

PEPTIDE Mass Sequencing

- 1) Stabilire l'identità di una proteina
- 2) Mappare i siti recanti modificazioni post-traduzionali
- 3) De novo sequencing



Sviluppo di strategie sperimentali per la combinazione di diverse tecniche separative (LC) con un'ampia varietà di spettrometri di massa.

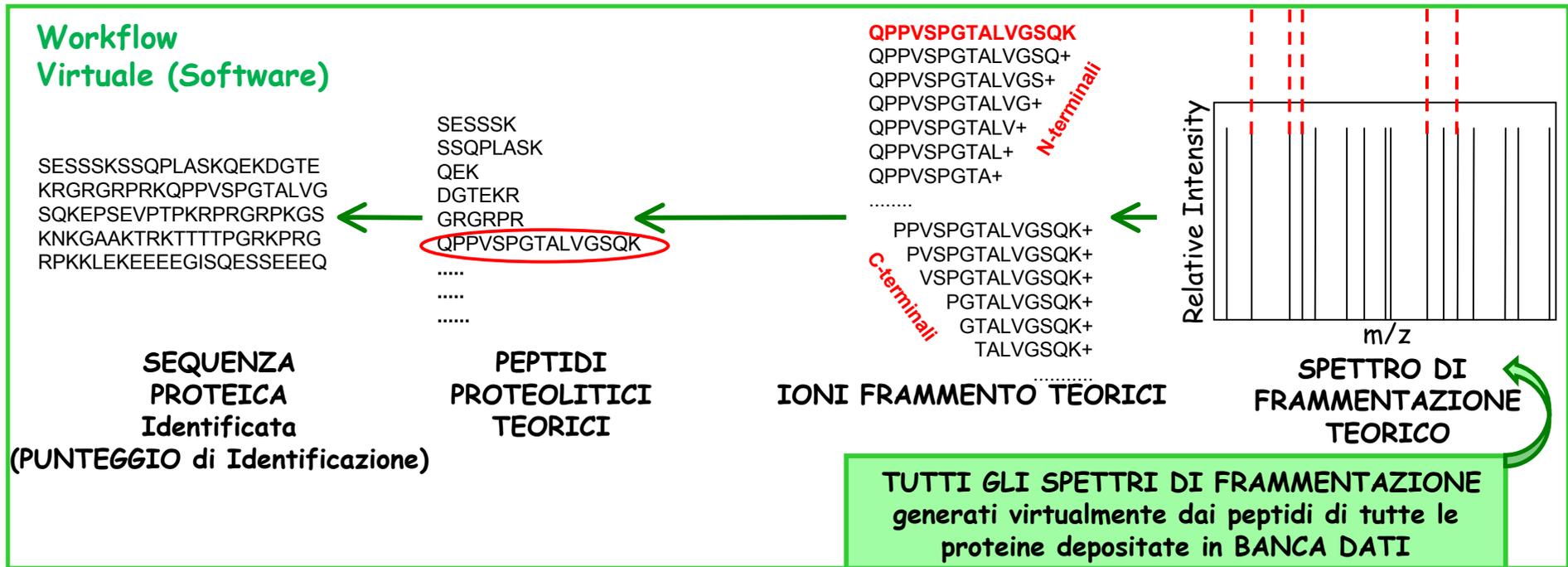
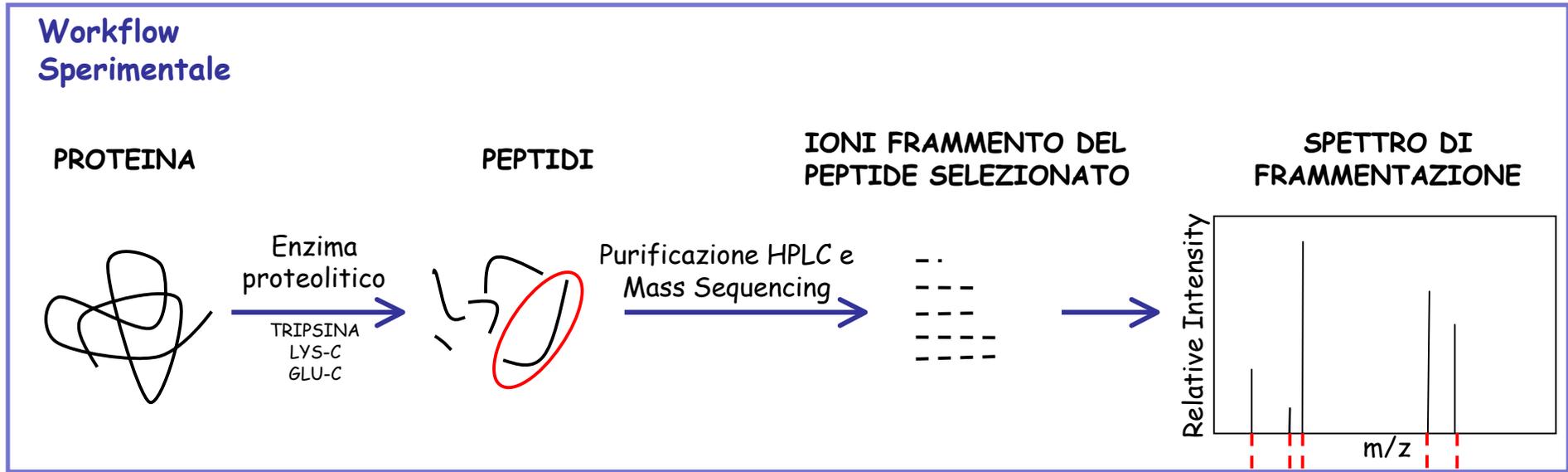


Sviluppo di software (database-supported and probability-based) per l'interpretazione degli spettri.

Sempre > "CONFIDENZA"

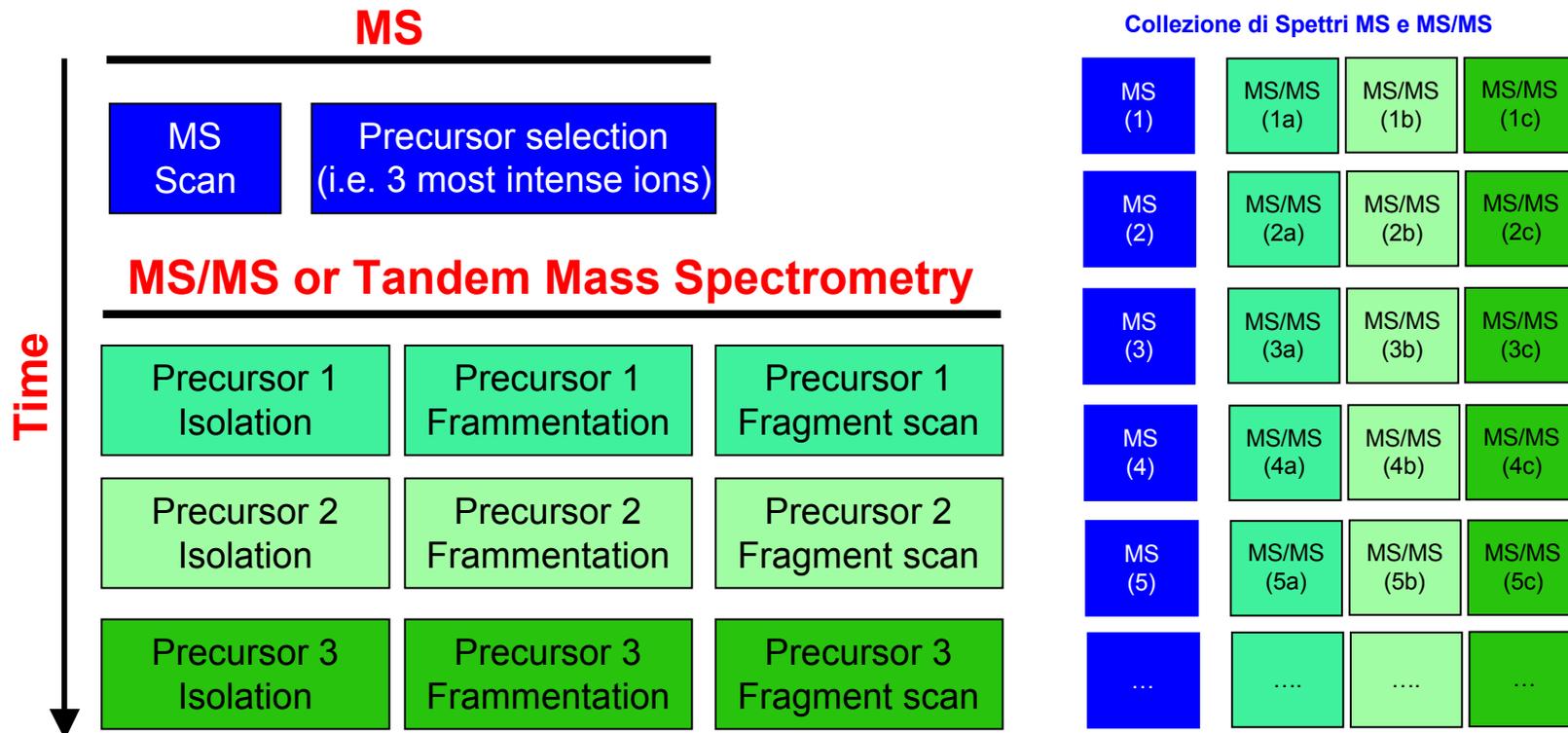


PEPTIDE Mass Sequencing

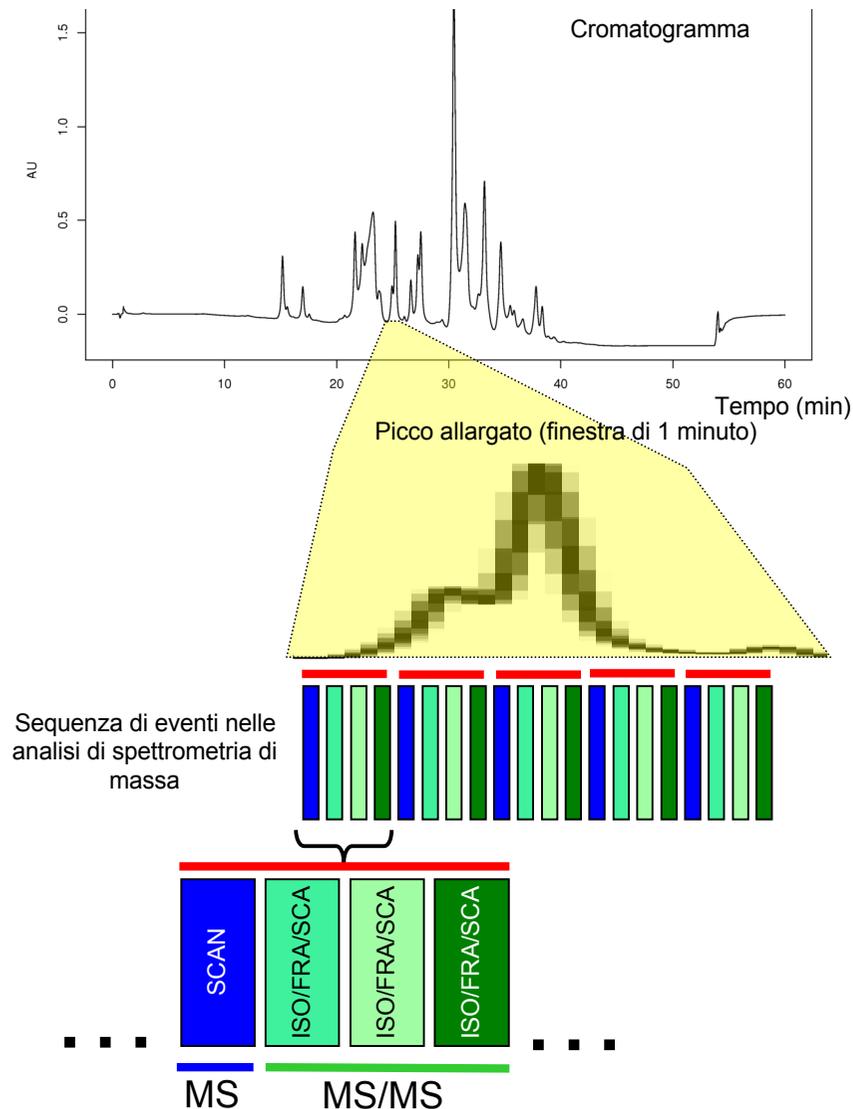


PEPTIDE Mass Sequencing

Dopo aver effettuato una scansione MS si possono identificare degli ioni precursori da sottoporre a frammentazione. Lo strumento passa automaticamente in modalita MS/MS e isola, frammenta e rileva i rapporti m/z dei frammenti generati. Nel caso del Peptide mass sequencing e qualora la metodica di frammentazione adottata sia la CID, i peptidi si trovano a rompersi prevalentemente a livello del legame peptidico (vedasi dopo la teoria del protone mobile). Le rotture non avvengono sempre a livello del medesimo legame peptidico, bensì, ciascuna diversa molecola (dello stesso peptide/molecola) è sottoposta ad una (ed una sola) rottura a livello diverso (chi tra il primo e il secondo residuo aa, chi tra il secondo e il terzo, chi tra il terzo e il quarto, etc.). Si generano in questo modo una serie di frammenti che tra di loro differiscono per singoli residui aa. Dal momento che il peso di ciascuno residuo all'interno di una catena polipeptidica è calcolabile, dalla differenza di peso (ricavato a partire dai valori m/z) tra i vari picchi dello spettro di frammentazione è possibile ricavare la sequenza aa del peptide sotto analisi. Bisogna però considerare che le rotture non avvengono in modo equivalente a livello di tutti i legami peptidici (questo perché l'intorno molecolare influenza la rottura - vedasi modello del protone mobile). Da questo ne consegue che lo spettro di frammentazione non è detto contenga tutti i frammenti.



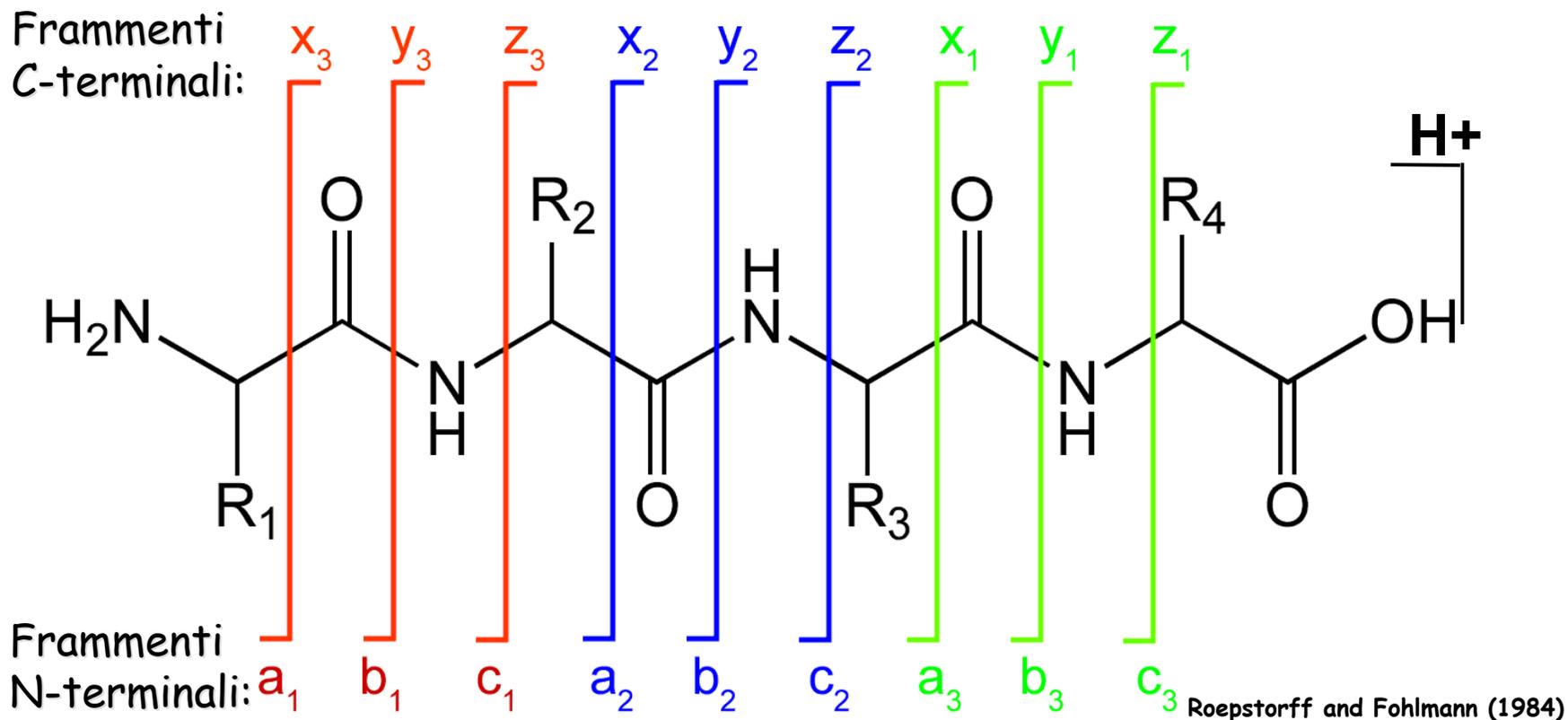
PEPTIDE Mass Sequencing - Analisi LC-MS/MS



Se la finestra allargata è di un minuto e se in essa sono stati acquisiti 4 cicli, ciascun ciclo (MS + 3 MS/MS) dura circa 15 secondi e il picco in questione è campionato in 4 punti. In una situazione reale, una scansione MS dura indicativamente 200 ms e sono necessari altri 200 ms per isolare, frammentare e scansionare, questo significa che per 3 MS/MS servono 600 ms il che porta in pratica ad avere nel caso specifico un tempo di circa 1 sec a ciclo (NB: i tempi forniti sono indicativi)

Nelle analisi proteomiche i digeriti triptici ottenuti dal campione che si vuole analizzare (i.e. ottenere informazioni circa le proteine presenti - identità - e le loro quantità) vengono solitamente analizzati mediante cromatografia liquida (LC) usando la metodica della RP-HPLC (dal momento che le condizioni che vengono utilizzate in questa metodica sono compatibili con la ionizzazione di tipo ESI). Lo spettrometro di massa analizza ciò che eluisce dalla colonna in base a quanto stabilito da noi. Una classica impostazione è chiedere allo strumento di fare una scansione (MS), vedere quali precursori (segnali m/z che potrebbero appartenere a peptidi) sono presenti nello spettro e sulla base di questo (per questa ragione questa modalità è anche chiamata DATA DEPENDENT ACQUISITION) passare in modalità MS/MS e quindi isolare, frammentare e fare la scansione dei frammenti generati a partire dai precursori scelti. Questa serie di operazioni si susseguono nel tempo lungo la corsa cromatografica. La velocità dello spettrometro di massa nel fare queste operazioni è essenziale per "campionare" il mio cromatogramma in modo dettagliato. Infatti, mentre lo spettrometro di massa è impegnato nel fare un'operazione le altre molecole continuano ad eluire e se lo strumento non è sufficientemente veloce il rischio è che alcune vadano perse (non vengono analizzate).

Le 3 principali famiglie di ioni frammento



La serie **a** e la serie **x**= rottura legame tra $C\alpha$ e C carbonilico

La serie **b** e la serie **y**= rottura legame amidico

La serie **c** e la serie **z**= rottura legame tra N amidico e $C\alpha$ aa successivo

(più rari: frammenti interni - doppia rottura sul backbone
 frammento e ioni satellite - rotture sulla catena laterale)

PEPTIDE Mass Sequencing

Determinazione sequenza aminoacidica mediante spettrometria di massa (MS/MS)

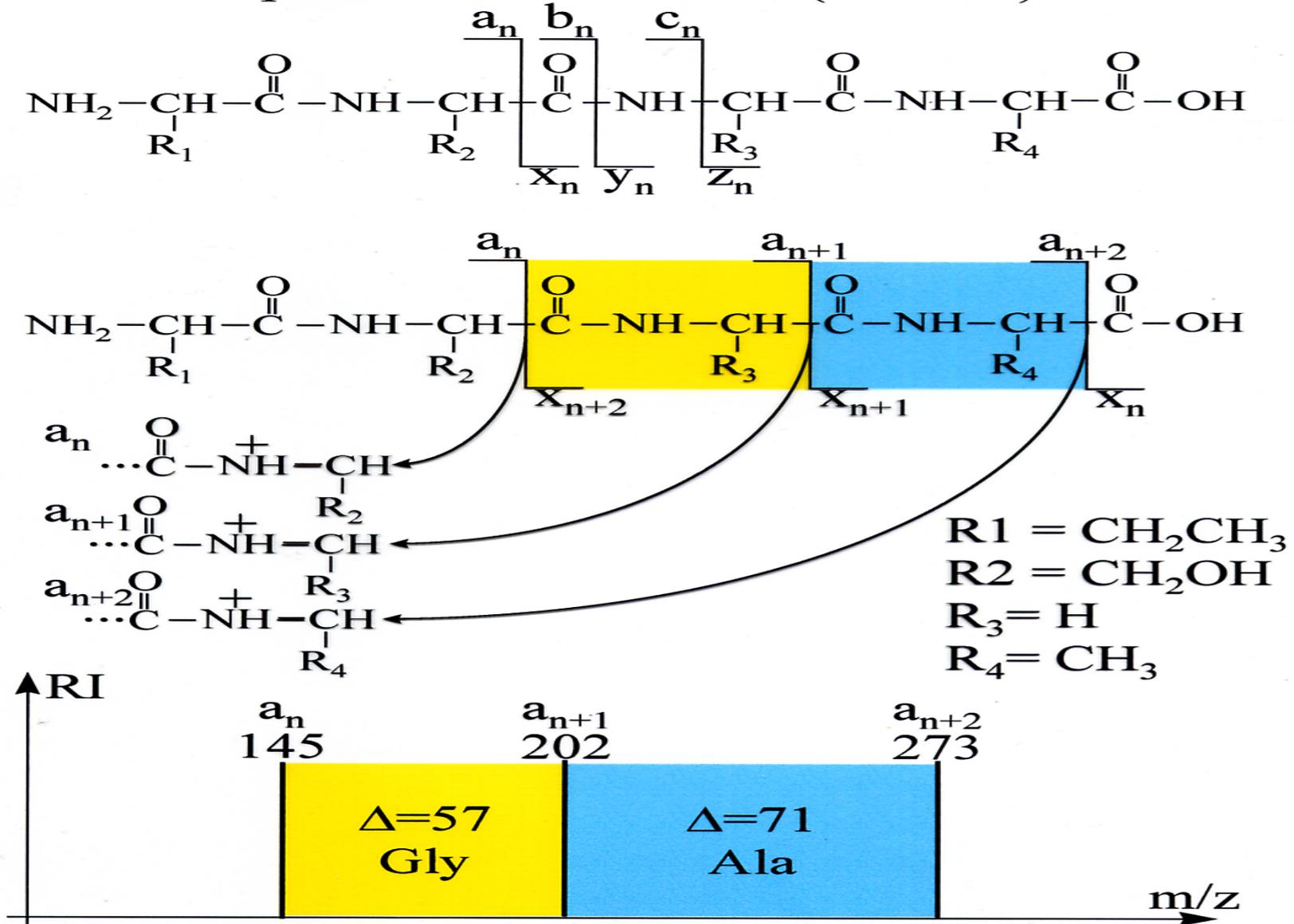
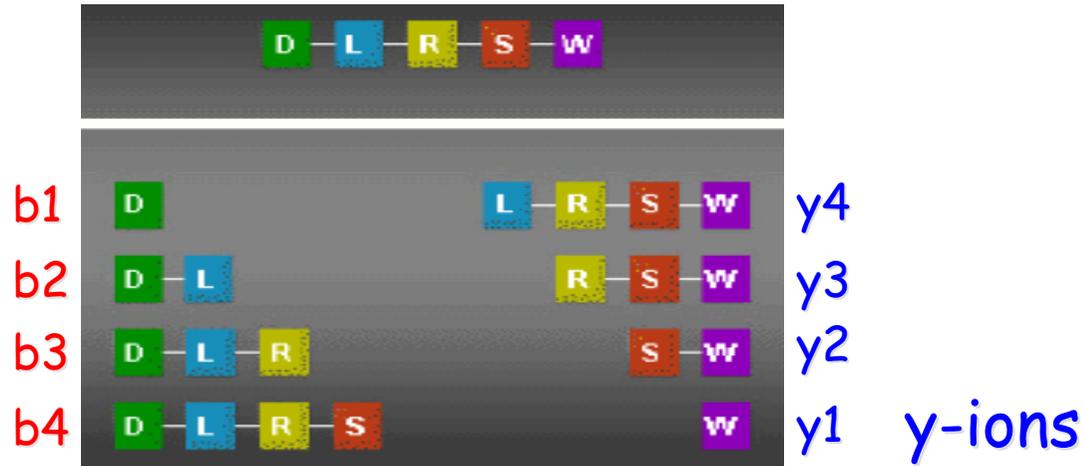


Table 7.3: Mass increments of the various amino acids				
Amino acid	Code (3 letters)	Code (1 letter)	Monoisotopic mass	Chemical mass
Glycine	Gly	G	57.02147	57.052
Alanine	Ala	A	71.03712	71.079
Serine	Ser	S	87.03203	87.078
Proline	Pro	P	97.05277	97.117
Valine	Val	V	99.06842	99.133
Threonine	Thr	T	101.04768	101.105
Cysteine	Cys	C	103.00919	103.144
Isoleucine	Ile	I	113.08407	113.160
Leucine	Leu	L	113.08407	113.160
Asparagine	Asn	N	114.04293	114.104
Aspartate	Asp	D	115.02695	115.089
Glutamine	Gln	Q	128.05858	128.131
Lysine	Lys	K	128.09497	128.174
Glutamate	Glu	E	129.04260	129.116
Methionine	Met	M	131.04049	131.198
Histidine	His	H	137.05891	137.142
Phenylalanine	Phe	F	147.06842	147.177
Arginine	Arg	R	156.10112	156.188
Tyrosine	Tyr	Y	163.06333	163.17
Tryptophan	Try	W	186.07932	186.213

NB:importanza della precisione strumentale

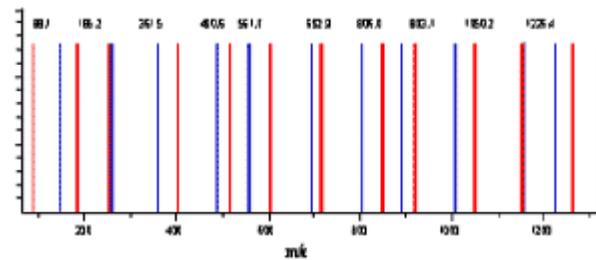
Gli ioni della serie b e della serie y

b-ions

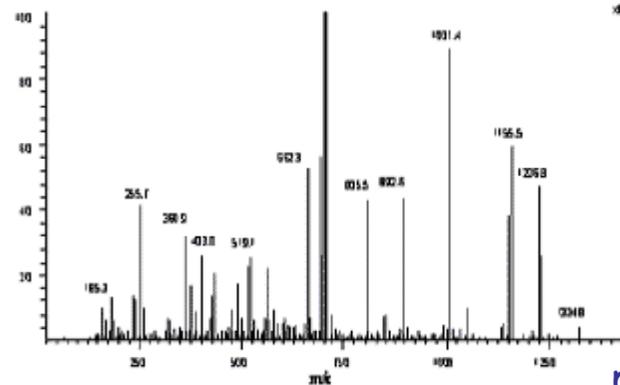


S-P-A-F-D-S-I-M-A-E-T-L-K
(protonated mass 1410.6)

<u>mass⁻</u>	<u>b-ions</u>	<u>y-ions</u>	<u>mass⁺</u>
88.1	S	P A F D S I M A E T L K	1323.6
185.2	SP	A F D S I M A E T L K	1226.4
256.3	SPA	F D S I M A E T L K	1155.4
403.5	SPAF	D S I M A E T L K	1008.2
518.5	SPAFD	S I M A E T L K	893.1
605.6	SPAFDS	I M A E T L K	806.0
718.8	SPAFDSI	M A E T L K	692.3
850.0	SPAFDSIM	A E T L K	561.7
921.1	SPAFDSIMA	E T L K	490.6
1050.2	SPAFDSIMAE	T L K	361.5
1151.3	SPAFDSIMAET	L K	260.4
1264.4	SPAFDSIMAETL	K	147.2



PREDIZIONE QUALITATIVA

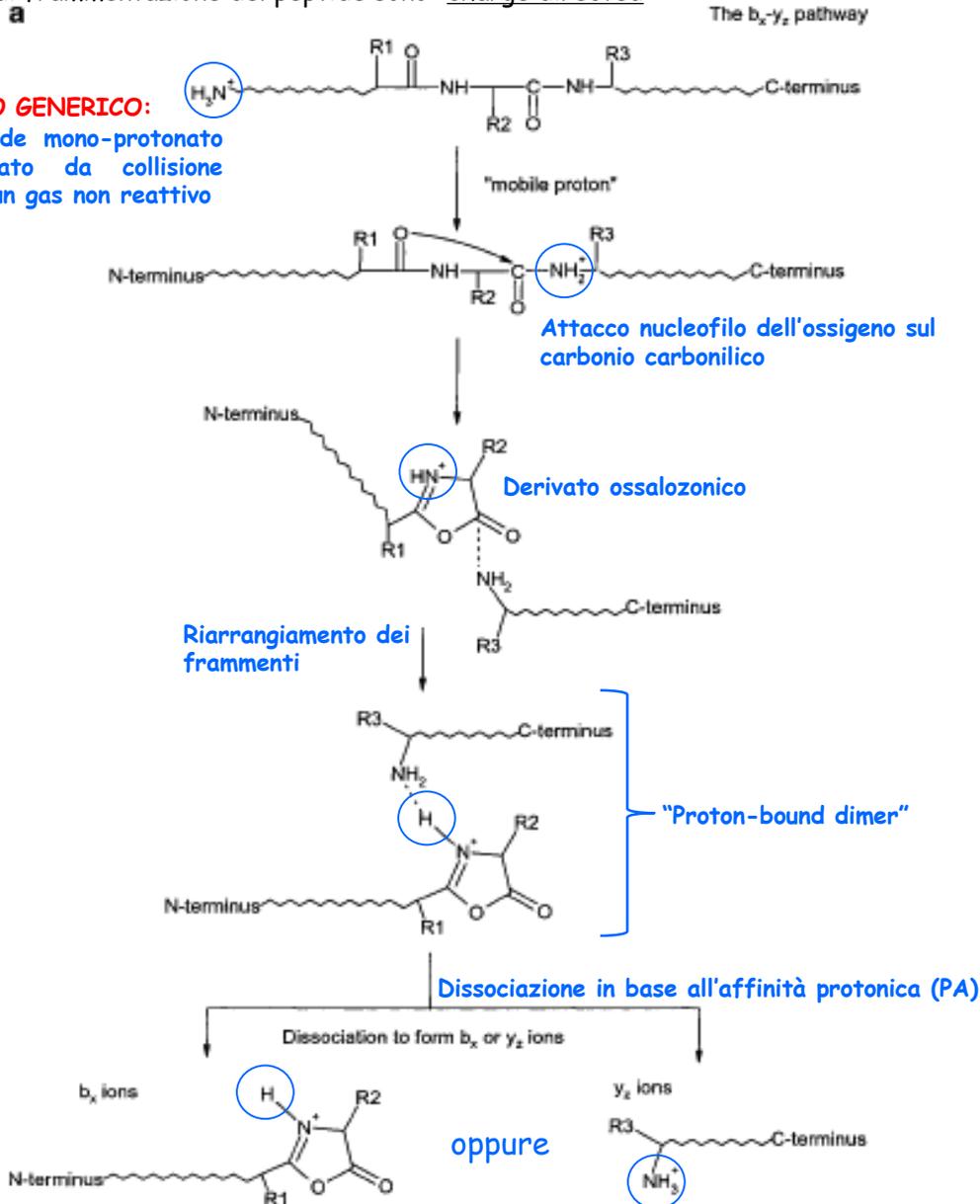


real spectrum

Quale CONFIDENZA nell'IDENTIFICAZIONE ?

IL MODELLO del "PROTONE MOBILE"

Le vie di frammentazione del peptide sono "charge directed"



SCHEME 3.

Reviewed by B.Paizs and S.Suhai, 2005

L'energia richiesta per la mobilitazione del protone dipende dalla composizione aminoacidica



Cosa accade se il peptide contiene un residuo R o K?



La catena laterale basica ha un'alta PA e "SEQUESTRA" il protone impedendone la mobilità lungo la catena principale



No Frammentazione

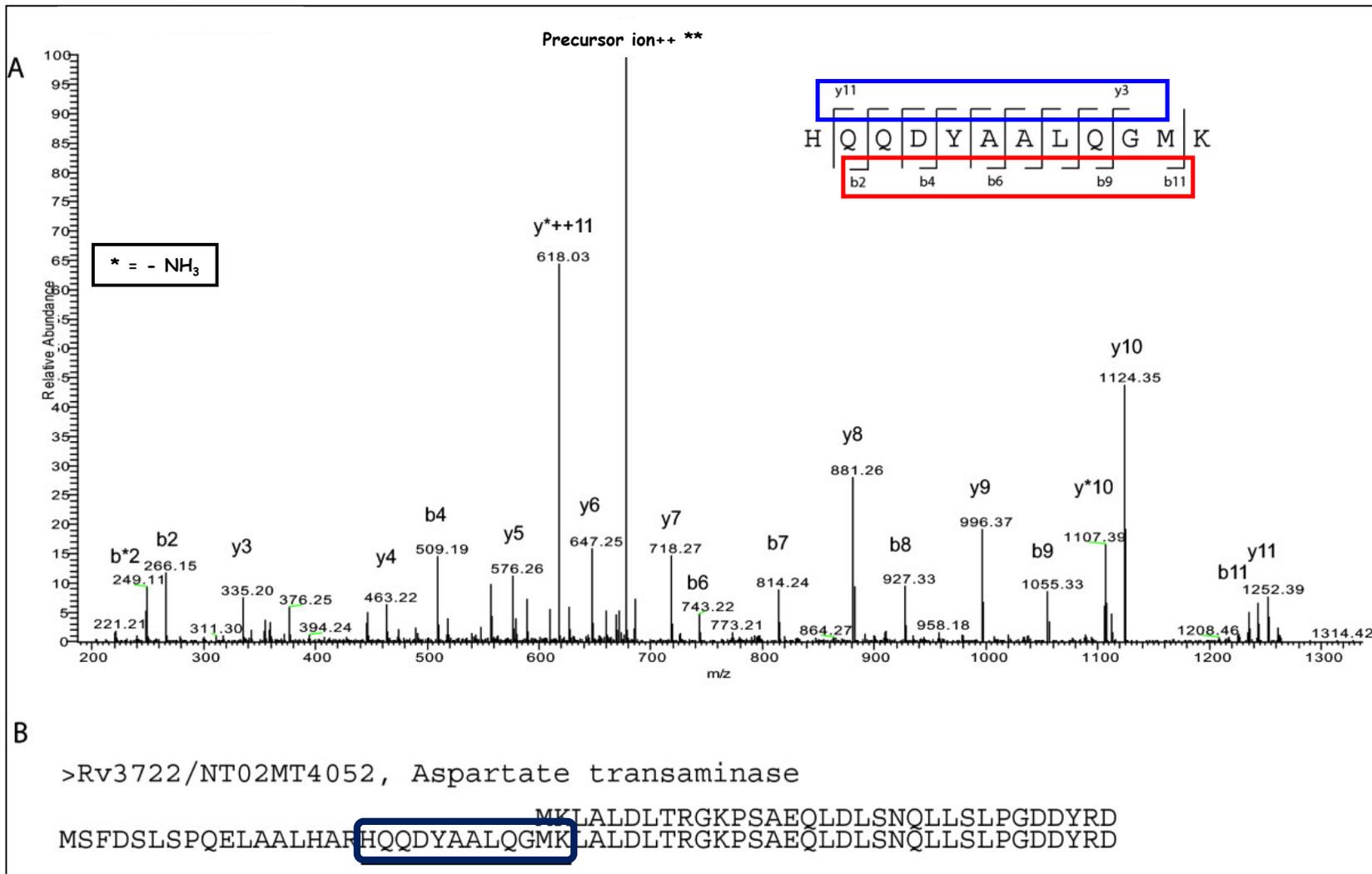
Frammentazione di PEPTIDI TRIPTICI BICARICHI

SE LA DIGESTIONE TRIPTICA E' COMPLETA

↓

1) C'è sempre 1 sola K o R, situata all'estremo C-terminale del peptide, che sequestra un protone;

2) L'altra carica situata all'N-terminale è libera, in seguito ad attivazione, di migrare e popolare i siti di protonazione energeticamente meno favoriti e indurre la frammentazione



b	Main Sequence Ions	y		
---	1	H	12	---
266.1248	2	Q	11	1252.5990
394.1833	3	Q	10	1124.5405
509.2103	4	D	9	996.4819
672.2736	5	Y	8	881.4550
743.3107	6	A	7	718.3916
814.3478	7	A	6	647.3545
927.4319	8	L	5	576.3174
1055.4905	9	Q	4	463.2333
1112.5119	10	G	3	335.1748
1243.5524	11	M	2	278.1533
---	12	K	1	147.1128

Identificazione proteica a partire dallo Spettro di Frammentazione di un peptide triptico

PEPTIDE Mass Sequencing

Software per l'identificazione proteica tramite mass sequencing

Spettro di frammentazione



- m/z (valori asse x)
- intensità (valori asse y)
- stato di carica (assegnato dopo deconvoluzione)



File in formato
.mgf



Ricerca con **Mascot MS/MS ion search**

http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS

Software per *de novo* sequencing

Peptov

<http://proteomics.ucsd.edu/LiveSearch/>

Software per la predizione dei frammenti sulla base della sequenza

Protein Prospector

<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct>

<http://expasy.org/tools/>

PEPTIDE Mass Sequencing

MASCOT MS/MS Ions Search

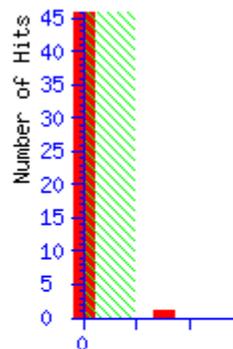
Your name	<input type="text"/>	Email	<input type="text"/>
Search title	<input type="text"/>		
Database(s)	<input type="checkbox"/> Fungi_EST <input type="checkbox"/> Environmental_EST <input checked="" type="checkbox"/> SwissProt <input type="checkbox"/> NCBInr <input type="checkbox"/> contaminants <input type="checkbox"/> ePAD	Enzyme	Trypsin <input type="button" value="v"/>
		Allow up to	3 <input type="button" value="v"/> missed cleavages
		Quantitation	None <input type="button" value="v"/>
Taxonomy	All entries <input type="button" value="v"/>		
Fixed modifications	<input type="button" value="v"/> --- none selected --- <input type="button" value="v"/>	<input type="button" value=">"/> <input type="button" value="<"/>	<input type="button" value="v"/> <ul style="list-style-type: none"> NIPCAM (C) Oxidation (HW) Oxidation (M) Phospho (ST) Phospho (Y) Propionamide (C) Pyridylethyl (C) Pyro-carbamidomethyl (N-term C) Sulfo (S) Sulfo (T) Sulfo (Y) <input type="button" value="v"/>
	Display all modifications <input type="checkbox"/>		
Variable modifications	<input type="button" value="v"/> --- none selected --- <input type="button" value="v"/>	<input type="button" value=">"/> <input type="button" value="<"/>	
Peptide tol. ±	1.2 <input type="button" value="v"/> Da <input type="button" value="v"/>	# ¹³ C	0 <input type="button" value="v"/>
		MS/MS tol. ±	0.6 <input type="button" value="v"/> Da <input type="button" value="v"/>
Peptide charge	2+ <input type="button" value="v"/>	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Data file	<input type="text"/>	<input type="button" value="Sfoggia..."/>	
Data format	Mascot generic <input type="button" value="v"/>	Precursor	<input type="text"/> m/z
Instrument	ESI-TRAP <input type="button" value="v"/>	Error tolerant	<input type="checkbox"/>
Decoy	<input type="checkbox"/>	Report top	AUTO <input type="button" value="v"/> hits
	<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>

MATRIX Mascot Search Results *SCIENCE*

User : eli
 Email : elisa.maurizio@gmail.com
 Search title : 090116-1
 MS data file : C:\Documents and Settings\Administrator\Desktop\MSMS A1a DNAPK mgf\A1a DNAPK 090116-1.mgf
 Database : NCBI nr 20090109 (7637019 sequences; 2625831693 residues)
 Taxonomy : Homo sapiens (human) (217656 sequences)
 Timestamp : 16 Jan 2009 at 10:55:57 GMT
 Protein hits : [gi|22208967](#) high mobility group AT-hook 1 isoform a [Homo sapiens]
 [gi|306873](#) high mobility group protein
 [gi|119624172](#) high mobility group AT-hook 1, isoform CRA_f [Homo sapiens]
 [gi|169163266](#) PREDICTED: similar to high mobility group protein [Homo sapiens]

Probability Based M

Ions score is $-10 \cdot \log(P)$
 Individual ions scores >
 Protein scores are derive



1. [gi|22208967](#) Mass: 11669 Score: 573 Queries matched: 51 emPAI: 24.99
 high mobility group AT-hook 1 isoform a [Homo sapiens]

Check to include this hit in error tolerant search

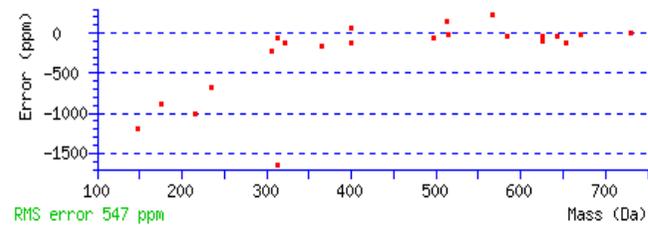
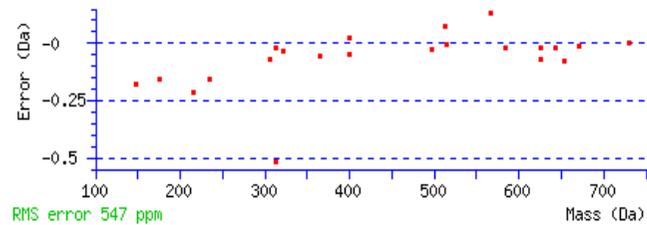
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 66	353.2390	704.4634	704.3453	0.1181	1	26	5	1	K.DGTEKR.G
<input checked="" type="checkbox"/> 97	407.2140	812.4134	812.3430	0.0705	0	30	2.3	1	K.TTTTPGR.K + Phospho (ST)
<input checked="" type="checkbox"/> 98	409.2600	816.5054	816.4341	0.0713	0	53	0.012	1	K.SSQPLASK.Q
107	430.8850	859.7554	860.4716	-0.7161	1	(22)	9.7	2	R.KTTTPGR.K
<input checked="" type="checkbox"/> 116	449.1980	896.3814	896.4004	-0.0190	0	(33)	1.1	1	K.SSQPLASK.Q + Phospho (ST)
122	471.2080	940.4014	940.4379	-0.0365	1	(24)	10	4	R.KTTTPGR.K + Phospho (ST)
123	941.4360	940.4287	940.4379	-0.0092	1	(9)	2.9e+02	6	R.KTTTPGR.K + Phospho (ST)
<input checked="" type="checkbox"/> 125	471.2420	940.4694	940.4379	0.0315	1	35	0.72	1	R.KTTTPGR.K + Phospho (ST)
133	511.2170	1020.4194	1020.4042	0.0152	1	(18)	35	2	R.KTTTPGR.K + 2 Phospho (ST)
<input checked="" type="checkbox"/> 137	532.2550	1062.4954	1062.4634	0.0320	0	38	0.36	1	K.EPSEVTPK.R + Phospho (ST)
<input checked="" type="checkbox"/> 138	532.2800	1062.5454	1062.4634	0.0820	0	(35)	0.71	1	K.EPSEVTPK.R + Phospho (ST)
146	364.2130	1089.6172	1089.5414	0.0757	2	14	92	2	K.QEKDGTEKR.G
<input checked="" type="checkbox"/> 150	560.2950	1118.5754	1117.6204	0.9551	2	(24)	9.5	1	K.TRKTTTPGR.K
153	597.8320	1193.6494	1193.5918	0.0577	1	16	62	2	K.TTTTPGRKPR.G + Phospho (ST)
<input checked="" type="checkbox"/> 154	400.2380	1197.6922	1197.5867	0.1055	2	(15)	75	1	K.TRKTTTPGR.K + Phospho (ST)
<input checked="" type="checkbox"/> 155	599.8560	1197.6974	1197.5867	0.1107	2	31	2	1	K.TRKTTTPGR.K + Phospho (ST)
<input checked="" type="checkbox"/> 169	661.8640	1321.7134	1321.6867	0.0267	2	31	2.3	1	R.KTTTPGRKPR.G + Phospho (ST)
<input checked="" type="checkbox"/> 183	501.5930	1501.7572	1501.6661	0.0911	1	(23)	14	1	M.SESSKSSQPLASK.Q + Phospho (ST)
<input checked="" type="checkbox"/> 184	751.8860	1501.7574	1501.6661	0.0914	1	51	0.024	1	M.SESSKSSQPLASK.Q + Phospho (ST)
<input checked="" type="checkbox"/> 193	797.4990	1592.9834	1592.8886	0.0948	1	62	0.0013	1	R.KQPPVSPGTALVGSQK.E
<input checked="" type="checkbox"/> 197	558.6520	1672.9342	1672.8549	0.0792	1	(29)	3.2	1	R.KQPPVSPGTALVGSQK.E + Phospho (ST)
<input checked="" type="checkbox"/> 198	837.4830	1672.9514	1672.8549	0.0965	1	(42)	0.16	1	R.KQPPVSPGTALVGSQK.E + Phospho (ST)

Monoisotopic mass of neutral peptide Mr(calc): 816.4341

Ions Score: 53 **Expect:** 0.012

Matches : 22/78 fragment ions using 27 most intense peaks ([help](#))

#	b	b++	b+	b+++	b0	b0++	Seq.	y	y++	y+	y+++	y0	y0++	#
1	88.0393	44.5233			70.0287	35.5180	S							8
2	175.0713	88.0393			157.0608	79.0340	S	730.4094	365.7083	713.3828	357.1951	712.3988	356.7030	7
3	303.1299	152.0686	286.1034	143.5533	285.1193	143.0633	Q	643.3774	322.1923	626.3508	313.6790	625.3668	313.1870	6
4	400.1827	200.5950	383.1561	192.0817	382.1721	191.5897	P	515.3188	258.1630	498.2922	249.6498	497.3082	249.1577	5
5	513.2667	257.1370	496.2402	248.6237	495.2562	248.1317	L	418.2660	209.6366	401.2395	201.1234	400.2554	200.6314	4
6	584.3039	292.6556	567.2773	284.1423	566.2933	283.6503	A	305.1819	153.0946	288.1554	144.5813	287.1714	144.0893	3
7	671.3359	336.1716	654.3093	327.6583	653.3253	327.1663	S	234.1448	117.5761	217.1183	109.0628	216.1343	108.5708	2
8							K	147.1128	74.0600	130.0863	65.5468			1



NCBI BLAST search of [SSOPLASK](#)

(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)

Other BLAST [web gateways](#)

All matches to this query

Score	Mr(calc):	Delta	Sequence
53.5	816.4341	0.0713	SSOPLASK
31.8	816.4453	0.0601	SSKPD RK
29.2	816.4341	0.0713	AGPSALSSK