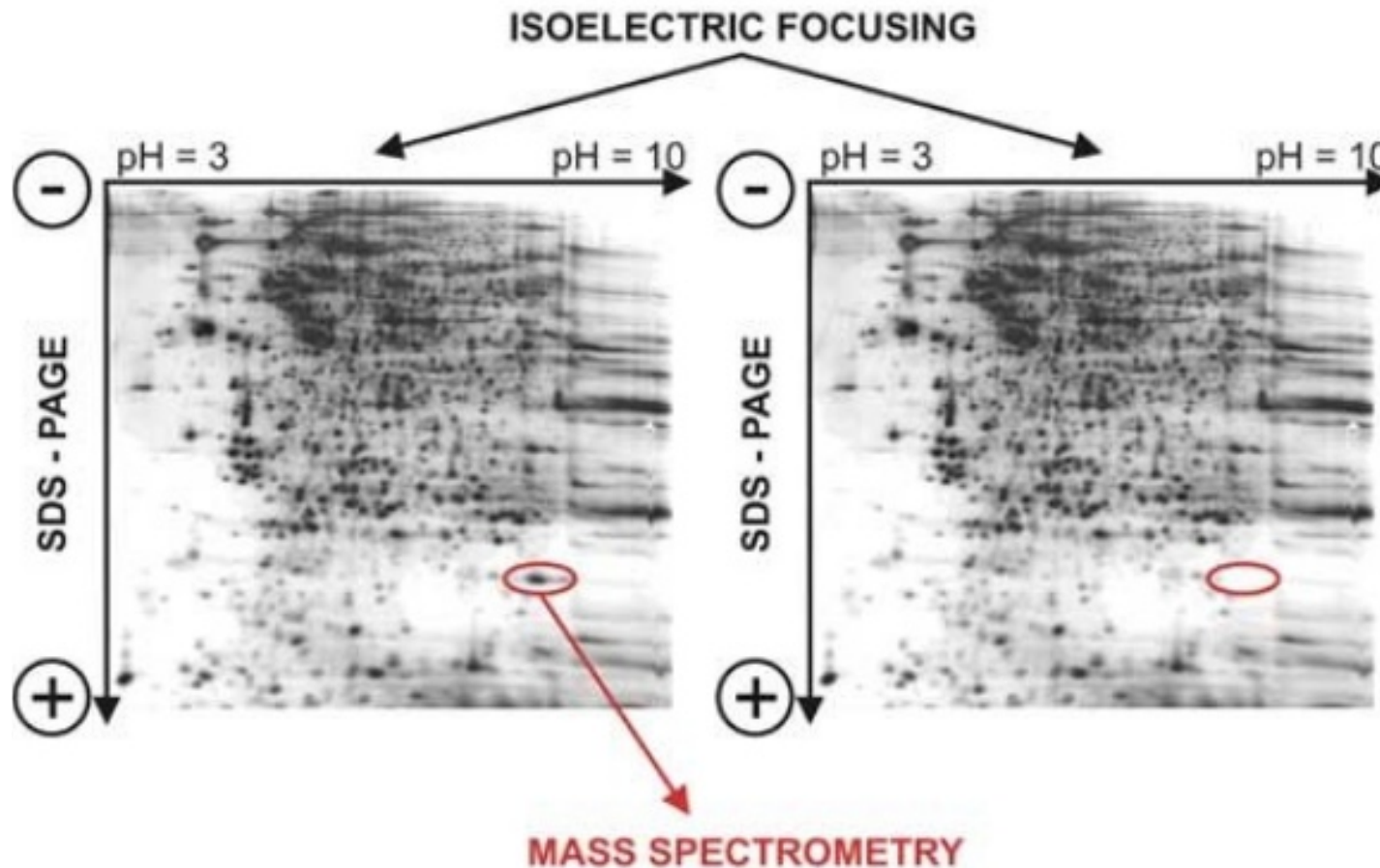


## Dallo spot all'identificazione

I passaggi analitici effettuati fino a questo punto ci hanno consentito di effettuare delle analisi elettroforetiche e di individuare delle differenze tra le condizioni prese sotto esame (i.e. controllo vs. stato patologico). Una volta rilevate le differenze, il punto è ottenere l'identità di queste differenze, ovvero assegnare un nome alle proteine che vediamo essere alterate. Per ottenere questa informazione dobbiamo isolare la proteina d'interesse, ricavare i peptidi e, tramite la spettrometria di massa e alcuni tool bioinformatici ricavare la loro identità.



## Dallo spot all'identificazione

Visualizzazione dei gel  Comparazione pattern di espressione proteica



Differenze

Qualitative (1) - presenza/assenza

Qualitative (2) - PTMs - spostamento di spot - profili tipici di determinate PTMs

Quantitative - variazioni nelle quantità

**Quali sono queste proteine?  
IDENTITA?**

Recupero degli spots



Decolorazione



Digestione endoproteolitica "in gel"



Identificazione mediante MS

# Recupero degli spots

## Manuale

Si effettua questa procedura in modo manuale quando il **numero di spot da prelevare non è elevato** e in genere quando la **processività richiesta non è così elevata** da giustificare l'utilizzo di sistemi robotizzati.

Un modo molto semplice di procedere è quello di collocare il gel colorato (argentica o CBB) su di un transilluminatore e utilizzando dei puntali da P200 tagliati in punta (in modo da avere un'apertura del diametro di circa 2-3 mm) incidere e prelevare gli spot d'interesse.

Attenzione alle possibili **contaminazioni da cheratine** dell'operatore.

Lavorare in ambienti puliti, sotto cappa a flusso, indossando camice e guanti.

Utilizzare materiale pulito per evitare contaminazioni proteiche.

**NB: queste attenzioni nei confronti della pulizia del materiale sono necessarie in qualsiasi operazioni che si effettua nello svolgimento di un'analisi bidimensionale, dalla preparazione del campione fino all'identificazione in MS**

## Robotizzata

In laboratori dedicati esclusivamente all'identificazione di proteine da gel bidimensionali tutte le operazioni tendono ad essere effettuate in modo robotizzato (anche i passaggi di colorazione e decolorazione dei gels).

In questi casi è possibile utilizzare dei sistemi integrati che prelevano gli spot, provvedono alla loro decolorazione, digestione in gel e alla analisi MS senza che l'operatore intervenga in modo diretto.

Questo velocizza di molto il processo d'identificazione e minimizza problemi di contaminazioni.

Attenzione alla deformazione dei gels dal momento dell'acquisizione dell'immagine al prelevamento degli spot: gel tendono a variare le loro dimensioni con il tempo.

Gel deve essere supportato su di una matrice indeformabile

=> dal momento della rilevazione al momento dello spot picking nessuna variazione dimensionale altrimenti lo spot picker robotizzato, che si basa su delle coordinate ricavate dall'immagine digitale, rischierebbe di prelevare pezzi di gel non corrispondenti all'effettiva posizione dello spot.

# Decolorazione spots

La decolorazione degli spot è necessaria per poter in seguito analizzare mediante MS i peptidi. Con i coloranti legati ciò non sarebbe possibile in quanto le molecole di colorante legato altererebbero la ionizzazione del peptide ed il suo peso molecolare, rendendone impossibile l'analisi in MS

## Spots da CBB

- 1) Spots proteici trasferiti in eppendorf;
- 2) decolorati con una soluzione:  
1/1 di H<sub>2</sub>O /CH<sub>3</sub>CN resa 2.5 mM BA  
BA: bicarbonato d'ammonio  
- effettuare diversi cambi di solvente fino quando il CBB non è stato completamente rimosso;
- 3) disidratare lo spot con CH<sub>3</sub>CN (2 x 10').

## Spots da Silver staining

- 1) Spots proteici trasferiti in eppendorf;
- 2) decolorati con una soluzione 1/1 di:  
Potassio ferricianuro 30 mM e Tiosolfato di Sodio 100 mM (circa 30' possono essere sufficienti - controllare che lo spot non sia più colorato);
- 3) effettuare diversi lavaggi in acqua per eliminare la soluzione di decolorazione. (circa 2h);
- 4) incubare 20' in bicarbonato d'ammonio 25 mM;
- 5) disidratare lo spot con CH<sub>3</sub>CN (2 x 10').

L'acetonitrile è più volatile dell'acqua pertanto e può essere eliminato dallo spot con maggiore facilità, o lasciando semplicemente all'aria lo spot oppure mettendo l'eppendorf in un evaporatore.

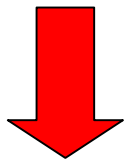
Lo spot completamente privo di solvente è ora pronto per essere sottoposto a digestione endoproteolitica.

Essendo completamente secco, al pari di quello che avviene per una strip da prima dimensione (IEF), si comporta come una **spugna** se messo a contatto con una soluzione acquosa. In questo modo **l'enzima proteolitico può essere portato all'interno dello spot** in modo da operare quella che viene definita l'**in gel digestion**.

## Digestioni endoproteolitica

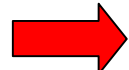
Scopo: ottenere frammenti peptidici dalla proteina d'interesse in modo da procedere alla sua identificazione:  
 a) attraverso **peptide mass fingerprinting** - b) attraverso **MS peptide sequencing**.

<i>Method of protein cleavage</i>	<i>Site of cleavage</i>	<i>Exception</i>	<i>pH range</i>
Trypsin	C-terminus of R-X, K-X	If X = P	7-9
Endoproteinase Glu-C (V8-DE)	C-terminus of E-X, D-X	If X = P	4-8
Chymotrypsin	C-terminus of F, Y, W, L, I, V, M	If X = P	7.5-8.5
Endoproteinase Lys-C	C-terminus of lysine, K-X	If X = P	8.5-8.8
Arg-C	C-terminus of arginine, R-X	If X = P	7.5-8.5
Elastase	Not very specific. C-terminal side of G, A, S, V, L and I.		8-8.5
Pepsin	C-terminus of F, L and E		2-4
Pronase	Pronase is a mixture of endo- and exo-proteinases. It cleaves almost any peptide bond.		7-8, dependent on proteases present



Digestione (i) condotta in modo da avere il **minor numero possibile di siti di taglio mancanti (missed cleavage)** ovvero in modo da avere una **digestione completa** e, ove possibile, (ii) effettuata utilizzando **tamponi volatili**, in modo da poterli eliminare alla fine della reazione di digestione (vedasi più avanti compatibilità con analisi LC-MS).

# In gel digestion - (di solito con tripsina)



Tripsina sciolta a 40 ng/ $\mu$ L in 50 mM ammonio bicarbonato (**tampone volatile**);



Allo spot viene aggiunto 10  $\mu$ l della soluzione di tripsina (400 ng - 0.4  $\mu$ g - eccesso di tripsina considerando le le quantità di proteina contenute solitamente in uno spot da bidimensionale - i.e meno di 100 ng). **La proteina è stericamente meno accessibile rispetto a quando è in soluzione e per questo è richiesta più tripsina - questione di efficienza;**



Dopo 45' la soluzione di tripsina che non è stata assorbita dal gel viene rimossa e sostituita con lo stesso tampone senza però l'enzima - effettuato in ghiaccio per minimizzare autolisi della tripsina;



Digestione lasciata procedere 12h - in modo da avere una digestione completa con meno siti di taglio mancanti i.e meno siti di missed cleavage;



Si aggiunge 20  $\mu$ L tampone e si sonica per 10' (favorire l'estrazione dei peptidi), si aggiunge 20  $\mu$ L CH<sub>3</sub>CN e si sonica altri 10' e alla fine si recupera il surnatante;



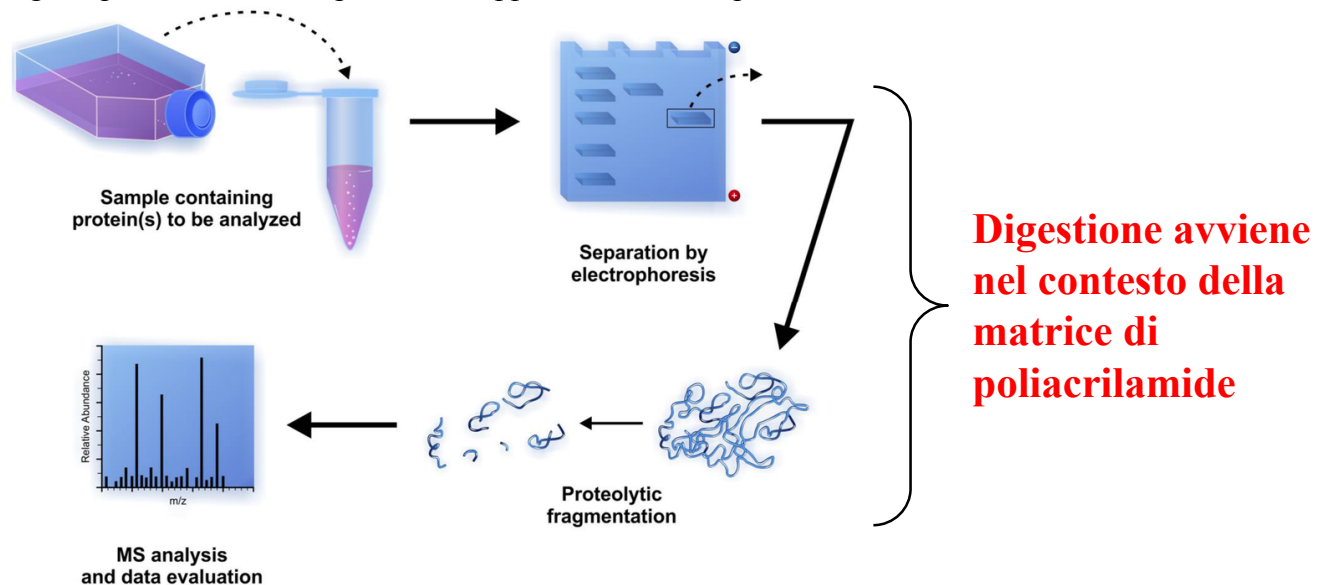
Si effettuano altre 2 estrazioni con la medesima procedura utilizzando prima H<sub>2</sub>O/CH<sub>3</sub>CN (1:1) - 2.5% TFA e successivamente solo CH<sub>3</sub>CN.



Si uniscono tutti i supernatanti raccolti nei quali sono eluiti i peptidi.

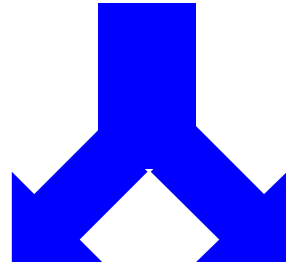


Peptidi solitamente vengono portati a secco e ripresi in un opportuno solvente per le analisi di LC-MS



## Strategie identificative

Le proteine vengono solitamente identificate a partire dai peptidi secondo due strategie



### Peptide Mass Fingerprinting

Sostanzialmente basato su analisi di tipo **MALDI-TOF**. Se i peptidi sono stati digeriti con un tampone volatile ed eluiti con soluzioni prive di sali possono essere analizzati direttamente (o dopo evaporazione se il volume con il quale sono stati eluiti era troppo elevato).

In alternativa, se la soluzione contenente i peptidi presenta sali, i peptidi possono essere purificati mediante Zip-Tip con fasi inverse. (vedasi più avanti)

### Peptide Sequencing

Se basato su analisi di tipo **MALDI-TOF/TOF**, valgono le indicazioni riportate sulla destra.

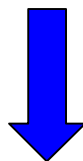
Se l'identificazione è basata su metodiche **LC-ESI-MS/MS** i peptidi vengono iniettati in colonna cromatografica di tipo RP (reversed phase) ed è necessario che la soluzione in cui sono disciolti sia compatibile con questo tipo di analisi (vedasi più avanti)

**NB:** ritorna il concetto evidenziato in precedenza circa la compatibilità dei reagenti con le analisi che seguono.

# Approcci identificativi basati sulla spettrometria di massa (MS)

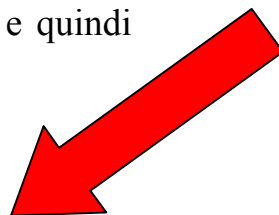
Per ragioni legate alla metodica utilizzata non è possibile ottenere sequenze dalle proteina intere (a dir il vero è difficile ottenere sequenze già con peptidi di circa 20-30 residui aminoacidici). La sola metodica che consente di ottenere informazioni di sequenza a partire da molecole proteiche intere è l'ETD ma il suo utilizzo è limitato ad un particolare tipo di strumento (trappole ioniche) e quindi trova non largo impiego

PROTEINA



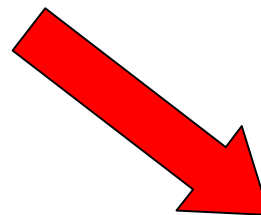
PEPTIDI

Il solo peso molecolare di una proteina non è sufficiente per fornire un dato identificativo. Le proteine infatti recano solitamente una serie eterogenea di modificazioni post-traduzionali che porta il peso molecolare a discostarsi dal semplice valore numerico che si può ricavare dalla somma dei pesi molecolari dei singoli residui aminoacidici.



## MS sequencing

Frammentazioni controllate ottenute mediante l'utilizzo di specifici spettrometri di massa (spettrometri di massa provvisti di una cella di collisione) può portare all'ottenimento di informazioni relative alla sequenza aminoacidica del peptide sottoposto a frammentazione. Dalla sequenza aminoacidica è possibile risalire all'identità della proteina da cui deriva il peptide mediante utilizzo di software di ricerca in banche dati dove sono depositate le informazioni (sequenze) delle proteine.

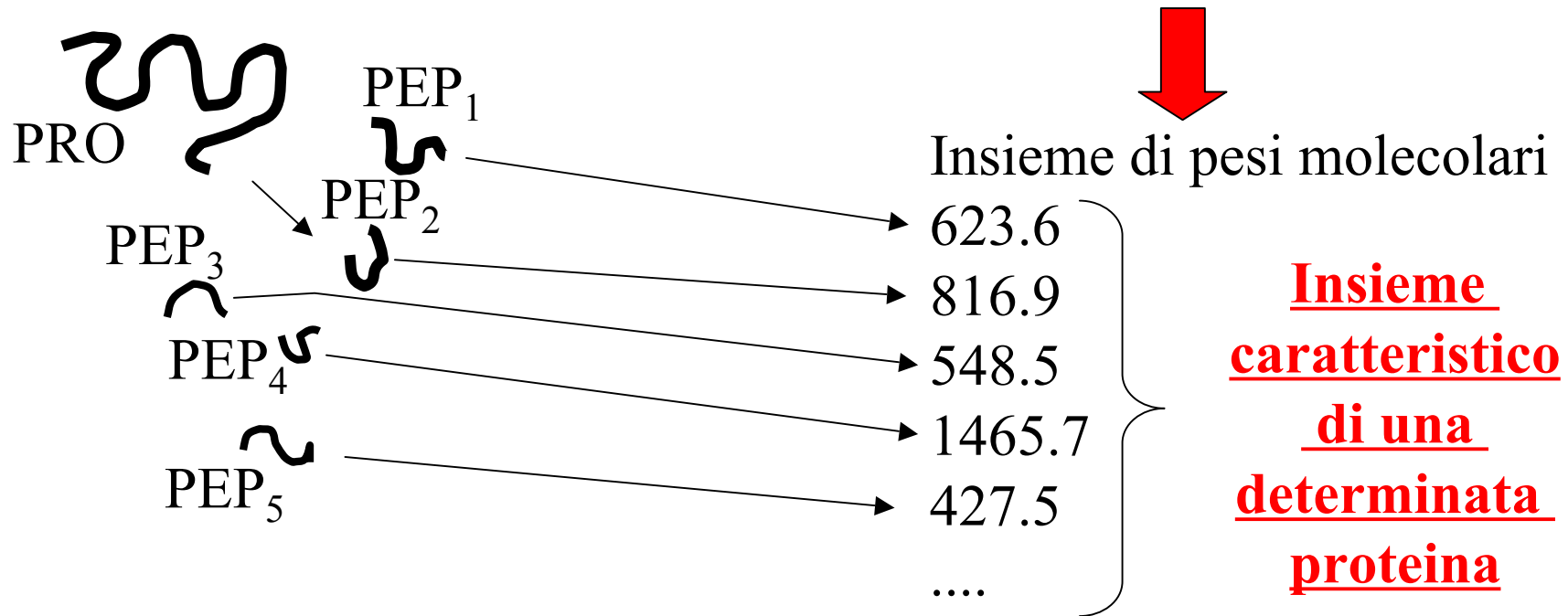


## Peptide mass fingerprinting

Da ogni proteina si ottiene una specifica serie di peptidi ciascuno dei quali è caratterizzato da un preciso valore di peso molecolare. La combinazione di pesi molecolari ottenuta può essere utilizzata per risalire all'identità della proteina. Trattandosi di peptidi, quindi di porzioni ristrette della proteina d'origine, è probabile che su di un elevato numero di essi non cadano PTMs o che il numero di queste sia limitato, tanto da poter esser considerate dai software di ricerca in banca dati.

# PEPTIDE MASS FINGERPRINTING

Digestione enzimatica  
SPOT → PEPTIDI → MS → Peso Molecolare dei singoli peptidi



PROTEINA: caratterizzata da una sua particolare e unica sequenza aa

↓  
PEPTIDE: caratterizzato da una sua particolare e unica sequenza aa

↓  
PEPTIDE: sequenza aa specifica => **peso molecolare specifico**



# PEPTIDE MASS FINGERPRINTING

Una proteina, digerita da un enzima proteolitico noto - è nota quindi la specificità di taglio - da origine ad una serie di peptidi e ciascuno di essi caratterizzato da uno specifico peso molecolare.

Anche se peptidi che presentano la stessa composizione aa ma sequenza diversa presentano pesi molecolari uguali è statisticamente poco probabile che da proteine diverse si generino più peptidi che presentano la stessa composizione aa ma diversa sequenza.

Questo significa che l'insieme dei pesi molecolari dei peptidi generatisi dalla digestione endoproteolitica di una proteina risulta essere in pratica una

**IMPRONTA DIGITALE**



**IDENTIFICAZIONE**

L'identificazione risulterà più certa quanto maggiore è l'**accuratezza** con la quale viene determinato il **peso molecolare** dei singoli peptidi e quanto maggiore è la **copertura** della intera **sequenza** della proteina di partenza  
(**sequence coverage**)

**Accuratezza** => **risoluzione strumentale e calibrazione strumentale**

**Sequence coverage** => **sensibilità strumentale ed efficienza nella in gel digestion e recupero dei peptidi**

# PEPTIDE MASS FINGERPRINTING

**"OK, I've got a match, but is it the *right* match?"**

**Confidence** in a peptide mass fingerprint result may come from having **independent supporting evidence**. For example, if the analyte originated from a spot at approximately 40 kDa on a 2D gel separation of yeast proteins, then the anticipated result of a peptide mass fingerprint is a 40 kDa yeast protein. If the top scoring protein fits this expectation, the search is deemed "successful". If the top scoring match is a 200 kDa protein from a different species, the initial reaction is likely to be that the search has "failed".

While this is a reasonable approach, Mascot provides additional guidance in the form of a significance level. By default, the significance level is set at 5%. That is, if the score for a particular match exceeds the significance level, there is less than a 1 in 20 chance that the observed match is a random event.

If the score is substantially **above the significance level**, look carefully before dismissing the result as spurious. Conversely, if the score is **below** the significance level, examine the match **sceptically**.

Taxonomy: unless the genome of the species of interest is completely sequenced, there is no guarantee that the true sequence of the analyte protein is actually present in the database. If it is missing, then high scoring matches from other species are of interest because they are likely to be homologous to the unknown.

It is the uncertainty in the mass of the intact protein which is the Achilles heel of a peptide mass fingerprint. This uncertainty is unavoidable, even when an accurate experimental mass for the intact protein is available, because it is unlikely that the mass of the expressed and processed protein will be exactly the same as that of the sequence entry in the protein database.

A peptide mass fingerprint can only provide the **statistically most probable identification**. This is a great step forward over simply counting peptide mass matches, which can only work when a ceiling is placed on the intact protein mass. Otherwise, the mega-proteins always come out top of the list due to random matches. Unfortunately, even with an ideal scoring algorithm, there may be insufficient matching mass values for a confident identification without making assumptions about the intact protein mass or the species.

One method of improving the specificity of a peptide mass fingerprint was first proposed by Peter James [[James, 1994](#)]. Simply do **additional digests using different proteases**. Seeing the same protein with a high score in two independent digests provides a similar degree of confidence to seeing multiple peptide matches in an MS/MS ions search.

## PTMs

**Table 7.1: Post-translational modifications and corresponding mass variations.**

Post-translational modification	Mass difference (Da)
Methylation	14.03
Propylation	42.08
Sulfation	80.06
Phosphorylation	79.98
<b>Glycosylations by:</b>	
Deoxyhexoses (Fuc)	146.14
Hexosamines (GlcN, GalN)	161.16
Hexoses (Glc, Gal, Man)	162.14
N-Acetylhexosamines (GlcNAc, GalNAc)	203.19
Pentoses (Xyl, Ara)	132.12
Sialic acid (NeuNAc)	291.26
Reduction of a disulfide bridge	2.02
Carbamidomethylation	57.03
Carboxymethylation	58.04
Cysteinylation	119.14
Ethylpyridylation	105.12
Acetylation	42.04
Formylation	28.01
Biotinylation	226.29
Farnesylation	204.36
Myristoylation	210.36
Pyridoxal phosphate Schiff condensation	231.14
Stearoylation	266.47
Palmitoylation	238.41
Lipoylation	188.30
Carboxylation of Asp or Glu	44.01
Deamidation of Asn or Gln	0.98
Hydroxylation	16.00
Met Oxidation	16.00
Proteolysis of a peptide bond	18.02
Deamination from Gln to pyroglutamic	-17.03

← Arg, Lys

← Ser, Thr, Tyr

← Cys

← Lys

Più comuni  
modificazioni  
post-traduzionali

Tutte queste modificazioni post-traduzionali alterano il peso molecolare della proteina, alcune di esse ne modificano anche il punto isoelettrico: Acetilazione e Fosforilazione  
 Acetilazione: rimuove una carica positiva  
 Fosforilazione: introduce due cariche negative

# PEPTIDE FINGER PRINTING SOFTWARE - Mascot

Alcune considerazioni sull'opzione PTMs.

Esempio:

Se in un'analisi di PMF viene inserita l'opzione della fosforilazione come possibile modificazione post-traduzionale, questo comporta che automaticamente viene ad essere amplificato il numero di masse con i quali noi facciamo la nostra analisi:

Ipotesi: 10 pesi molecolari inseriti - fosforilazione come possibile modificazione post-traduzionale => Ciascun peptide generato in modo virtuale dalle banche dati viene considerato come potenzialmente fosforilabile se ha nella sua sequenza S, T, o Y (si può scegliere anche se S/T o Y fosforilazione).

Il numero di possibili match in questo caso aumenta e si può verificare il caso che un valore da me trovato sperimentalmente che non è effettivamente modificato vada a combaciare con un valore di un peptide fosforilato nella digestione virtuale => falso positivo. In pratica si aumentano le possibilità di trovare un match.

Nella pratica di laboratorio quello che solitamente si fa è fare delle analisi bioinformatiche in serie, partendo prima con una situazione in cui non vengono incluse PTMs nella ricerca.

Sulla base di quanto trovato da questa prima analisi, si cerca di assegnare ulteriori peptidi eventualmente modificati.

Una buona norma è quella di procedere con livelli di complessità crescenti.

Solitamente una analisi di PMF fornisce un'identificazione anche senza includere PTMs.

Questo non toglie però che si possono avere informazioni maggiori dalle proprie analisi.