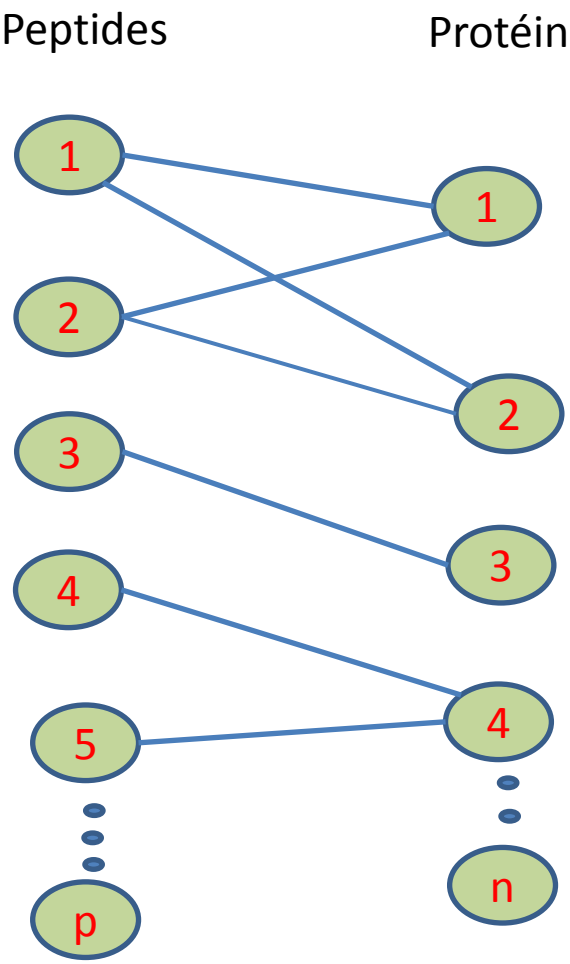


FIG. 3. **An example of a protein family.** Eleven tryptic peptides are identified that are shared between the members of the α-tubulin family. None of the proteins is identified by a peptide that is unique to it, thus making it impossible to determine which particular member(s) of the family is present in the sample.

Nesvizhskii, A.I. and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem, *Mol Cell Proteomics*, **4**, 1419-1440.

Le problème est en général formalisé sous la forme d'un graphe biparti



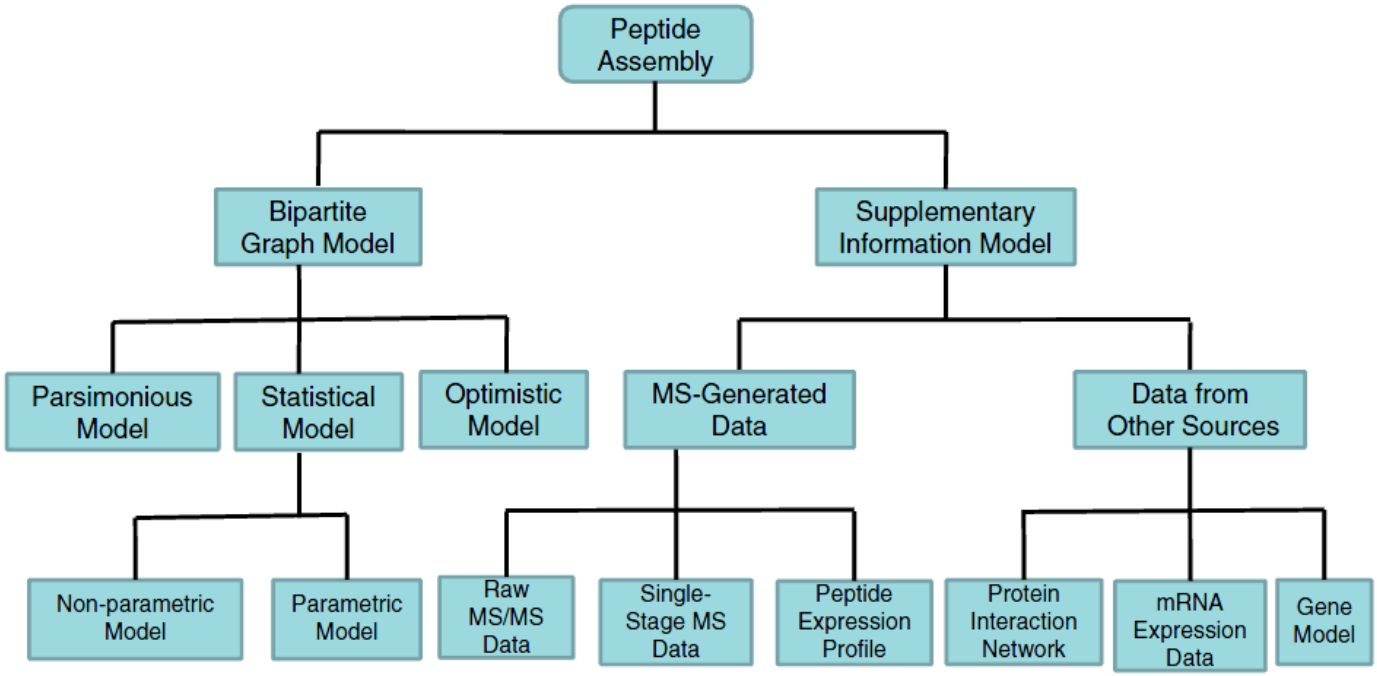
m = nombre max de protéines

n = nombre min de protéines
⇒ Set coverage problem (NP-complet)
⇒ Principe de parcimonie

Le problème des protéines avec 1 seul peptide

Gupta, N. and Pevzner, P.A. (2009) False discovery rates of protein identifications: a strike against the two-peptide rule, *J Proteome Res*, **8**, 4173-4181.

Les différents algorithmes

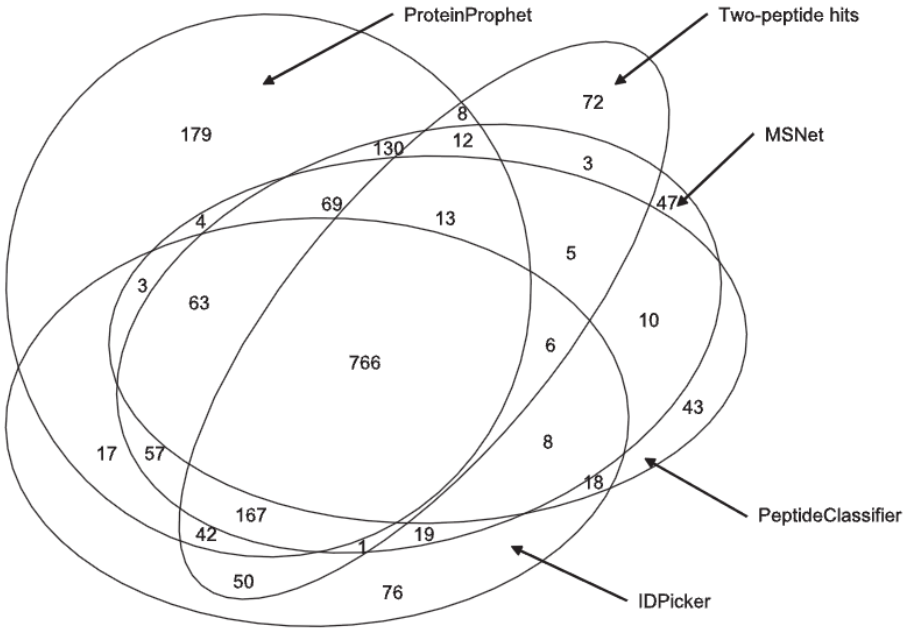


Huang, T., Wang, J., Yu, W. and He, Z. (2012) Protein inference: a review, *Brief Bioinform*, **13**, 586-614.

Table 2: The performance of five methods

Data	IDPicker	ProteinProphet	MSNet	PeptideClassifier	Two-peptide hits
Yeast.DI	**1293** (1152)	**1530** (1254)	**1376** (1319)	**1008** (977)	**1112** (978)
Yeast.D2	**154** (116)	**322** (162)	**291** (166)	**232** (145)	**317** (133)

Two protein samples undergo MS/MS analysis to generate ten lists of proteins identified by five peptide assembly algorithms. The figures in brackets indicate the number of proteins present in the reference set.



Huang, T., Wang, J., Yu, W. and He, Z. (2012) Protein inference: a review, *Brief Bioinform*, **13**, 586-614.

PSM	UPS2	Yeast YPD	iPRG09(Red)	iPRG09(Yellow)
Total MS/MS spectra observed	74,602	240,781	69,416	70,970
SEQUEST	32,651 (87)	57,955 (268)	9,524 (98)	9,492 (83)
X!Tandem	27,264 (210)	74,244 (332)	15,147 (117)	15,366 (112)
MyriMatch	26,262 (79)	41,179 (106)	9,706 (88)	9,134 (46)
InsPecT	25,618 (64)	69,341 (414)	12,691 (202)	13,295 (216)
Union	40,829 (434)	95,315 (1053)	21,764 (505)	21,684 (455)
MSblender	39,273 (336)	99,814 (1011)	23,580 (153)	23,717 (177)
1 engine	4,043 (190)	10,441 (100)	2,138 (38)	2,073 (52)
2 engines	7,389 (89)	16,861 (546)	3,768 (76)	3,878 (74)
3 engines	5,560 (35)	32,111 (203)	6,820 (24)	6,816 (21)
4 engines	22,202 (24)	38,257 (18)	10,830 (3)	10,826 (4)
Protein	UPS2	Yeast YPD	iPRG09(Red)	iPRG09(Yellow)
Total proteins	48	6,698	4,417	4,417
SEQUEST	38	1,391	757	749
X!Tandem	38	1,459	870	847
MyriMatch	36	1,241	722	657
InsPecT	29	1,527	877	902
Union	44	1,873	999	1,024
MSblender	42	1,911	1,185	1,147

Kwon, T., Choi, H., Vogel, C., Nesvizhskii, A.I. and Marcotte, E.M. (2011) MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines, *J Proteome Res*, **10**, 2949-2958.

Résultat à 0.5% de FDR

Plan de la présentation

Le contexte

L'association spectre-peptide par approche comparative (Peptide-Spectrum Match)

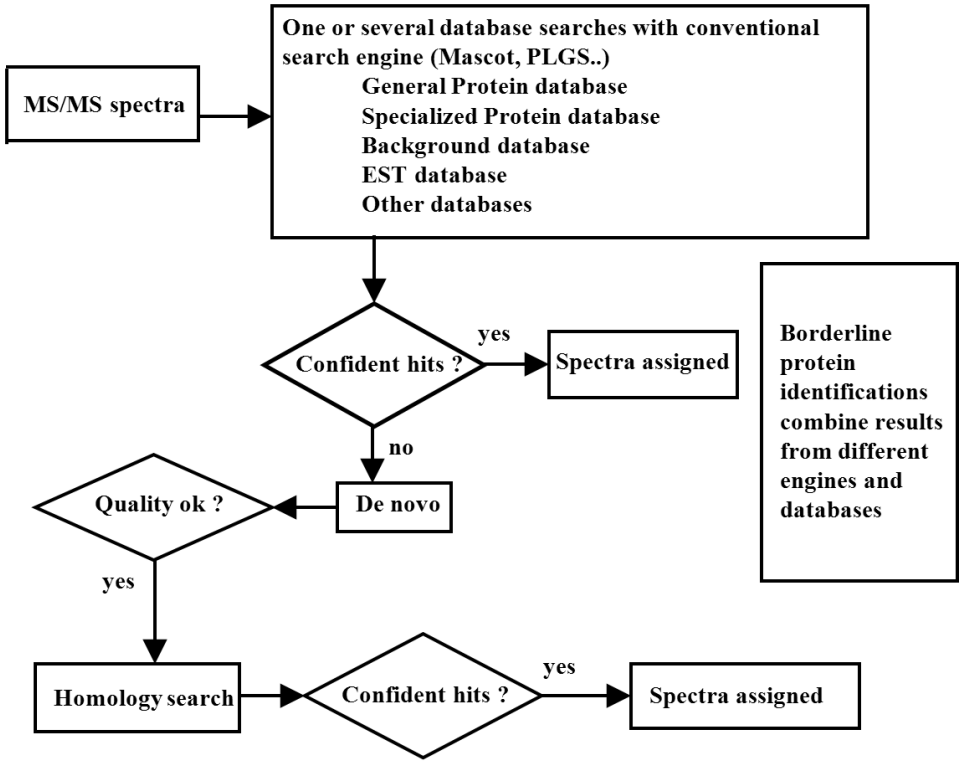
L'association spectre-peptide par approche *de novo*

L'identification des protéines

La définition de « pipeline » d'analyse

Développement de processus d'analyse adaptés aux questions de recherche

Développement du logiciel OVNIp



Remerciements

*La composante de spectrométrie de masse de la plateforme **BIBS**, en particulier Hélène Rogniaux, Audrey Geairon*

Freddy Cliquet (doctorant), Guillaume Merceron (master 2)

*L'équipe ComBi du **LINA**, en particulier Guillaume Fertin, Irena Rusu*

*Colette Larré (unité **INRA-BIA**)*

*Le groupe bioinformatique de **BIBS** : Virginie Lollier, Stéphane Bansard, Marc Vasseur*