POLITECNICO DI TORINO

Collegio di Ingegneria Chimica e dei Materiali

Corso di Laurea Magistrale in
Ingegneria Chimica e dei Processi Sostenibili

Tesi di Laurea Magistrale

A CE-MALDI-MS method for bioanalysis
Focus on the glycosylation of osteopontin

Relatore

prof. Francesca Bosco

Candidato

Aldo Cingolani

Marzo 2018

Abstract

The quest for biomarkers with clinical applications is a major drive for the thorough study of proteins and their post translational modifications. In this work, an investigation of the glycosylation of osteopontin was conducted. In the first part of the work, a CE-MALDI-TOF-MS hyphenation was optimized for fractionation and analysis of biomolecular species, with the help of a robotic arm and a positioning table. A proof of concept was developed on a three-protein-mixture. In the second part of the work, the glycosylation of osteopontin was investigated through glycopeptides and released glycans, by MALDI-TOF mass spectrometry. The research was focused on an acidic core-1 O-glycan structure weighing 965.3 Da, hypothesized by previous studies. No evidence was found for the targeted glyco-moiety in the glycopeptide analysis, but multiple clues for its presence were obtained by the analyses on released glycans. Furthermore, an indirect purification method for released glycans was developed, which led to a 4- to 7-fold increase of the signal to noise ratio of relevant peaks. The work is intended to provide better knowledge on the post translational modifications of osteopontin for future work related to the activity of the protein in the upbuilding of the inorganic matrix of bones.

Contents

0. Riassunto

0.1 Introduzione

I sistemi biologici sono intrinsecamente complessi: miliardi di molecole ricoprono un ruolo specifico nel processo di nascita, crescita e morte delle cellule, così come nel controllo del loro corretto funzionamento. La piena comprensione dei sistemi viventi è impensabile se si ragiona in compartimenti stagni: è necessario invece accorpare informazioni ricavate da studi interdisciplinari per ottenere una visione completa. Nel recente passato, proteomica, glicomica e metabolomica (indicate comunemente come scienze "omics") hanno raggiunto un buon grado di complementarietà, permettendo una buona comprensione dei sistemi biologici tramite l'analisi di biomolecole, anche supportata dal forte progresso tecnologico/strumentale. Il più forte impatto delle strategie "omics" si è registrato nella ricerca di potenziali biomarker per la diagnosi di malattie difficilmente rilevabili. Ad esempio, la determinazione del profilo glicosidico nel siero umano ha permesso di scoprire diversi marker per il cancro. Allo stesso modo, lo studio dei glicani di alcune proteine ha fornito importanti informazioni relative a patologie del fegato Tali aree scientifiche si fondano principalmente su analisi in spettrometria di massa (MS): grazie ai recenti progressi, le piattaforme MS sono diventate strumenti imprescindibili, specialmente per le loro elevate sensibilità e accuratezza.

Questo lavoro si inquadra nell'area sopra descritta: nella caratterizzazione di proteomi e nella ricerca di nuovi biomarker, è necessario lo sviluppo di metodi analitici che assicurino la separazione e facilitino l'analisi di biomolecole in miscele complesse (*e.g.* sangue, plasma, fluido cerebrospinale ecc.). L'obiettivo del lavoro consiste nello sviluppo di un metodo che metta insieme separazione ed analisi dei glicani dell'osteopontina (OPN), una glicoproteina multifunzionale coinvolta principalmente nel metabolismo osseo. Il target iniziale era il confronto del pattern glicosidico di OPN in campioni biologici di plasma di individui sani e individui affetti da sindrome dolorosa regionale complessa (CRPS); ma le bassissime concentrazioni di OPN nei fluidi biologici, così come la mancanza di informazioni relative alla glicosilazione della proteina avrebbero reso il lavoro troppo proibitivo. Pertanto, tutte le analisi su OPN sono state condotte su proteina ricombinante, in modo da lavorare con concentrazioni più facilmente rilevabili.
La prima parte del lavoro è incentrata sulla separazione analitica di biomolecole e successiva deposizione di singole specie per analisi MS. Il metodo è ideato per future applicazioni in glicomica, ma è stato sviluppato come proof of concept su un campione contenente tre proteine (citocromo C (Cyt-C), lisozima (Lys) e α-lattoalbumina (Lact), di massa tra 12 e 15 kDa). Le proteine sono dapprima separate con elettroforesi capillare (CE), ed in seguito pressurizzate fuori dal capillare per essere depositate. L'estremità terminale del capillare viene controllata in remoto da un braccio meccanico programmato con Arduino, in modo che le singole specie di proteine possano essere depositate su spot distinti di una piastra matrix-assisted laser desorption/ionization (MALDI), e analizzate in tempo di volo (TOF). Il processo di deposizione di singole specie su singoli spot è chiamato frazionamento.
La seconda parte del lavoro è incentrata direttamente sullo studio della glicosilazione di OPN,

come potenziale biomarker per CRPS. Un primo approccio all' indagine del pattern glicosidico è stata l'analisi della proteina in seguito a digestione enzimatica con tripsina o chimotripsina, in modo da sfruttare la maggiore facilità di analisi dei peptidi in MS rispetto alle proteine intatte. Il secondo approccio scelto per lo studio del pattern glicosidico di OPN è stato l'analisi di glicani intatti in seguito a rilascio chimico (β-eliminazione con eccesso di agente riducente), purificazione e concentrazione. E' stato suggerito che OPN possieda cinque O-glicani, tutti dello stesso tipo, formati da una struttura core-1 e recanti due acidi sialici.. Tre metodi di rilascio e purificazione di zuccheri sono stati impiegati e confrontati in MALDI-TOF-MS al fine di testare se il glicano supposto sia effettivamente presente sulla backbone della proteina. Una maggiore chiarezza sulla glicosilazione di OPN, così come lo sviluppo di metodi rapidi ed efficaci per separare ed analizzare biomolecole in generale - e glicani in particolare - sono la chiave verso la comprensione dei ruoli di OPN nello sviluppo di malattie ossee difficilmente diagnosticabili.

0.1.1 Osteopontina

Il lavoro è incentrato sull'osteopontina (OPN), una glicoproteina acida multifunzionale e fortemente fosforilata che è espressa in diversi organi. L'OPN è coinvolta principalmente nel metabolismo e nella fisiologia ossee, motivo per cui è stata largamente studiata in relazione a biomineralizzazione e rimodellamento osseo in seguito a traumi. Numerosi lavori sono stati dedicati alla possibile attività di OPN nello sviluppo di osteoporosi, sclerosi multipla e artrite reumatoide, specialmente per via delle sue proprietà regolatrici di adesione, migrazione e proliferazione cellulare. Si suppone che le modifiche post-traduzionali (PTM) di OPN siano strettamente connesse con tali funzioni regolatrici, e che pertanto siano direttamente coinvolte nell'evoluzione delle patologie. In particolare, è stato suggerito che le glicosilazioni di OPN possano fungere da "interruttore" regolatore della mineralizzazione ossea. Queste ipotesi rendono OPN una interessante proteina da studiare in relazione al metabolismo osseo. L'OPN ricombinante utilizzata nel presente lavoro consiste di 298 amminoacidi per un totale di 33.7 kDa. Considerando le PTM, il peso si aggira invece intorno a 39 kDa.
La glicosilazione dell'OPN non è mai stata caratterizzata: è stato determinato che l'OPN abbia cinque siti di O-glicosilazione, tutti distribuiti in una porzione molto limitata della proteina (lungo 18 amminoacidi) ma non esistono profilazioni che ne caratterizzino le strutture. Alcuni studi hanno formulato delle ipotesi, la più completa delle quali suggerisce una struttura core-1 acida del tipo [GalNAc-Gal-NeuNAc$_2$] (Christensen et al., 2008), di massa 965.3 Da, la cui struttura è visualizzabile in figura 1.

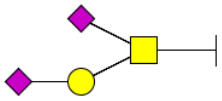

Figura 1: Struttura del glicano GalNAc-Gal-NeuNAc$_2$ proposta da Christensen *et al.*, 2008. Quadrato giallo: GalNAc; Cerchio giallo: Gal; Rombo viola: NeuNAc.

L'obiettivo della seconda parte di questo lavoro è capire se la struttura proposta da Christensen sia effettivamente presente tra le glicosilazioni di OPN.

0.1.2 Elettroforesi Capillare

L'elettroforesi è la migrazione di ioni in soluzione sotto l'azione di un potenziale esterno. CE è invece la separazione di ioni in soluzione sotto l'azione di un potenziale applicato alle estremità di un capillare, generalmente in silica. Il vantaggio di questa tecnica di separazione consiste nell'alto potere risolvente, nei minimi volumi coinvolti e nella grande versatilità. Infatti, CE

offre diverse modalità operazionali. L'elettroforesi capillare a zone (CZE), usata in questo lavoro, è la più semplice e comune. In tale tecnica, il capillare è riempito con una soluzione elettrolitica di background (BGE), e il campione viene iniettato dall'inlet del capillare per via elettrocinetica o idrodinamica. Quando viene applicato il campo elettrico, gli analiti migrano verso l'elettrodo di carica opposta – *i.e.* gli anioni si muovono verso l'anodo e i cationi verso il catodo. I soluti nel capillare vengono separati per via delle loro mobilità, dipendenti da raggio di Stokes, carica e viscosità della soluzione BGE. Un detector è posizionato lungo il capillare per registrare il passaggio dei campioni. Il segnale è processato da un software che fornisce in output l'elettroferogramma.
Una rappresentazione schematica di un setup CZE è mostrata in figura 2.

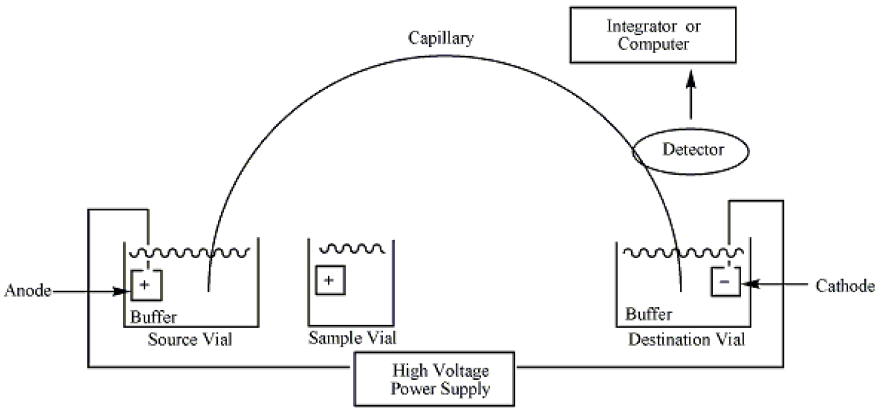

Figura 2: Schema esemplificativo di un apparato CE. Nel caso riportato in figura si tratta si uno strumento operato in modalità positiva, come evidente dal simbolo '+' sulla vial di ingresso e sulla vial contenente il campione. Fonte: *https://upload.wikimedia.org/wikipedia/commons/3/37/Capillaryelectrophoresis.png*

0.2 Metodi

0.2.1 Frazionamento

Il frazionamento di biomolecole è stato raggiunto per mezzo di un setup sperimentale composto da quattro blocchi: generatore, inlet box CE, UV detector e area di conrollo della piastra MALDI (piattaforma XY e braccio meccanico). Più in dettaglio, il sistema di accoppiamento tra CE a MALDI è formato da una piattaforma di posizionamento (su cui è fissata la piastra MALDI), che controlla due gradi di libertà (X e Y), ed un braccio meccanico che muove l'estremità terminale del capillare lungo Z. In questo modo, tramite il moto del capillare lungo l'asse verticale, o il moto della piastra sul piano XY, si ha il pieno controllo dei 3 gradi di libertà nello spazio. Durante una normale elettroforesi ("modo I" in figura 3), il capillare si estende dalla vial di inlet (nel box CE) verso il detector a luce UV (posto a metà capillare), per terminare nella vial di outlet, posizionata all'esterno del detector. Durante il frazionamento ("modo II" in figura 3), invece, l'outlet del capillare è fissato nel braccio meccanico e mantenuto entro l'area della piattaforma XY, per favorire le deposizioni. Il contatto elettrico è garantito facendo in modo che la piastra abbia lo stesso potenziale della vial di uscita. Lo schema in figura 3 mostra le principali differenze tra le due modalità.
Tutti i frazionamenti iniziano con una normale iniezione elettrocinetica di 10 secondi a un potenziale di 10 kV, seguita da separazione elettroforetica dei campioni sotto l'azione di un potenziale di 26 kV, con acetato d'ammonio 20 mM e pH 4.15 come BGE. La separazione avviene in modalità positiva in un capillare rivestito internamente con idrossipropil cellulosa. Durante la separazione, quando un picco di interesse viene rilevato dal detector (situato a metà capillare), si inseriscono i dati relativi a tempo di migrazione ed estensione del picco in un

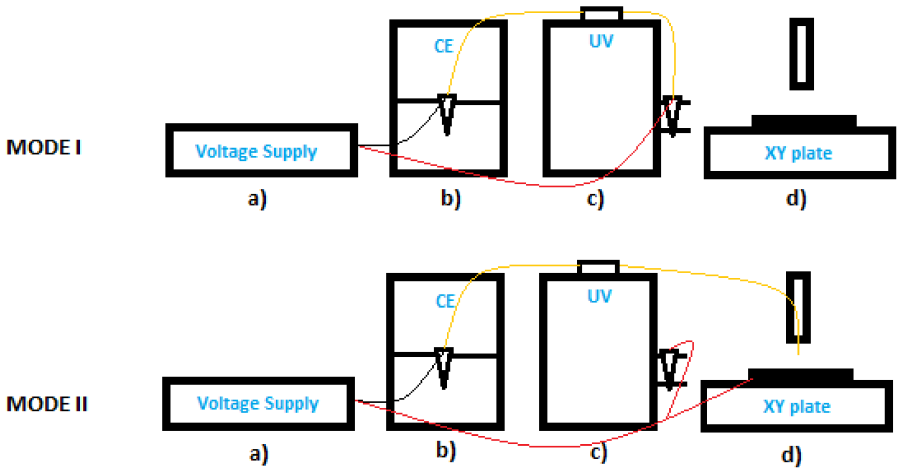

Figura 3: Confronto del sistema CE nei due setup utilizzati nel presente lavoro. Il capillare è rappresentato in arancione. Il tratto nero che parte dal generatore rappresenta il polo negativo, mentre quello rosso il positivo (ground). Modo I: il capillare termina nella vial di outlet (messa a potenziale zero). Modo II: l'outlet del capillare è fissato nel braccio robotico, che controlla la deposizione. La piastra è a potenziale zero.

foglio Excel che fornisce in output i parametri necessari a programmare la deposizione frazionata. Tali valori vengono inseriti nell'interfaccia Arduino che controlla robot e piattaforma di posizionamento. Da questo momento in poi, la deposizione viene condotta automaticamente secondo i tempi e le modalità date in input al programma. In seguito a deposizione e previa applicazione di matrice sulle specie frazionate, la piastra è pronta per essere analizzata in MS. Il metodo è stato dimostrato ed ottimizzato su un campione proteico composto da Cyt-C 0.4 μM, Lys 0.4 μM e Lact 1.2 μM.

0.2.2 Analisi di Glicopeptidi

In breve, dopo la purificazione e la concentrazione di campioni di plasma (da donatori sani) arricchiti con OPN, la proteina viene digerita con tripsina o α-chimotripsina, e i peptidi digeriti sono analizzati con MALDI-TOF-MS. Gli enzimi addetti alla digestione sono stati scelti sulla base della selettività rispetto al sito di clivaggio, in modo da ottenere una o più glicosilazioni sullo stesso peptide. In particolare, la tripsina taglia le proteine subito dopo gli amminoacidi R e K, mentre la α-chimotripsina taglia dopo Y, W, F, M, L (solo se non dopo P). Le digestioni in tripsina sono state effettuate in bicarbonato d'ammonio 40 mM in un rapporto in peso compreso tra 1:20 e 1:30 enzima:substrato. Le soluzioni sono state incubate per 17 ore a 37 °C, e spente ad alta temperatura. Le digestioni in α-chimotripsina sono state effettuate in un buffer Tris-HCl 100 mM a pH 8, contenente cloruro di calcio 10 mM, con un rapporto enzima:substrato tra 1:50 e 1:60 in peso. Le soluzioni sono state incubate per 24 ore a 37 °C, e spente portando il pH sotto 2.0 con HCl.

0.2.3 Rilascio e Purificazione di Glicani

I glicani sono stati rilasciati tramite β-eliminazione secondo tre metodi, ispirati a lavori precedenti. In particolare, sono stati usati come riferimento i lavori di Gao *et al.*, 2013, Ozcan *et al.*, 2013, Turyan *et al.*, 2014. Le reazioni sono avvenute in NaOH con $NaBH_4$ come agente riducente nei primi due casi, e in etanolammina con $(NH_4)_2CO_3$ come agente riducente nel secondo. Tutti i rilasci sono avvenuti a 42 °C per il 17 ore in campioni da 50 μL in volume contenenti 1 μL di OPN 1 mg/ml, e quenchati con HCl 6M in un bagno di ghiaccio facendo scendere il pH sotto 3.0.
I glicani rilasciati sono stati purificati in due modi:

- metodo diretto: purificazione con cartucce in carbone grafitico poroso (PGC) Thermo Scientific™ HyperSep Hypercarb™ SpinTips cartridges, 10-200 μL. Glicani affini alla fase stazionaria, dunque ritenuti, eluiti, concentrati a secco e ricostituiti a 6 μL in H_2O per analisi in MALDI-TOF-MS. L'analisi è stata condotta con acido 2,5-diidrossibenzoico come matrice.

- metodo indiretto: purificazione con cartucce Pierce® C18 Tips, 100 μL bed. Glicani non affini alla fase stazionaria, dunque non ritenuti, ma concentrati a secco e ricostituiti a 6 μL in H_2O a partire dalla soluzione privata delle specie affini alla fase stazionaria (proteine, eventuali peptidi, sali, ecc.). I glicani sono stati analizzati in MALDI-TOF-MS con acido 2,5-diidrossibenzoico come matrice.

0.3 Risultati

0.3.1 Frazionamento

Le procedure di deposizione sono state testate e modificate diverse volte per ottenere un metodo di frazionamento ottimale. Cyt-C ha la mobilità più alta tra le tre proteine, mentre Lact è la meno veloce. L'ordine di migrazione atteso è pertanto: Cyt-C, Lys, Lact. I primi tentativi erano volti a depositare una singola specie su numerosi spot, in modo da avere certamente almeno alcuni degli spot con solo l'analita desiderato. Infatti, siccome i tempi di migrazione di Cyt-C e Lys sono molto vicini tra loro, la distribuzione di un analita contenuto in un singolo picco su molteplici spot avrebbe assicurato di ottenere analiti puri almeno in alcune delle aree di deposizione, sebbene in quantità ridotte. Questo era in teoria un principio funzionante, ma i risultati sono stati poco promettenti poiché la quantità già limitata di analita nel capillare, ulteriormente ripartita su molti spot, avrebbe portato a una scarsa rilevabilità in MS. Pertanto, il metodo è stato ottimizzato per depositare un intero picco (vale a dire una singola specie) su un massimo di tre spot, sebbene questa procedura necessiti di maggiore precisione.
I frazionamenti del campione di tre proteine sono stati condotti nel seguente modo:

- La separazione viene normalmente avviata e condotta in "modo I". Non appena i primi due picchi hanno superato la finestra di rilevamento (il che significa che hanno percorso metà capillare), il tempo di picco e la larghezza del picco vengono registrati nel foglio di calcolo Excel, che fornisce parametri quali tempo di inizio frazionamento e tempo di deposizione per ogni specie.

- 30 secondi prima di iniziare a depositare, in base all'output del foglio Excel, il braccio robotico solleva il capillare dalla vial di uscita e lo sposta sul punto in cui la deposizione deve iniziare (è possibile associare la coordinata zero ad un punto qualsiasi della piastra MALDI tramite la piattaforma Arduino). Questo processo corrisponde alla commutazione del sistema da "modo I" a "modo II".

- All'inizio della deposizione, il capillare viene pressurizzato con 0.125 bar per velocizzare la fuoriuscita degli analiti dal capillare. Il resto della deposizione è controllato dal software che governa il braccio robotico e la piattaforma XY.

Sono stati utilizzati tre spot per Cyt-C e tre per Lys, dato che hanno mobilità simili e pertanto migrano in tempi quasi uguali. E' stato quindi necessario un certo margine nel numero di deposizioni per assicurarsi che almeno in uno o due dei tre spot le proteine fossero pure. Si è poi fatto uso di spot "spazzatura" con lo scopo di raccogliere il BGE che eluiva tra le prime due proteine e la terza, perché Lact raggiungeva la fine del capillare molto dopo Cyt-C e Lys.

In particolare, si è scelto di utilizzarne due, in modo da sventare la possibilità che eventuali residui di Lys in ritardo fossero depositati sullo stesso spot di frazioni di Lact in anticipo. Infine, è stato utilizzato uno spot singolo per frazionare Lact.
In totale sono stati utilizzati nove spot per ogni frazionamento: tre per Cyt-C, tre per Lys, due "spazzatura" e uno per Lact. Uno spot singolo è stato sufficiente per Lact perché, essendo molto più lenta delle altre due, non sussisteva il rischio che le altre proteine ne compromettessero la purezza sulla piastra. La figura 4 mostra il risultato di un frazionamento condotto con la procedura descritta.

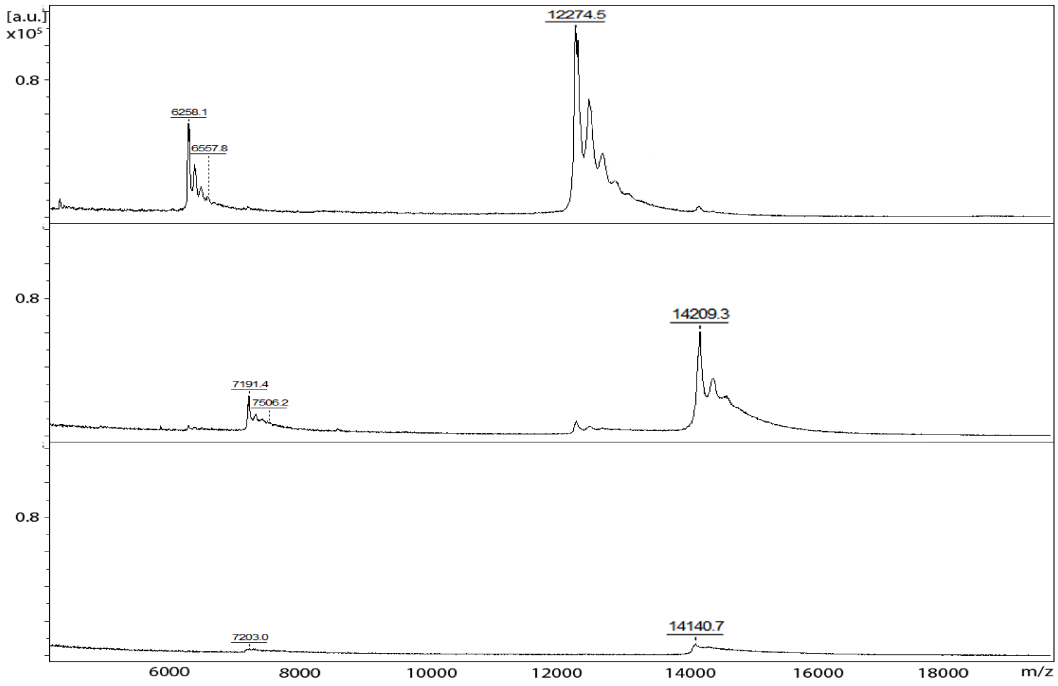

Figura 4: Confronto di tre spettri MS di proteine frazionate con il metodo presentato. Sopra: citocromo C, in mezzo: lisozima. In basso: α-lattoalbumina. I picchi sotto 10000 m/z sono ioni con doppia carica. Matrice: acido sinapinico. Numero di colpi: 3000. Frequenza: 2000 Hz.

Il metodo presentato sopra è stato ulteriormente ottimizzato scegliendo di mantenere la tensione spenta durante le deposizioni. Ciò ha avuto un impatto positivo sui frazionamenti; infatti, mentre nel primo caso si registrava una forte instabilità nella corrente dovuta al movimento del capillare, una deposizione pilotata solo dalla pressione non avrebbe necessitato della continuità elettrica per far si che le eluzioni si verificassero a velocità costante. In particolare, i frazionamenti condotti con tensione applicata causavano una caduta di corrente ogni volta che il capillare veniva sollevato da uno spot e spostato verso il successivo. Lo stesso, con un'estensione temporale maggiore, avveniva quando il capillare veniva spostato dalla fiala alla piastra (ovvero nel passaggio da "modo I" a "modo II").
Il metodo ha permesso di eluire separatamente le biomolecole, separandole e spottandole direttamente su piastra MALDI, specie per specie, con conseguente guadagno in termini di sensitività e risoluzione.

0.3.2 Analisi di Glicopeptidi

Tripsina e α-chimotripsina hanno diverse selettività rispetto ai siti di clivaggio. La seconda taglia più legami peptidici della prima, e perciò produce peptidi più corti. Le digestioni di OPN con entrambi gli enzimi sono state simulate tramite apposite piattaforme per studiare come sarebbero state distribuite le glicosilazioni sui peptidi prodotti.
La digestione con tripsina produce, tra gli altri, un peptide recante tutte e cinque le glicosila-

zioni, dal peso di 9179.6 Da (più i cinque glicani, che contribuiscono per 948.3 Da ciascuno). L'α-chimotripsina invece porta alla divisione dei cinque siti su due peptidi diversi: uno da 1551.8 Da con tre glicosilazioni, e uno da 1091.6 Da con due glicosilazioni.
Nessuna delle digestioni ha portato a risultati soddisfacenti in MS. Dapprima sono state condotte le reazioni con tripsina, ma gli spettri non hanno mostrato picchi relativi a ioni con massa superiore a 8000 Da (mentre il glicopeptide, comprese le glicosilazioni, dovrebbe pesare oltre 13 kDa). E' stato concluso che i glicopeptidi, specie già difficilmente ionizzabili in MALDI, avrebbero ionizzato ancor meno se pesanti. Per questo motivo è stata presa in considerazione la digestione con chimotripsina, che avrebbe fornito glicopeptidi più corti. Anche in questo caso, però, non sono stati rilevati segnali relativi ai peptidi di interesse, neanche considerando parziali deglicosilazioni o desialilazioni. La mancanza di risultati è stata giustificata considerando che l'intensità dei glicopeptidi in MALDI-MS sia generalmente inferiore a quella dei peptidi non glicosilati, sia per la loro inferiore efficienza di ionizzazione, che a causa di soppressione ad opera di altri e più abbondanti peptidi, specialmente se i glicani in oggetto sono sialilati.

0.3.3 Analisi di Glicani

Se analizzati in MALDI, i glicani tendono a complessare metalli alcalini, soprattutto se hanno sialilazioni. Nessuno dei tre metodi testati ha portato a picchi relativi alla massa protonata del glicano intatto ($[M+H]^+$ = m/z 966) o del glicano intero con addotti metallici ($[M+Na]^+$ = m/z 988 o $[M+K]^+$ = m/z 1004). Invece, sono stati riscontrati picchi relativi al glicano mono-desialilato con addotti sodio e potassio (m/z 697 e 713, rispettivamente). È stato inoltre identificato un picco a m/z 384, probabilmente relativo al glicano completaemnte desialilato (*i.e.* la struttura core-1 protonata). Conferma di tale ipotesi è stata ottenuta in MALDI-TOF-TOF-MS, in cui la frammentazione dello ione parente a m/z 384 con tecnologia LIFT™ ha portato a una perdita neutrale di 43 m/z, possibile indicazione della perdita di un gruppo acetile. Tale perdita può essere motivata considerando un clivaggio induttivo sul legame C-N sulla catena laterale del GalNAc.
Le purificazioni dei glicani prima delle analisi sono state condotte con carbone grafitico poroso (PGC), secondo letteratura, e con C18, secondo una procedura innovativa di purificazione indiretta. Il primo caso non ha portato a risultati di rilievo, mentre il secondo metodo ha non solo incrementato il rapporto segnale/rumore fino a 7 volte (n=3) per i picchi di interesse, ma ha anche permesso l'assegnazione di nuovi picchi, non visibili in nessuno dei metodi basati su PGC. In particolare, in seguito a purificazione indiretta, è stato possibile osservare un picco a m/z 675, corrispondente al glicano mono-desialilato protonato. Il risultato di questo metodo di purificazione è riportato in figura 5. La tabella 1, invece, riepiloga i segnali riscontrati in MALDI-TOF-MS che confermerebbero la presenza di un O-glicano core-1 tra le glicosilazioni di OPN.
Le analisi effettuate sono una chiara prova della possibile presenza della struttura glicosidica core-1 come O-glicano di OPN.

0.4 Conclusioni e Prospettive

La mancanza di un metodo di concentrazione ben consolidato per l'OPN, così come la mancanza di informazioni attendibili sulla glicosilazione della proteina hanno reso l'analisi sui campioni biologici (plasma) un percorso inaccessibile. E' fondamentale espandere e portare avanti lo studio per far approdare l'analisi in campo biologico.
Il punto d'arrivo del presente lavoro sarebbe l'unione dei due blocchi principali che sono stati affrontati nella tesi. Più specificamente, sarebbe auspicabile che i glicani potessero essere prima separati in CE, e quindi spottati sulla piastra MALDI attraverso il metodo di frazionamento. Se questa idea funzionasse, contribuirebbe in maniera sensibile alla comprensione

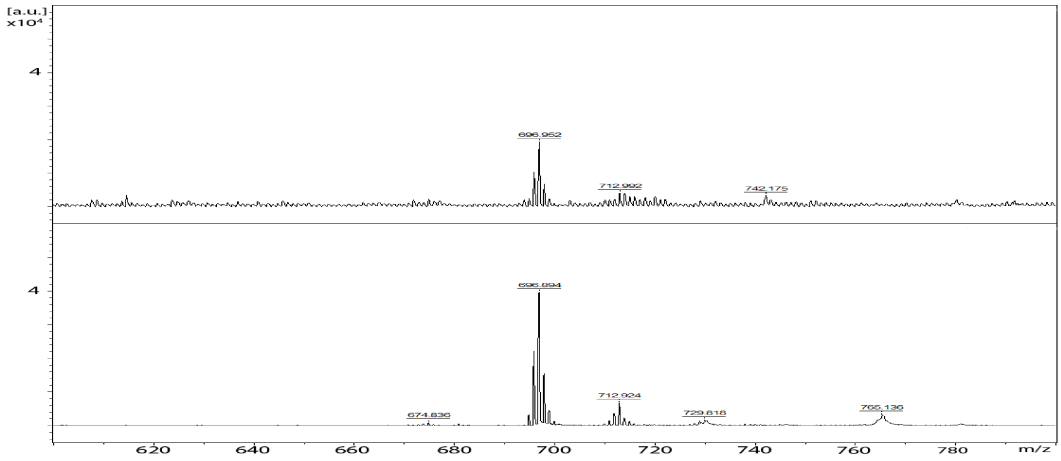

Figura 5: Close-up nella regione m/z 640 - m/z 740 per un campione di glicani purificato conPierce® C18. E' possibile apprezzare il sensibile miglioramento in S/N del campione purificato (sotto) rifetto a quello non purificato (sopra). Il metodo ha portato a un aumento del segnale per il picco a 713 m/z di undici volte nel caso riportato. Inoltre, è stato possibile apprezzare il picco a 675 m/z, relativo al glicano mono-desialilato [GalNAc-Gal-NeuNAc + H^+]. Matrice: acido 2,5-diidrossibenzoico 20 mg/ml. Somma di 6000 colpi in modalità reflectron. Frequenza: 2000 Hz.

Tabella 1: Riepilogo dei picchi rilevati in analisi MALDI-TOF-MS di glicani rilasciati da OPN. La struttura proposta è [$GalNAc - Gal - NeuNAc_2$].

m/z	**Glicano**	**Addotto**	**Struttura**
384	[GalNAc-Gal]	H^+	
675	[GalNAc-Gal-NeuNAc]	H^+	
697	[GalNAc-Gal-NeuNAc]	Na^+	
713	[GalNAc-Gal-NeuNAc]	K^+	

della glicosilazione dell'OPN, così come di molte altre proteine. L'unione di queste due parti ha tuttavia alcuni ostacoli importanti. In primo luogo, i glicani non sono assorbenti UV. Pertanto, sono necessari passaggi aggiuntivi in modo diretto (*e.g.* derivatizzazione, che comunque non è raccomandata perché spesso porta al rilascio di acidi sialici) o indiretto (*e.g.*con assorbimento indiretto; ovvero usando un BGE che assorba UV e rilevando i glicani come picchi negativi). In secondo luogo, non c'è molto lavoro svolto sui glicani in CE; quindi tutto il metodo dovrebbe essere adattato al meglio a questo tipo di campione. Infine, anche supponendo che venga trovato un metodo di concentrazione adeguato, non vi è certezza che il LOD di CE sarebbe abbastanza basso da garantire il rilevamento dei glicani. È necessaria una rilevabilità sufficiente in MS per valutare la qualità del frazionamento.

Uno studio complementare all'analisi CE-MALDI-MS dei glicani rilasciati potrebbe essere l'analisi LC-ESI-MS dei glicopeptidi. I glicopeptidi non hanno funzionato perché MALDI-MS non è la più indicata per tali specie, specialmente se i peptidi sono pesanti e le unità glicaniche sono sialilate. Molti hanno invece lavorato con i glicopeptidi in MS mediante una sorgente di ionizzazione elettrospray (ESI), grazie ai vantaggi che si possono ottenere dall'accoppiamento online con CE e in particolare con LC. Questo approccio combinato potrebbe fornire un'idea più chiara di tutti i glicani possibili presenti in OPN: in questo lavoro l'attenzione è stata

posta su una sola struttura, perché questa struttura (figura 1) è stata presentata in letteratura (Christensen *et al.* 2008) come quella che più probabilmente rappresenta l'unico glicano di OPN. D'altra parte, ci sono stati alcuni lavori in disaccordo con questa ipotesi; sono quindi necessarie ulteriori indagini.
Potrebbe inoltre essere interessante investigare quanto si possa scendere in concentrazione di OPN utilizzando gli stessi metodi qui presentati. In questo lavoro, l'attenzione non è mai stata sulle concentrazioni, perché l'obiettivo principale è stato quello di comprendere la glicosilazione dell'OPN da un punto di vista qualitativo. L'obiettivo era capire se il glicano GalNAc-Gal-NeuNAc$_2$ fosse effettivamente presente in OPN. Al fine di ottenere ulteriori conferme per la presenza del glicano core-1, si potrebbe valutare l'aggiunta di sali di litio al campione. Lo ione con addotto litio, $[M + Li]^+$, potrebbe essere un'ulteriore prova della presenza di glicani, anche se non si presenta in maniera così comune come gli addotti sodio e potassio. Pertanto, si potrebbe pensare di confrontare lo stesso campione di glicano con e senza sali di litio, e osservare se si rivela un picco in più. La stessa idea potrebbe essere sfruttata semplicemente aggiungendo sali di sodio, che creerebbero un maggior numero di addotti, e valutando la differenza nel segnale $[M + Na]^+$ mediante quantificazione con, ad esempio, standard interni.

Con questo progetto sono state fornite nuove informazioni sul pattern di glicosilazione dell'OPN, che si spera possano aggiungersi alle conoscenze già disponibili sul coinvolgimento della glicoproteina nella formazione e il riassorbimento osseo e possibilmente aprire la strada verso la sua istituzione come biomarker per le malattie ossee.

1. Introduction

Biological systems are intrinsically complex: billions of molecules tirelessly play their role in different places and different times to keep cells alive and correctly functioning. It is impossible to understand biology by conducting single, compartmented experiments: data from various techniques and different areas of science have to be incorporated in order to gain a wider and clearer perspective over biological systems. In recent years, a combined perspective involving proteins, glycans and metabolites has been brought up, strongly backed by the technical advances on the instrumental side, which allow for deeper understanding and thorough analysis of biomolecules. This level of information has specifically come off in the so-called "omic" strategies, such as proteomics (see section 1.1) and metabolomics, as promising ways to discover *e.g.* potential biomarkers for disease diagnosis. For instance, global glycan profiling in serum has led to the discovery of numerous promising markers in cancer, and protein glycosylation has also been relevant to understand liver pathologies [1, 2]. "Omics" strategies are mainly based on mass spectrometry (MS) platforms [3]. MS has undergone incredible development over recent years, especially with respect to sensitivity of analysis and accuracy of results, turning into an indispensable tool in biological analysis. [3]
This project is framed within the above described area: in the quest for new biomarkers and in proteome characterization, it is crucial to develop working analytical methods to perform efficient separation of proteins from complex mixtures (*e.g.* blood, serum, cerebrospinal fluid, *etc*). The first part of this work focuses on the development of a capillary electrophoretic method aimed at the separation of proteins of similar mass. The proof of principle has been conducted on cytochrome C, lysozyme and α-lactalbumin, all ranging between 10 and 15 kDa. Proteins are separated in positive mode under the drive of an electric field, and they are let flow out of the capillary onto a matrix-assisted laser desorption/ionization (MALDI) plate by pressure.The capillary is automatically controlled by a robotic arm, so that single protein species can be directly deposited on a MALDI plate prior to MS analysis. The second part of this work is focused on the glycosylation study of osteopontin (OPN) as a potential biomarker for Complex Regional Pain Syndrome (CRPS). The key idea of such a study is to compare the glycosylation pattern of the OPN present in healthy plasma donors against the one in CRPS-diseased patients. A first approach chosen to tackle this investigation is through the analysis of OPN glycopeptides, obtained after enzymatic digestion. After purification and concentration of OPN spiked plasma samples, the protein was digested with Trypsin or α-Chymotrypsin, and the digested peptides were analysed with MS using a MALDI ion source and a time of flight (TOF) analyzer. The second approach chosen for the study consists of whole-glycan analysis after chemical release (beta-elimination with excess of a reducing agent). This path was followed because glycans are suspected to be involved in the signaling and development of several types of diseases.

1.1 Proteomics

"Proteomics" aims to the large-scale characterization of the proteome of a cell line, tissue or organism. [4] This research area was born as a natural and necessary consequence of genomics: researchers understood that gene sequencing alone would not be enough to gain full understanding of complex processes such as aging, diseases, and environmental effects. In fact, the so-called "post-genomic" era blossomed when the classical biunivocal relation between genome and proteome was replaced by the idea that more proteomes can be a product of a single genome [5]. Proteome differentiation is mainly attributable to post translational modifications (PTMs) [6]; the most common are displayed in Table 1.1. PTMs are covalent processing events that affect the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids after the protein is successfully translated [7]. PTMs play major roles in determining the protein's activity, localization, turnover, and interactions with other proteins [7]. Hence, focus was shifted from genotype to phenotype, and many works started to be conducted not only on how protein isoforms (i.e. proteins with the same amino acid sequence, but different PTMs) relate to different activity and function, but also how protein themselves are affected by external factors [7]. Today, proteomics is strongly established in several disciplines (from biochemistry to bioinformatics) as a powerful instrument to perform protein mapping, protein characterization, PTM analysis, and study of proteome aging [4,8]. In fact, the proteome of an organism is highly dynamic, and it is affected by the environment wherein it is considered. "*Thus, examination of the proteome of a cell is like taking a 'snapshot' of the protein environment at any given time.*" [4]. As an example, a cancerous cell can express proteins in different amount or with diverse PTM than a healthy one. These factors might well indicate that such a cell could be targeted as a potential disease indicator. The latter is a good example of what a biomarker does: according to MacNeil J. *et al.*, a biomarker is an objectively measurable molecule that indicates the presence of an abnormal condition in a patient; it can either be a gene, or a protein, or a metabolite, that has been proved to be related to specific diseases [9]. It is thereby very straightforward to imagine why proteomics has gained so much attention in recent years.

Table 1.1: List of some common Post-Translational Modifications. Adapted from [4].

Type of PTM	**ΔMass (Da)**	**Stability***	**Function**
Phosphorylation			Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acetylation	+42	+++	Protein stability, protection of N terminus. Regulation of protein-DNA interactions (histones)
Methylation	+14	+++	Regulation of gene expression
Glycosylation			
N-linked	>800	+/++	Cell-cell recognition/signaling
O-linked	>800	+/++	Regulatory functions
Sulfation (sTyr)	+80	+	Modulator of protein-protein and receptor-ligand interactions

*Stability in tandem mass spectrometry: + labile, ++ moderately stable, +++ stable.

Interest in the subject has grown much within the the last 20 years, as a direct result of the technical and instrumental advances that have taken place in recent times. Among the most important developments, MS deserves to be mentioned, for its growing sensitivity towards PTMs. [10, 11]

1.1.1 Glycans and Glycoproteomics

Many of the secreted proteins in our body, such as hormones or plasma proteins, are glycoproteins, that are conjugated proteins containing covalently bound saccharides. [12]
Glycosylation is known as one of the major PTMs [13], and it presents a high grade of complexity. Even though the term "glycan" is defined by the IUPAC "Compendium of Chemical Definitions" as a synonym of polysaccharyde, it is mostly used when referring to the monosaccharide sidechain of a glycoconjugate [14]. Glycans are responsible for modulating protein structure and functionality, and are responsible for conferring cell-specificity [15]. Moreover, glycans are responsible for the regulation of both the folding and degradation of proteins. Finally, they glycans are involved in cell-to-cell signaling by mediating interactions with cells of the same kind or with pathogenic organisms such as viruses, bacteria, and multicellular parasites. Their potential utility as biomarkers for pathological conditions is a major driver for characterization of the glycome. [16]
The carbohydrates moieties attached to biomolecules are of three kinds. They can either be conjugated to lipids or to proteins through a nitrogen atom (N-linked), or to proteins through an oxygen atom (O-linked) [15]. Glycoproteins are therefore N-glycosylated or O-glycosylated. O-linked moieties are mostly attached to the protein through serine (Ser) or threonine (Thr) residues. On the other hand, N-linked saccharides are always attached to Asparagine (Asn), and they are found after consensus sequence Asn-Xaa-Ser/Thr (where is Xaa is any amino acid, except Proline (Pro). O-glycans are usually shorter than N-glycans, which rarely exceed 15 simple sugars in chain length [12]. Glycans are composed of sugars of a very restricted class. Specifically, the building blocks of glycoprotein side chains are:

- Glucose (Glc)
- Fucose (Fuc)
- Mannose (Man)
- Galactose (Gal)
- N-Acetylglucosamine (GlcNAc)
- N-Acetylgalactosamine (GalNAc)
- N-Acetylneuraminic acid (NeuNAc)

Most of the O-glycans belong to the so-called "mucin type", indicating sidechains that always start with a GalNAc residue. There are eight mucin-type fundamental structures (cores), as shown in figure 1.1. Despite the fixed mucin-type cores, a large heterogeneity exists. For instance, it is very common to find glycans with sialic acid cappings [17].

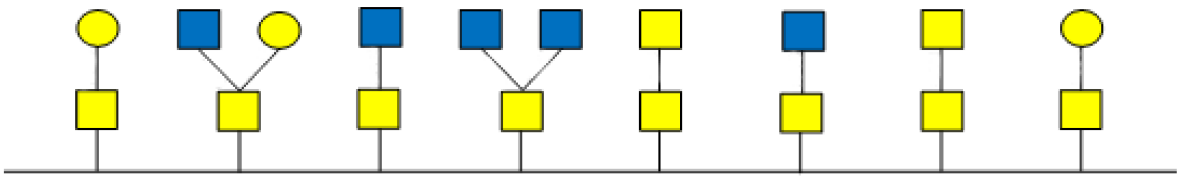

Figure 1.1: A schematic representation of the eight mucin-cores of O-Glycans. All glycans are intended to be linked to Ser/Thr amino acid residues. Cores that are made by the same monosaccharides have different carbon atoms involved in the bonds, or different optical isomery. Yellow square: GalNac; yellow circle: Gal; blue square: GlcNAc.

N-glycans are instead less strictly classifiable into core structures, but they always start with two GlcNAc residues, the first of which is linked to Asn, as mentioned above. Depending on what follows the GlcNAc sugars, three main subtypes can be identified: [18]

- High-mannose (many Man residues linked to the GlcNAc residues).

- Complex (any residues can show up).
- Hybrid (a forked structure with a high-mannose structure on one branch and a complex on the other).

1.1.2 Osteopontin and Complex Regional Pain Syndrome

CRPS is a disease of the central nervous system, involving changes in central sympathetic, somatosensory, and motor systems [19]. This disease generally develops after physical trauma such as fractures or strokes. An affected patient shows a range of symptoms varying from pain and swelling, to oversensitivity and skin color changes [20]. CRPS is a debilitating disease without a generally working cure. Early treatment is beneficial, however over and under diagnosis is very common. [21] Presently, the underlying cause of CRPS is unknown: researchers greatly lack in understanding this disease. The Budapest criteria were put into place in order to have a more precise diagnosis. This includes several diagnostic indexes including sensitivity and specificity, allowing categorization of the patients‘ pain [22]. However, this method presents its limitations, as based solely on clinical signs and symptoms which can easily be misinterpreted leading to over- and under-diagnosis. Being among the highest forms of chronic pain that exist in medical science, it could be helpful to find an effective noninvasive method of detection for this disease. It has been observed that many patients have localized bone density loss in the affected limb. This does not mean that the individual has generalized osteoporosis (excessive bone porosity and fragility) or is at an increased risk of developing osteoporosis at a later date. [23] Proteins such osteoprotegerin have been already specifically associated to CRPS [24]; but the studies conducted on the relations between OPN and bone remodeling, bone resorption and osteoporosis [25–27] suggest that this protein could as well represent a biomarker for bone conditions.
OPN is a multifunctional calcium binding-phosphorylated-acidic glycoprotein. It is expressed in several organs and it is involved in multiple biological cellular functions [28]. It has largely been investigated for its roles in bone physiology and biomineralization, bone remodeling after injuries or infections, as well as involvement in the pathogenesis of acute inflammation [29]. Furthermore, several studies have focused on its possible activity in the development of osteoporosis, multiple sclerosis and rheumatoid arthritis, especially due to its regulating properties of cell adhesion, migration and proliferation. [24, 25, 27, 29–34] Moreover, OPN is subjected to significant PTMs, such as phosphorylations and glycosylations [35]. In particular, OPN phosphorylation is thought to increase cell adhesion properties against osteoblasts [36]; O-glycosylations as well are thought to be involved in cell adhesion and migration properties [35]. In spite of the vague information upon the functional roles of O-glycans in OPN, a broad spectrum of involvement in biological functions and diseases emerges. Among others, its roles in bone mineralization makes OPN an interesting target to study with respect to bone conditions. It has been suggested that such PTMs could act in OPN as regulatory switches in promotion or hindering of mineralization [37]. This is where the idea of studying the glycosylation pattern of OPN came from: assessing the difference in OPN glycosylation between healthy individuals and diseased ones could be the key for *e.g.* early diagnosis of bone conditions. Extensive research has been made on the glycosylation pattern of OPN. Human urinary OPN [38], as well as rat bone OPN [39], mouse osteoblasts OPN [40] and milk OPN [35] seem to have in common 5 sites of O-glycosylation, all located in a very restricted region of less than 20 amino acids ($Thr^{118}, Thr^{122}, Thr^{127}, Thr^{131}, Thr^{136}$), and no N-glycosylations [37]. Reported concentrations of OPN in healthy individuals are of the order of magnitude of $10^1 - 10^2 \mu$g/L [30, 41]. It is evident that with such concentrations purification and concentration steps are needed prior to analysis, in order to make biological OPN detection possible. For this reason, the glycosylation study was firstly conducted on a pro-

tein standard OPN, with the intention of gaining knowledge that can be exploited for future studies on biological samples.
The standard OPN used in this work is made of 298 amino acid, with a total backbone weight of 33.7 kDa. Here follows the sequence of the residues:

IPVKQADSGS SEEKQLYNKY PDAVATWLNP DPSQKQNLLA PQNAVSSEET
NDFKQETLPS KSNESHDHMD DMDDEDDDDH VDSQDSIDSN DSDDVDDTDD
SHQSDESHHS DESDELVTDF PTDLPATEVF TPVVPTVDTY DGRGDSVVYG
LRSKSKKFRR PDIQYPDATD EDITSHMESE ELNGAYKAIP VAQDLNAPSD
WDSRGKDSYE TSQLDDQSAE THSHKQSRLY KRKANDESNE HSDVIDSQEL
SKVSREFHSH EFHSHEDMLV VDPKSKEEDK HLKFRISHEL DSASSEVN

The five sites of O-glycosylations are underlined in the sequence above). Christensen *et al.* proposed a possible common core-1 structure for all five sites [38]. The proposed structure is illustrated in figure 1.2.

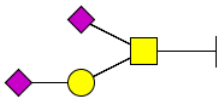

Figure 1.2: GalNAc-Gal-NeuNAc$_2$ glycan structure proposed by Christensen *et al.* [38]. Yellow square: GalNAc; Yellow circle: Gal; Purple rhombus: NeuNAc.

The mass of the glycan in figure 1.2 is 965.3 Da if it is a standalone molecule; it is instead 948.3 if attached to the backbone of the protein. This is because the glycosidic bond releases one water molecule. The main goal of the glycan analysis in this work is to test whether this structure is a plausible one among the possible glycosylations of OPN.

1.2 Capillary Electrophoresis

Electrophoresis is the migration of ions in solution driven by an applied external potential. CE is the separation of ions in solution under the influence of a potential applied from end to end of a fused-silica capillary. The advantage of this separation technique lays in its high resolving power, in the minute volumes of sample involved, and its wide versatility [42]. In fact, CE features different modes of operations. Capillary Zone Electrophoresis (CZE), which is used in this work, is the simplest of these modes. In CZE, the capillary is filled with a background electrolyte solution (BGE) and the sample is injected at one end of the capillary, after the electric field or a hydrodynamic push have been applied. The solutes therefore move towards the electrode of opposite charge; i.e. anions move towards the anode and cations move towards the cathode [42]. Figure 1.3 shows a schematic drawing of a CZE setup. A detector is placed along the capillary (tipically a UV-vis detector), with the aim to monitor the time each species takes to move from inlet to detector. The signal is then processed by a computer which outputs the electropherogram [42]. The on-capillary detection ensures high efficiency, since no dead-volume mixing affects the measurement. Capillaries usually range between 25 μm and 150 μm in diameter, and are 150 mm to 1000 mm long. The small dimensions involved determine the injection of a low volume of sample during each run. Depending on the specific capillary, injected samples have a volume of 1 to 50 nL. [42] Solutes in the capillary are separated depending on the difference in their velocities. The velocity v of a solute is determined by the following:

$$v = \mu_e E \tag{1.1}$$

Where μ_e is the electrophoretic mobility, and E is the electric field. The electrophoretic mobility is dependent on the contrasting action of two opposite forces: the electric force that the species experience in the medium, and the frictional drag. μ_e can be evaluated by:

$$\mu_e = \frac{q}{6\pi\eta r} \tag{1.2}$$

Where q is the ion charge, η is the solution viscosity and r is the hydrodynamic radius (or Stokes radius). Besides the electrophoretic movement, there are other factors affecting the movement of solutes in the capillary. In fact, the inside walls of silica capillaries induce a radial distribution of charges, which provoke a pumping action in the longitudinal way when subject to an electric field. This happens because silanol groups are easily deprotonated above pH 2-3 [42], thus becoming negatively charged (Si-OH $\rightarrow$ Si-O^-). Positive ions therefore are attracted to the capillary wall. A tighter first layer of positive charges (Stern layer) neutralizes a fraction of the silanol ions (the extent of neutralization depends on the concentration of the buffer), and a subsequent diffused layer neutralizes the remaining negative charge. Under the influence of the applied voltage, the diffused layer experiences a momentum and starts moving longitudinally, creating a uniform plug flow, called electroosmotic flow (EOF).

$$v_{eof} = \frac{\epsilon\varsigma}{\eta}E \tag{1.3}$$

Where ϵ is the solution's dielectric constant, ς is the zeta potential and η is the solution viscosity. The corresponding mobility (μ_{eof}) is the right-hand term in equation 1.3 without the electric field E. As a consequence, solutes in the capillary will move depending on the sum of their own electrophoretic mobility and the electroosmotic one. This resulting mobility is called apparent mobility, μ_{app} where:

$$\mu_{app} = \mu_e + \mu_{eof} \tag{1.4}$$

Analytes will all move towards one electrode irrespective of their charge, because the electroosmotic contribution is usually larger than the electrophoretic one (when the EOF is present). Species having positive charge will move faster than the EOF, negative ones will migrate slower than the EOF and neutral ones will move at the speed of the EOF. pH has a significant influence on separation, as it influences the degree of ionization of the silanol groups on the capillary wall, thereby changing the zeta potential and the EOF. The EOF is also affected by the ionic strength of the buffer, which impacts directly on the degree of neutralization of wall charges.

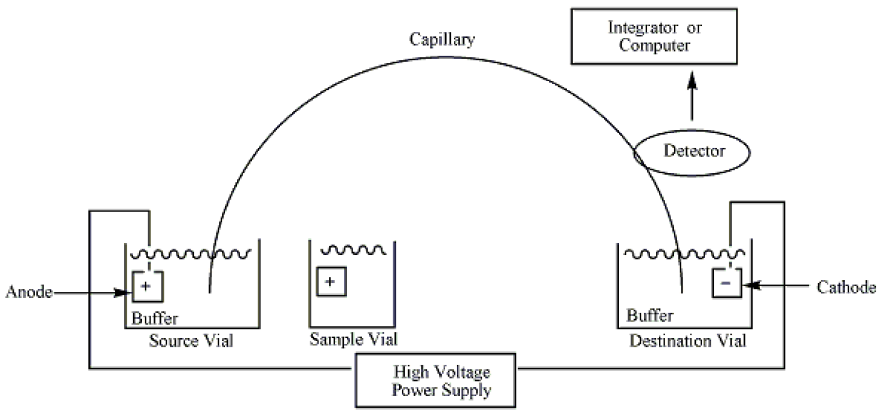

Figure 1.3: Simplified apparatus for a capillary zone electrophoresis instrument. In the specific case, the plus symbols on the inlet vial and on the sample indicate a positive-mode apparatus. Source: *https://upload.wikimedia.org/wikipedia/commons/3/37/Capillaryelectrophoresis.png*

Proteins are formed by amino acids connected by peptidic bonds. Their global charge is given by the side chains of its amino acids. A protein will be neutral at its isoelectric point (pI). When proteins are at pH below the pI, they will have a positive net-charge whilst at pH above the pI the net-charge of the protein will be negative. It is therefore fundamental to control the pH of the buffer used in CE.
When working with proteins as analytes, absorption of proteins onto the capillary walls occurs. This happens as a consequence of the multiple charges that these large macromolecules bear. Even when they carry a net negative charge, some sites on the molecules might still present a positive charge, and the wall-protein interaction is thus possible [42]. In order to circumvent such problems, the capillary walls can be coated either dynamically or covalently.

1.2.1 Wall Modifications

Inner capillary walls are coated for multiple reasons; among them, the suppression of the EOF, and the inhibition of wall-analyte interactions. Both effects are dependent on the neutralization of the silanol anions on the silica groups, which are ionized in aqueous solutions, bringing a charge to the surface and causing EOF during CE runs [43] Instead, when the net charge on the walls is brought to zero, no zeta potential is formed and therefore there is no pumping action in the liquid bulk. Also, there are less (if not no-) charged sites, which analytes could be attracted to and therefore be retained at the wall. Wall modifications are achieved by:

- Dynamic modification: the surface modifier is part of the BGE and the coating is equilibrium driven.
- Permanent modificaton: chemical bonding or physical adherence. [42]

Dynamic coatings are performed by adding coating agents to the separation medium [43]. This is advantageous because it provides self-regeneration, as the CE buffer is continuously replaced. These kinds of coatings are therefore chosen because they are easy to obtain, often inexpensive, and quite versatile; but on the other hand they might interfere with the optical detection, and also they sometimes bring issues to post-column analysis such us MS, which may be sensitive to some additives. Permanent coatings generally consist of polymeric materials, which are let to polymerize onto the inner walls of the capillary. The outcome is similar to the one achieved with dynamic modification, but the coating is on the other hand permanent, and requires no maintenance (except possible regeneration and/or re-coating after a fixed amount of runs). This is usually a more expensive alternative than the dynamically coated capillaries. [42]
In this work, a permanent hydroxypropyl cellulose (HPC) coating has been chosen as wall modification to circumvent protein-wall interactions. HPC was selected according to previous works with similar purposes as the present one [44].

1.3 Mass Spectrometry

Mass spectrometry is a powerful tool to identify unknown compounds [45]. In MS, molecules are firstly desorbed from condensed phases, ionized, and then accelerated towards the detector by a strong electric field. This allows for species separation according their mass-to-charge ratio (m/z), which in turn yields the molecular weight of the original compound. The charge given to analytes is dependent on the ion source. If the ion is singly charged, the m/z value will correspond to the real mass plus one ($[M+H]^+$). If it is doubly charged, the real mass will be double as the output value minus two, and so on. [46] More specifically, a mass spectrometer is qualitatively representable with five main blocks [45]:

- Sample inlet
- Ion source
- Analyzer
- Detector
- Software

The inlet can be at atmospheric pressure or under vacuum, while analyzer and detector generally work under high vacuum. Ion sources will vary their conditions depending on the technique. Several combinations of ion sources, analyzers and detectors are possible, which guarantee a high versatility and allow for dedicated studies on a wide range of organic and inorganic samples. For example, depending on type of analytes, one can choose if a "hard" method (i.e. a method which yields a high degree of fragmentation) or a "soft" method is appropriate (i.e. a method which yields a low degree of fragmentation). Macromolecules are usually analysed with soft methods, such us Matrix-Assisted Laser Desorption/Ionization - MS or ElectroSpray Ionization - Mass Spectrometry.

Matrix-Assisted Laser Desorption/Ionization - Time of Flight Mass Spectrometry

MALDI-TOF-MS is a widespread spectrometric technique in the study of biological molecules. Analyses are conducted based on the time that analytes take to fly to the detector after having been ionized.

A MALDI ion source is based on the absorption of laser light by a solid sample layer, deposited on a metallic plate. The samples absorb energy which causes them to desorb from the solid layer and ionize. Laser light is focused on small spots that usually do not exceed 0.2 mm in diameter. The sample is to be co-crystallized in an organic matrix, which serves to absorb part of the energy of the laser, and to assist analyte ion formation. [45] The exact mechanism of ion formation is still quite debated in the scientific community, but the matrix is thought to be involved in the process [45]. Aromatic matrices are often chosen, as they play a better role in absorbing the high amount of energy shone on the sample. In positive-ion mode, $[M+H]^+$ ions predominate, while $[M-H]^-$ ions are usually formed in negative mode [45]. The choice of a matrix and optimized conditions of sample preparation have substantial influence on the analytical value of MALDI spectra. Examples of the most common matrices for MALDI are: 2,5-dihydroxybenzoic acid (DHB, gentisic acid); α-cyano-4-hydroxycinnamic acid (α-CHC, α-CHCA, 4-HCCA, CCA,); 3,5-dimethoxy-4-hydroxycinnamic acid (SA, sinapinic acid). According to Bruker (Bruker Daltonics (Bremen, Germany) these matrices are commonly used for peptides, glycans and glycopeptides in the first case; peptides and smaller proteins in the second case; proteins in general in the latter one. DHB yields peripheral crystals, which are usually concentrated in hot spots. SA and 4-HCCA produce more diffused layers of matrix. MALDI matrices are often applied on the plate according to well-defined procedures (*e.g.* "dried droplet": pre-mixing of equal amount of sample and matrix before spotting; "dried sample": spotting of sample first, and matrix after the solvent of the sample droplet has evaporated; "double layer": application of first thin layer, followed by the "dried droplet" procedure; *etc.*) MALDI is often preferred as an ionization technique, because being a "soft" one it allows for the analysis of biomolecules. It also has the advantage that multiple re-examination of the same sample is often possible, because not all co-crystals are degraded after a single analysis. It is a very sensitive technique (sensitivity reaches down to the attomole level [47]), but it is not quantitative.

TOF analyzers are long, straight, evacuated tubes in which ions are let to fly under the influence of an electric field. Analytes are accelerated by a strong voltage applied to the backplate and pushed towards the drift region, where there is no electric field. Figure 1.4 shows a simplified scheme of the instrument.

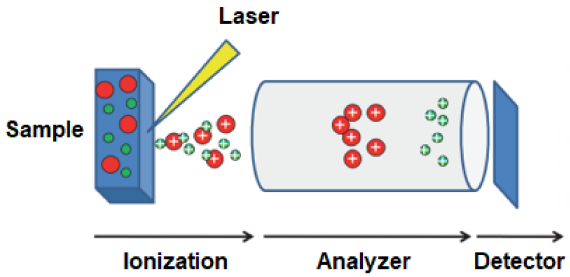

Figure 1.4: Schematic representation of MALDI-TOF mass spectrometer. Source: https://upload.wikimedia.org/wikipedia/commons/2/28/MALDI_TOF_EN.png

Theoretically, all ions start off with the same kinetic energy, provided by the field. Being the kinetic energy $1/2mv^2$, ions with different masses will travel with different speeds (*i.e.* will take different times to reach the detector). Practically, not all analytes emerge from the plate with the same energy, as they can be formed at slightly different distances from the backplate. This results in ions being accelerated through different voltage differences, and hence in ions gaining unequal kinetic energies. The consequent loss in resolving power is usually circumvented by operating the TOF in "reflectron" mode: a series of rings are added at the end of the tube, which are held a increasing positive potentials [45]. When ions approach the rings, they start slowing down more and more, till they stop and eventually reverse their motion, gaining progressive speed in the opposite direction. Ions are therefore accelerated once, then slowed down, stopped and re-accelerated in the opposite direction, as shown in figure 1.5. The more kinetic energy ions have when they emerge, the longer will they penetrate into the

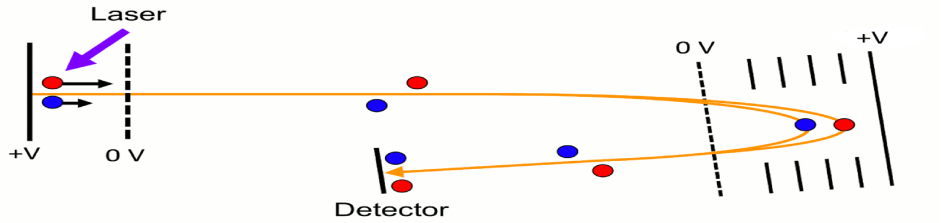

Figure 1.5: Working principle of a reflectron. Adapted from: K. K. Murray - Own work, GFDL, https://commons.wikimedia.org/w/index.php?curid=59785578

electric field before they are stopped and their motion is reversed. This way, all differences are compensated, and analyses are more reliable.

As mentioned above, MALDI-TOF-MS is often chosen for biomolecule analyses. In this work, it will be employed for the detection of proteins after their fractionation, and for the investigation of the glycosylation of OPN. MALDI is in fact widespread in the analysis of glycans [48–50].

2. Experimental

2.1 Chemicals

All water used in the present work was purified in a Millipore Synergy 185 (Bedford, MA, USA) to a resistivity of 18.2 MΩ cm at 25°C. acetonitrile (>99.8%), ammonium acetate (>98%), ammonium carbonate, hydroxypropylcellulose (80 kDa), α-chymotrypsin (type VII, TLCK treated), cytochrome-C (from equine heart, 12.4 kDa, >95%), ethanolamine, α-lactalbumin (from bovine milk, type I, 14.1 kDa, >85%), lysozyme (from chicken egg white, 14.3 kDa, >90%), osteopontin (human recombinant expressed in HEK 293 cells, >97%), quinine (>99%), trifluoroacetic acid (>99%), tris(hydroxymethyl)-aminomethane (>99.9%), trypsin (type XIII, TPCK treated) were purchased from Sigma-Aldrich. Hydrochloric acid (37%) and sodium hydroxide (>99%) were purchased from Merck. Calcium chloride was purchased from Mellinckrodt Chemical Works. Acetic acid was purchased from Honeywell Riedel-de Haen. Sodium borohydride (>98%) was purchased from Lancaster Synthesis, England. DHB, 4-HCCA, SA were purchased from Bruker Daltonics.

2.2 Capillary Electrophoresis Setup

All electrophoretic separations were conducted in a home-built instrument composed of 3 main blocks: high voltage supply, main CE box (where che inlet was located), UV detector (Spectra 100, Spectra-Physics, 15 Newport (Darmstadt, Germany). The capillary (CM Scientific (Silsden, UK) exited of the CE box towards the UV detector, and ended in the outlet vial, which was located outside the main CE box. During fractionations, the capillary ended on the XY-plate area instead, where the robotic arm controlled the outlet tip of the capillary (see section 2.4 for details). Therefore, unlike a commercial CE instrument, the inlet vial was never located in the same environment as the outlet one. Instead, a carton-built vial holder was crafted and glued to the side of the UV detector in a way that the liquid level in the outlet vial would be the same as the one in the inlet vial. All signals were recorded with HP Chemstation, (Hewlett-Packard (Stockholm, Sweden). The instrument was operated in two modes:

- Standard mode: closed-system runs operated from the inlet vial (in the CE box) to the outlet vial, located on the side of the UV detector. This mode will henceforth be referred to as "mode I", and it is represented by the top scheme of figure 2.1.
- Deposition mode: open-system runs in which no outlet vial is used. The capillary runs from the inlet vial (in the CE box), and the outlet is held by the robotic arm (see section 2.4). This mode will henceforth be referred to as "mode II", and it is represented by the bottom scheme of figure 2.1.

The capillaries used in this project were made of fused-silica and externally coated with polyimide (CM Scientific (Silsden, UK). More specifically, all capillaries used in the present

work had a total length of 80 cm, and an effective length of 40 cm. with an internal diameter of 50 μm and an external diameter of 375 μm. A coating treatment with HPC was conducted in order to hinder protein-wall interactions. The negative electrode was put in contact with the inlet BGE vial, whereas the outlet vial was grounded. In mode II, the MALDI plate was grounded too. A sketch of the two dispositions can be appreciated in figure 2.1. A flask containing calcium chloride was put in the CE box to promote water absorption. Keeping the humidity of the CE environment low was essential to hinder current dispersions in the inlet environment. All runs were performed in a 20 mmol ammonium acetate (BGE) buffer at pH 4.15, under a voltage of 26 kV. All injections were performed electrokinetically, with a 10 s injection, and a voltage of 10 kV. New capillaries were washed with methanol (MeOH) for 5 minutes, 1 M NaOH for 5 min, 0.1 M NaOH for 5 min, H_2O for 5 min before coating, according to Romand et al. [51]. Before use, every day, the capillary was rinsed with H_2O for 10 min and conditioned with BGE for 15 min. BGE was also used to flush the capillary in between runs for 5 minutes. Also, before shutting down the system every day, the capillary was rinsed with water for 10 minutes and with air for 5 minutes, to remove all liquids from inside. Once dried, the capillary was stored in air. Every day, fresh BGE was taken from a stock solution. Stock solutions were prepared on a monthly basis. BGE vials were always ultra-sonicated before use, and all samples were vortexed at 500 rpm for 5 minutes before use. All flushes were conducted with a pressure of 1 bar.

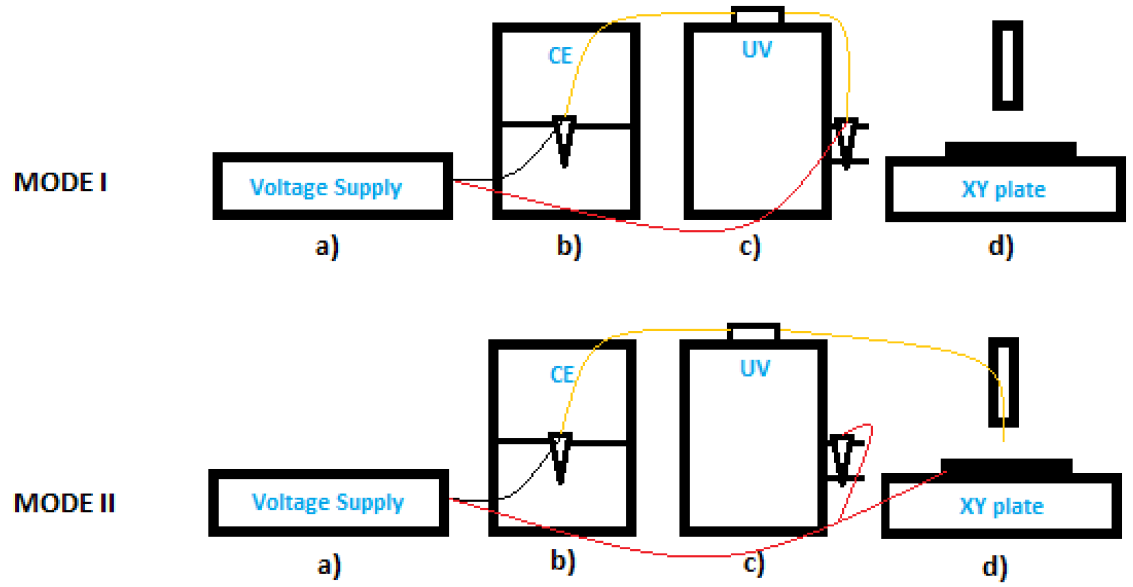

Figure 2.1: Comparison of the CE system in the two setups used for this work. The capillary is sketched in orange. The black line stands for the negative pole and red is ground. Mode I: the capillary end is in the outlet vial (which is grounded). Mode II: the capillary end is fixed in the robotic arm for deposition. The MALDI plate is grounded.

Wall Coating

The capillaries were coated with HPC according the procedure of Shen *et al.* [52]. Briefly, HPC powder was dissolved in water to a concentration of 5% wt. The polymer solution was left for 10 min under vacuum to eliminate bubbles. The capillary columns were filled with the polymer solution using a syringe, and the excess of polymer solution was removed using a N_2 flow at 4 bar for a few minutes. Immobilization of the polymer was conducted through heating the capillary in a gas chromatography oven (Agilent 6890 (G1530A) Plus Gas Chromatograph, HP Agilent, US) from 60°C to 140°C at a rate of 5°C/min and then

at 140°C for 20 min. N_2 was let to flow through the capillary throughout the process, still with a pressure of 4 bar. The coating procedure was conducted one time, or twice to form a double-layer coating. When the coating procedure was to be performed on a capillary that had been previously coated, preparative steps were performed before applying the procedure above. In particular, 1 M NaOH was flushed for 15 minutes, followed by 0.1 M HCl for 15 min to re-protonate the wall, and H_2O for half an hour.

2.2.1 Stability

Daily repeatability and reproducibility over multiple days were assessed for the CE system by means of an organic marker. Quinine (figure 2.2) was chosen for the purpose because it is an aromatic, therefore UV-absorbing, molecule and because it would be positively charged at the working pH (4.15, see section 2.2). Detection was performed at 250 nm, according to literature data on absorption wavelengths for quinine [53]. The migration time of quinine was measured every day before any other CE run, so that its mobility could be calculated and taken as a reference.

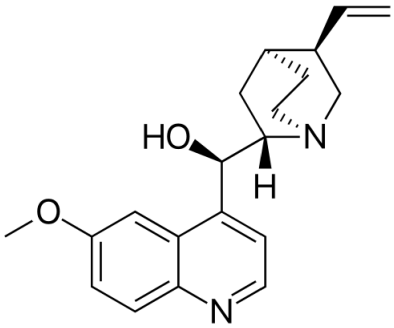

Figure 2.2: Molecular structure of quinine.

2.2.2 Intact Proteins Sample

CE separations and fractionations were conducted with a three-protein sample composed of 1.2 μM α-lactalbumin (Lact), 400 nM lysozyme (Lys) and 400 nM cytochrome C (Cyt-C), with respective molecular weights of 14.1, 14.3 and 12.4 kDa, and pIs of 4.5, 9.3 and 9.7. The green marks in figure 2.3 show the pI values of the proteins on a pH scale. The blue square indicates the pH value of the system. Detection for all proteins was performed at 200 nm.

2.3 MALDI-MS Setup

MALDI-TOF-MS analyses were conducted on a UltrafleXtreme system (Bruker Daltonics (Bremen, Germany). Spectra were acquired in linear mode for protein samples and in reflector mode for peptide and glycan samples, and laser was always shot with an intensity of

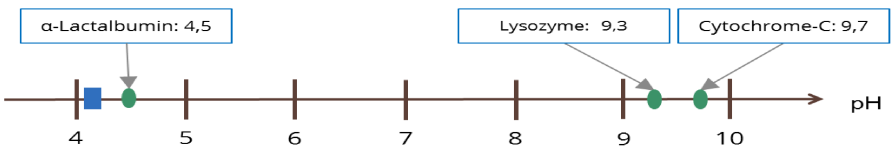

Figure 2.3: The green dots indicate the pI of the proteins in the sample. The blue square indicates the pH of the buffer.

70% or above with or without partial random walk. Spectra were recorded and analysed on FlexAnalysis: a S/N value of three was set as a cutoff to identify a peak. Each saved spectrum was a sum of 3000 shots for intact proteins and peptides, and 6000 for glycans. Calibration was performed before every analysis set by means of peptide/protein samples supplied by the manufacturer (Peptide calibration standard and Protein calibration standard I (Bruker daltonics (Bremen, Germany)). All samples were applied onto MTP 384 ground steel BC targets (Bruker Daltonics (Bremen, Germany)) using DHB, 4-HCCA or SA as matrices depending on the type of sample. Matrices were prepared according to the recommendations of the manufacturer, with little modifications. In particular:

- 4-HCCA: saturating a solution of TA30 (30 : 70 (v/v) Acetonitrile : 0.1% Trifluoroacetic Acid (TFA) in water) with 4-HCCA powder. Spotting on the target was performed by a "dried sample" method, depositing 0.5 μl of sample on an anchor first and 0.5 μl of matrix then (the "dried sample" method was reputed more suitable and handy for the specific sample, compared to the "dried droplet" method recommended by the manufacturer).
- SA: a slightly modified version of the "double layer" method was used: After the application of a first thin layer, the "dried sample" procedure was applied. Two different solutions were used when SA was employed as a matrix. The first one was SA saturated solution in ethanol (EtOH) - solution I, the second SA saturated in TA30 - solution II. A thin layer of solution I was formed with approximately 0.15 μl, then 0.5 μl of sample was deposited, and then 0.5 μl of solution II after sample solvent evaporation.
- DHB: a 20 mg/ml solution of DHB powder in TA30 was used. Spotting on the target was performed by a "dried sample" method, depositing 0.5 μl of sample on an anchor first and 0.5 μl of matrix then.

2.4 Hyphenation

CE and MS were coupled off-line with the help of a positioning table and a robotic arm. Hyphenation of the two above mentioned systems requires the CE to switch from a mode I operation to a mode II one during the run (with the help of an Excel sheet that calculates time and duration of depositions on each anchor of the MALDI plate). More specifically, the capillary is held by the robotic arm and it is positioned in the outlet vial from the beginning of the run, and the separation takes place. When a peak of interest shows up in the signal output on the computer (*i.e.* as soon as an analyte passes through the detection window), the Excel sheet is filled with the relevant data inherent to the considered peak. The Excel sheet calculates when and for how long depositions should be programmed in the Arduino platform that controls the deposition on all axes (for further detail check appendix A). A few seconds before deposition, the robotic arm moves the capillary away from the output vial and moves it to the starting point of deposition on the MALDI plate. At this point, a command is given to start the process and the deposition takes place according to the selected program. Different programs comprehend, among other things, deposition along a single row, deposition along a single column, or customizable patterns wherein a specific sequence of anchors can be selected, as well as the time of deposition on each anchor. This can be adapted to the deposition of a single peak, or to fractionation of multiple peaks. All depositions took place with a constant pressure of 0.125 bar.

2.4.1 Robotic arm and XY Plate

X and Y axes were controlled by a TIXY200 positioning table (Newport (Darmstadt, Germany); the Z axis was controlled by a MM-3M-EX extended Motorized MicroMiniTM Stage

(National Aperture (Cambridge, UK). An Arduino DUE (Ivrea, Italy) microcontroller together with a homemade circuit board were used to interface the three axes with a terminal. The control program in Arduino available from previous works [54] was revised and extended for optimal adaptation to the present work.

2.4.2 Pressure System

An air pressure system was developed and manufactured in order to ensure reproducible flushing and deposition conditions. All flushes were conducted at a pressure of 1 bar, whereas depositions were performed with a 0.125 bar pressure. Since the precision of the regulators was moderate, manually tuning pressure from 1 bar to 0.125 bar would not ensure reproducible conditions, which are crucial especially for depositions. Therefore, two independent pressure lines were used so that the above-mentioned pressure environments could co-exist. A schematic plot of the pressure system is represented in figure 2.4. The pressure line on the left, labelled P_1 supplied air at a pressure of 2 bar, which was then laminated through a downstream regulator (P_3) to the requested pressure of 0.125 bar. The second pressure line (P_2), drawn on the right side of figure 2.4, was set to 1 bar. This line supplied air both to the CE capillary and to a secondary capillary, which was exclusively used to deposit matrices on the MALDI plate. A three-way valve was used to switch between the two pressure environments depending on the need (flush or deposition).

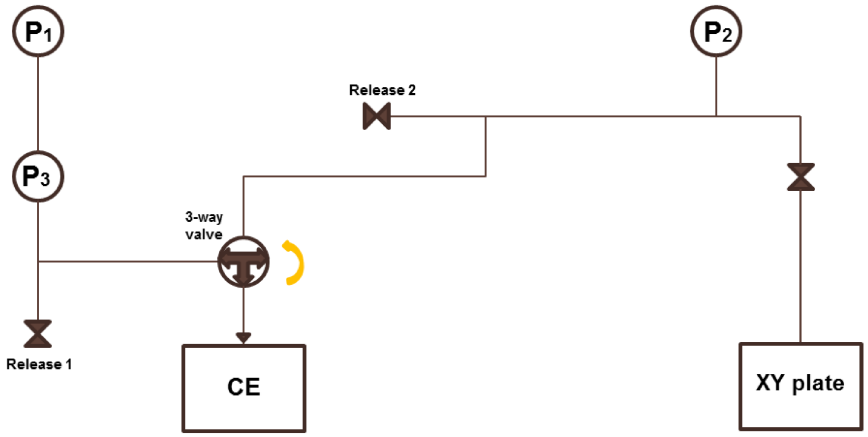

Figure 2.4: Sketch of the system used to flush and deposit on the MALDI plate.

2.4.3 Digestion of Proteins

Lyophylized OPN was either dissolved in H_2O or spiked in human plasma samples. The former samples were digested directly, whereas the plasma ones underwent a concentration/purification step[1] before digestion, to try to detect OPN from biological samples. Digestions were conducted according to standard protocols provided by Sigma-Aldrich, with slight modifications [55, 56]. In particular, trypsin digestions were conducted by first reconstituting trypsin in 1 mM HCl at a concentration of 1 mg/ml, and then adding it together with 1 μL OPN 1 mg/ml into a vial containing 40 mM ammonium bicarbonate (the manufacturer recommends 100 mM, but a lower concentration of salts was more convenient with respect to MS analysis) in a ratio (w/w) of 1:20 to 1:30 enzyme:substrate. Digestion solutions were incubated at 37°C for 17 hours, and they were quenched at 95°C for 3 minutes. All solutions were 50 μL in volume. Chymotrypsin digestions were conducted at a ratio (w/w) of 1:50 to 1:60 enzyme:substrate. Digestions took place in a 100 mM Tris(hydroxymethyl)aminomethane-HCl

[1] This method will not be presented in this work, as it is currently a submitted manuscript

(Tris-HCl) buffer at pH 8, containing 10 mM $CaCl_2$. Proteins were incubated for digestion for 24 hours at 37°C, and digestions were terminated by adjusting the pH to below 2.0 with HCl.

2.5 Glycan Analysis

2.5.1 Glycan Release

Glycans were released by β-elimination according to three procedures, based on the literature [57–59]. All solutions were 50 μL in volume, containing 1 μL OPN 1 mg/ml. The β-elimination reaction steps were followed as described in the papers.

- Method 1 was applied inspired by the work of Gao *et al*, 2013 [57]. Briefly, proteins were added to 50 μL of a 2 M $NaBH_4$ in 0.1 M NaOH. The mixture was incubated at 42°C for 18 h.
- Method 2 was applied inspired by the work of Ozcan *et al*, 2013 [58]. Briefly, proteins were added to 50 μL of a 0.1 M $NaBH_4$ in 1 M NaOH. The mixture was incubated at 42°C for 18 h.
- Method 3 was applied inspired to the work of Turyan *et al*, 2014 [59]. Briefly, proteins were added to 50 μL ethanolamine consisting of 20 mg/ml $(NH_4)_2CO_3$. The mixture was incubated at 42°C for 18 h.

All reactions were quenched with HCl 6M which was slowly added until the pH of each solution dropped below 3. During this procedure, the solutions were kept in an ice bath. Solid phase extraction (SPE) was performed to extract and enrich the glycans.

2.5.2 Glycan Purification through PGC Cartridges

SPE was conducted by means of porous graphitized carbon (PGC) Thermo Scientific™ HyperSep Hypercarb™ SpinTips cartridges, 10-200 μL, obtained by Thermo Fisher Scientific (Rockford, US). All extractions were performed by keeping the same ratio between unit load volumes and amount of stationary phase as the one used in the reference works mentioned in the previous subsection. All samples were 50 μL in volume. Before SPE extraction, the mixtures were vortexed for 1 minute.

- Method 1: Cartridges were preconditioned with 150 μL of 80% ACN containing 0.05% TFA, followed by 300 μL H_2O. In the loading phase, the β-elimination product was added. In the desalting phase, 750 μL H_2O was passed through the column. Then 150 μL of 10% ACN was used for elution.
- Method 2: Cartridges were conditioned with 300 μL of 80% ACN containing 0.01% TFA, and equilibrated by 300 μL H_2O. In the loading phase, the β-eliminated product was added. In the desalting phase, 600 μL H_2O was passed through the column. Then 150 μL of 40% ACN containing 0.05% TFA was used for elution.
- Method 3: Cartridges were prepared with 150 μL NaOH followed by 300 μL H_2O, and 150 μL acetic acid followed by 300 μL H_2O. They were conditioned with 150 μL of a 50% ACN solution containing 0.1% TFA (solvent A), followed by 300 μL of a 5% ACN solution containing 0.1% TFA (solvent B). The β-eliminated product was then added to the cartridge and let to equilibrate for 15 minutes. Washing was conducted with 150 μL H_2O followed by 150 μL of solvent A. Then 300 μL of solvent B were used for elution.

All flows in the procedures above were approx 2 ml/min. All eluates were collected and concentrated to 6 μL in an Eppendorf concentrator 5301 (Eppendorf AG, Hamburg, Germany) before MS analysis.

2.5.3 Glycan Purification through C18 Cartridges

An indirect purification method was developed, which involved Pierce® C18 Tips, 100 μL bed, purchased from Thermo Fisher Scientific (Rockford, US). The protocol was adapted from the manufacturer's one for peptides, without major changes. Each step involved a volume of 100 μL, except the loading of the sample, which always were 50 μL. Briefly, after adjusting samples to 0.25% TFA, tips were wetted by aspirating 50% ACN in H_2O for 4 times; then they were equilibrated by aspirating 0.1% TFA for 4 times. Sample was then loaded in the tip, dispensing and aspirating multiple times (10-15 times) for maximum adsorption efficiency. The extra steps in the protocol (rinsing of the tips for 3 times with 5% ACN plus 0.1% TFA, and eluting the retained compounds with 50% ACN plus 0.1% TFA) were not taken into consideration because the tips were always discarded after unloading of the sample.
The unloaded samples were collected, concentrated till dryness in an Eppendorf concentrator 5301 (Eppendorf AG, Hamburg, Germany) and then reconstituted to 6 μL with H_2O before MS analysis.

3. Results and Discussion

3.1 CE-MS Fractionation of Intact Proteins

3.1.1 Reproducibilty

The mobility of quinine was used as a reference to make sure that the system would work reliably and that the polymeric coating would be intact over time. Quinine has a first dissociation constant at pH 9.7 and a second at pH 5.07 (at 18 °C) [60], therefore it carries two positive charges under the aforementioned working conditions. The dissociations involve the nitrogen atoms shown in figure 2.2: the non-conjugated nitrogen has a higher degree of alkalinity than the one in the aromatic ring, and it will be more easily protonated. The former is therefore prone to take up a proton below pH 9.7; whereas the conjugated nitrogen is more likely to get protonated at more acidic pH (< 5.07). Hence, in the system used for this work, quinine was always doubly charged.
The degradation of the inner coating would lead to increased wall interactions and affect peak reproducibility, with respect to migration and shape. The latter was calculated by registering the migration of quinine, and calculating its velocity by taking $v = s/t$, where s is the effective length of the capillary in cm, t is the migration time in s, and v is the velocity in cm/s. Equations 1.1 and 1.4 were used, assuming an EOF mobility equal to 0. This was an accurate approximation because tests of electrophoretic runs with neutral analytes did not lead to any peak after 60 minutes of waiting. The resulting mobilities with relative standard deviation (RSD) values are presented in table 3.1.

The capillary was re-coated with the HPC treatment described in section 2.2 after a total of approximately 40 runs (the twelfth run on quinine was the fortieth overall). The coating was restored both because the capillary had been unused for ca. 15 days, and because the last runs performed before run 12 were slower than the ones done when the capillary was freshly coated. The re-coating procedure was preceded by preparative steps intended to improve the final result. NaOH 1 M was flushed to deprotonate eventual exposed silica groups, as well as for the removal of possible contaminations. HCl 0.1 M was used to re-protonate the wall, and H_2O was intended to remove any possible ions retained on the walls, if any. This pre-treatment was necessary because HPC coatings tend to degrade after a few CE runs; hence it is not surprising that silica groups become exposed again, after an extensive use of the capillary. There is no agreement upon the minimum number of runs to start the process of wall degradation, especially because the extent of deterioration depends on the operating conditions (pH of the BGE, liquid pressure, voltage, etc.). Up to 100 stable runs have been reported in previous works under HPC-coated capillary isoelectrofocusing conditions [52]. The RSD for the runs conducted before run 12 was 2.44% in migration times and 2.41% in mobility (n=11). A single layer polymeric coating in HPC ensured sufficiently reproducible conditions for a total of 40 runs. After re-treating with a double layered HPC coating, the RSD shifted down to 1.73% in times and 1.72% in mobility (n=8). At the time of the last value reported in table 3.1, more than 30 runs were conducted. Also, the capillary coating

Table 3.1: Quinine apparent mobility calculated over 19 runs spread across one month. Apexes indicate the days when the runs were performed. Temperature of the environment: 21.5°C - 22.5 °C.

Day	**Run**	**t** $[min]$	**RSD - t**	μ_{app} $[cm^2/(Vs) \cdot 10^{-4}]$	**RSD -** μ_{app}
1	1	6.778		3.03	
1	2	6.723		3.05	
1	3	6.696		3.06	
1	4	6.713		3.06	
4	5	6.597		3.11	
4	6	6.653		3.08	
5	7	6.919		2.96	
5	8	6.913		2.97	
11	9	7.023		2.92	
12	10	7.091		2.92	
12	11	7.091		2.92	
26	12*	6.779		3.03	
26	13	6.923		2.96	
26	14	6.819		3.01	
26	15	6.600		3.11	
27	16	6.883		2.98	
27	17	6.719		3.05	
28	18	6.652		3.08	
29	19	6.655		3.08	
			2.14%		2.11%

*The capillary coating was restored on this day.

has not been restored ever since, and it has shown indication of good reproducibility after 3 months, and approximately 80 runs (results not shown).

3.1.2 CE Analyses of Intact Proteins

Before starting depositions, the protein mix was run with CE in "mode I" to check if all peaks were visible, and to assess the degree of separation. Figure 3.1 shows the electropherogram from a CE run of the three-protein mix, namely Cyt-C, Lys and Lact.

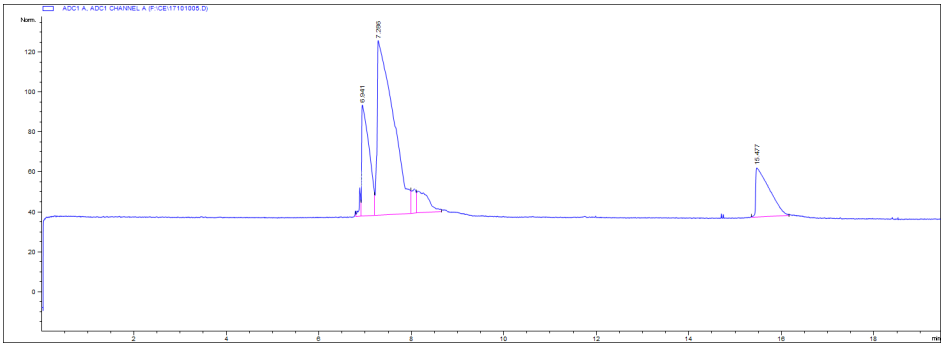

Figure 3.1: electrophoretic run of the three protein mix. Voltage: 26 kV, Background electrolyte: ammonium acetate 25 mM, pH: 4.15. Elektrokinetic Injection with 10 kV for 10 s. First peak: cytochrome C, second peak: lysozyme, third peak: α-lactalbumin.

The three marked peaks represent Cyt-C, Lys and Lact from left to right. The order of migration is in agreement with the pIs of the proteins (see section 2.2.2): Lact has the lowest

pI of the three, and it is almost equal to the buffer pH. In these conditions, the protein carries a very mild charge, and hence it has the lowest mobility. The size is also important in the assessment of the electro-migration, but being the mass of the proteins in the same range (12.4 kDa to 14.3 kDa) it can be assumed that they have comparable Stokes radii. Cyt-C and Lys have a similar pI value, and this is confirmed by their migration time to the detector, which is almost the same. Cyt-C migrates slightly faster due to the higher pI, which will indicated a higher charge at the running pH. Tables 3.2 and 3.3 show migration times and apparent mobilities for Cyt-C and Lys. Average mobilities resulted $(2.77 \pm 0.11) \cdot 10^{-4}$ cm^2/(Vs) and $(2.75 \pm 0.08) \cdot 10^{-4}$ cm^2/(Vs), respectively.

Table 3.2: Apparent mobility of Cyt-C calculated on 9 runs in 5 separate days spread along a three-weeks period. Double layer HPC-coated capillary. Temperature of the environment: 21.5°C - 22.5 °C.

Day	**Run**	**t** $[min]$	**RSD - t**	μ_{app} $[cm^2/(Vs) \cdot 10^{-4}]$	**RSD -** μ_{app}
1	1	6.813		3.01	
2	2	7.110		2.89	
8	3	6.942		2.95	
8	4	6.886		2.98	
8	5	6.909		2.97	
18	6	7.230		2.84	
21	7	7.607		2.70	
21	8	7.392		2.78	
21	9	7.378		2.78	
			3.87%		3.82%

Table 3.3: Apparent mobility of Lys calculated on 9 runs in 5 separate days spread along a three-weeks period. Double layer HPC-coated capillary. Temperature of the enviroment: 21.5°C - 22.5 °C.

Day	**Run**	**t** $[min]$	**RSD - t**	μ_{app} $[cm^2/(Vs) \cdot 10^{-4}]$	**RSD -** μ_{app}
1	1	7.186		2,85	
2	2	7.546		2.72	
8	3	7.287		2.81	
8	4	7.237		2.83	
8	5	7.270		2.82	
18	6	7.446		2.75	
21	7	7.866		2.61	
21	8	7.630		2.69	
21	9	7.618		2.69	
			3.06%		3.03%

It is evident that the system showed less reproducibility for the protein sample than for quinine. Though being still below 5%, the RSD value was double compared to the previous case. This might possibly be due to proteins being very prone to interact with each other: inter- or intra-chain interactions could then lead to a wider distribution of migration times. Interactions between proteins and coated walls are not to be excluded as well.

3.1.3 MALDI-MS Limit of Detection for Intact Proteins

The limit of detection for the protein sample was investigated in order to make sure that a sufficient amount of analyte would be deposited on each anchor during fractionation. Therefore, it was crucial to evaluate a cutoff protein amount above which a signal was produced in the MS. The three-protein pool was deposited directly on the MALDI plate in different concentrations to detect the limit of detection (LOD) for this protein sample. SA was used as matrix because it is suitable for intact proteins, according to the manufacturer. Moreover, in spite of the longer deposition procedure, SA forms on anchor a diffused and well distributed layer, compared to DHB which only forms peripheral crystals. With DHB, it is actually more time-consuming to seek for hot spots where samples reached a good level of crystallization. With SA, instead, good signals will be obtained even when the laser is shot randomly on the anchor. Four sample amounts were analyzed: 500 fmol, 250 fmol, 50 fmol and 25 fmol; the spectra from these four samples are shown in figure 3.2.

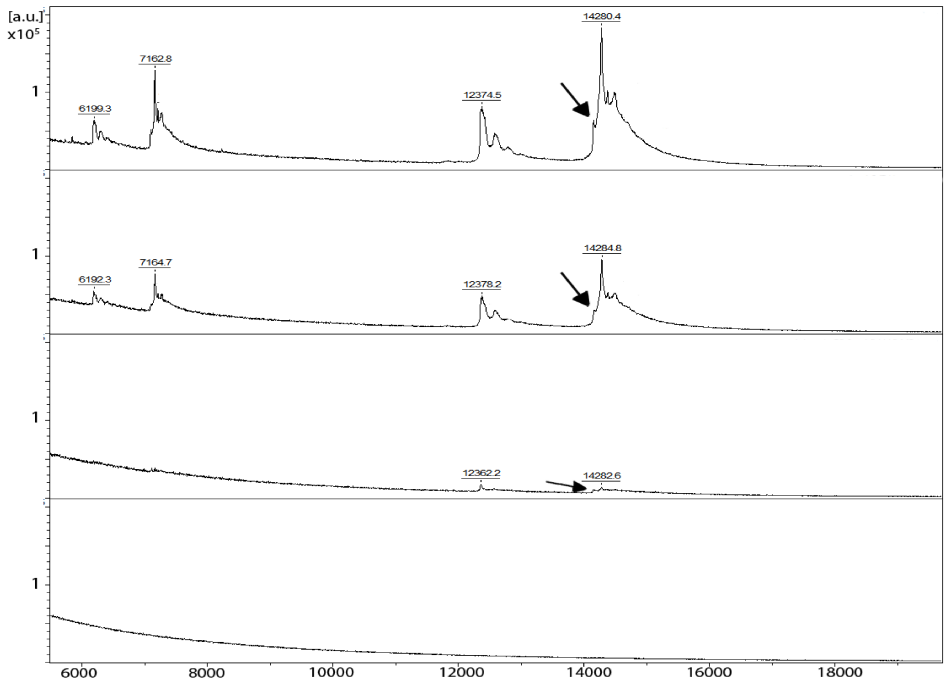

Figure 3.2: Spectra of the protein mix (Cytochrome C: 12.4 kDa, Lysozyme: 14.3 kDa and α-lactalbumin: 14.1 kDa - indicated with the black arrow) directly spotted on the MALDI plate to assess the detection limit. The peaks below 10000 Da in the two top spectra are doubly charged ions of the three proteins. Amounts deposited from top to bottom: 500 fmol, 250 fmol, 50 fmol, 25 fmol. Matrix: sinapinic acid. Number of shots: 3000.

Clear signals are still visible for all three proteins down to an amount of 50 fmol, whereas no detection was recorded for the 25 fmol sample. It was therefore concluded that all fractionations should be programmed on the Arduino platform such that each deposition would spot a higher amount than 50 fmol per anchor.

3.1.4 Fractionation of Intact Proteins

Different deposition procedures were tried to find an optimal fractionation method. Early attempts aimed at depositing the analyte related to one peak onto multiple anchors, in order to increase the chances to obtain fractionated species on different anchors. This choice was made because the migration times of Cyt-C and Lys were very close to each other, and splitting the analyte contained in a single peak over many anchors would ensure to get small amounts of pure analytes at least in some of the anchors. This was in theory a well working

principle, but it hardly ever brought interesting results, as the limited amount of analytes in the capillary would be split onto multiple spots, leading to poor detectability in the MS analysis. Therefore, the method was optimized to deposit a whole peak (*i.e.* a single species) on a maximum of three anchors only, even though this procedure would require much more effort. Fractionations were always performed with the help of a deposition sheet.
Fractionations of the three proteins sample were conducted with the following procedure:

- Separation was normally started and conducted in "mode I". As soon as the first two peaks passed the detection window (meaning that they had migrated half the total length of the capillary), peak time and peak width were recorded and given as input in the Excel deposition sheet.

- 30 seconds before the start of depositions, according to the output of the deposition sheet, the capillary was lifted from the outlet vial by the robotic arm, and automatically moved onto the point where the deposition would start. This process corresponds to switching the system from "mode I" to "mode II".

- When the starting time came, according to what calculated by the deposition sheet, a pressure of 0.125 bar was turned on to help push the analytes out of the capillary. The rest of the deposition was automatized by the Arduino platform controlling the robotic arm.

Six anchors were used for Cyt-C and Lys, because they would reach the end of the capillary right after each other. Some margin was therefore needed to make sure that at least in one or two out of the three anchors the proteins would be pure. A "waste" was used because Lact was reaching the end of the capillary several minutes after the first two proteins. Thus, two anchors were always employed to discharge the plug of BGE that followed Lys and forewent Lact. In total, nine anchors were used per fractionation: three for Cyt-C, three for Lys, two wastes and one for Lact. One spot only (with longer time of deposition) was sufficient for Lact because the protein was very slow in migration compared to the other two. Therefore there was no risk that other proteins with similar mobility would compromise the purity of the Lact fractions. Figure 3.3 shows the result of a successful fractionation with the described procedure.
The method presented above was further optimized by keeping the voltage off during depositions. This had a positive impact on fractionations; in fact, while the movement of the capillary during depositions would lead to a very unstable current, a pressure-only-driven deposition would not need to keep any electrical continuity in order for elutions to happen at a constant rate. In particular, fractionations conducted with the voltage on caused a drop in current every time the capillary was lifted from one anchor and moved to the adjacent one. The same, with even higher extension in time, was happening when the capillary was first moved from the vial to the plate (switch from "mode I" to "mode II"). These instabilities affected directly the velocity of the proteins in the capillary in a slight way, due to the small differences in migration times compared to the expected ones. The electropherogram in figure 3.4 shows the effects of current instability in the output signal. The sudden drop after 12 min corresponds to the time the capillary is lifted from the vial, and the subsequent oscillation is a consequence of the missing electric contact (*i.e.* the capillary is far both from the vial and from the plate, while it is moved from the former to the latter). Moreover, the vertical spikes after 13 min are due to the capillary moving away from an anchor and reaching a new one.
On the other hand, when the voltage is kept off, the variations that take place in the velocities of proteins can be neglected. Also, the analytes would not significantly change their position when the capillary is lifted from the vial and moved to the plate, because the voltage is off and the pressure is not on yet. A possible contribution that might represent a source of error

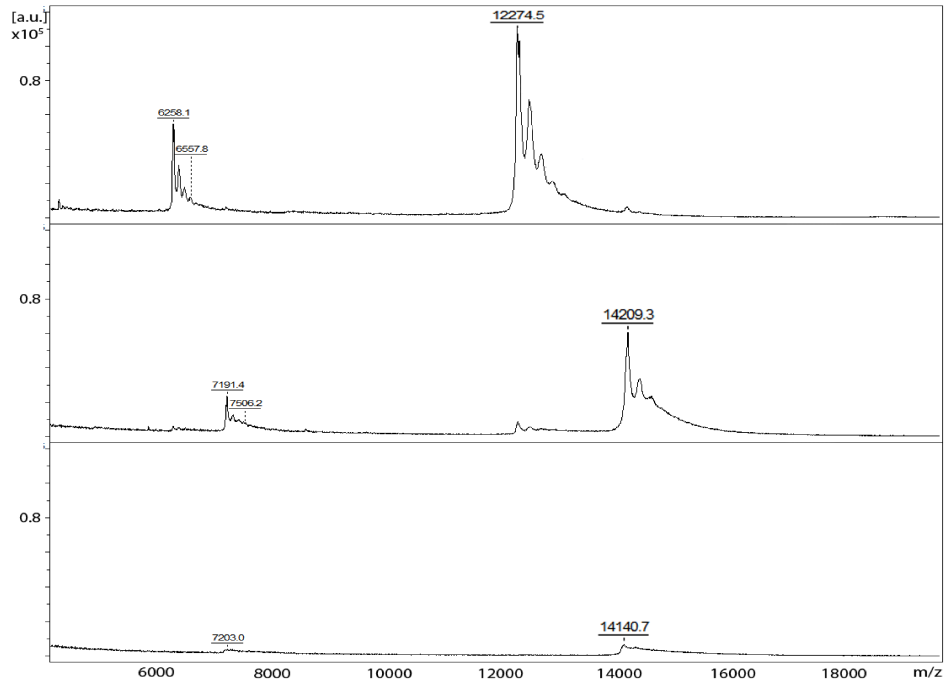

Figure 3.3: Fractionation of Cytochrome C (top), Lysozyme (middle) and α-lactalbumin (bottom). Matrix: SA. Number of shots: 3000.

is siphoning; but this is very difficult to avoid because the relative heights of inlet and outlet change multiple times during sample spottings.
All in all, pressure-only-driven depositions make fractionations slower, yet more precise. The deposition sheet was also slightly changed in order to account for pressure as the unique drive for the elution of analytes. Please refer to appendix A for a detailed explanation of how deposition times were calculated.

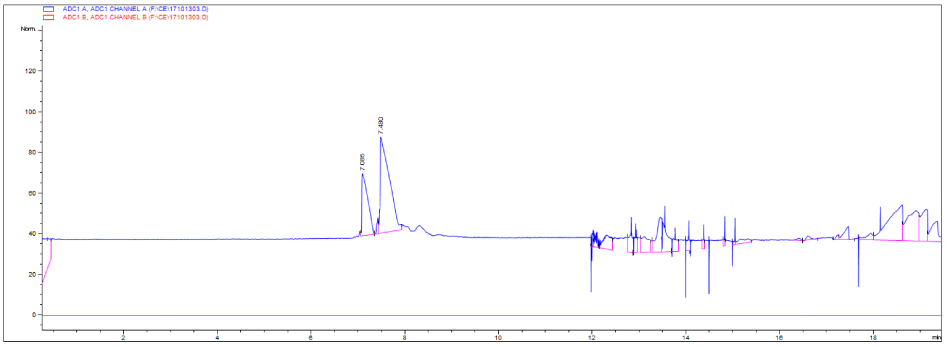

Figure 3.4: Electropherogram acquired during a deposition ("mode II") run with both pressure and voltage on. Voltage: 26 kV. Background electrolyte: ammonium acetate 25 mM. pH: 4.15. Elektrokinetic injection with 10 kV for 10 s. Pressure drive: 0.125 bar

Calculating the amount of sample spotted on each anchor was also crucial in order to forecast whether the deposited peak would be detectable in MALDI or not. In fact, the volumes injected in the capillary are extremely low, and they are not always enough to provide good signals after the sample is deposited on the plate. In order to estimate the amount of sample that was spotted on each anchor, a few assumptions were made:

1. the concentration of the proteins in the sample was assumed constant after every run. In reality, after an elektrokinetic injection, the concentration in a sample slightly decreases

as a consequence of charged analytes entering the capillary.

2. the amount of sample introduced in the capillary was assumed equal to the value given by the equation below:

$$Q = \frac{(\mu_e + \mu_{EOF})V\pi r^2 C t_{inj}}{L} \tag{3.1}$$

Where:

Q is the amount of injected analyte; mol

μ_e is the mobility of the analyte; $m^2/(V\ s)$

μ_{EOF} is the mobility of the the EOF; $m^2/(V\ s)$

V is the voltage applied during the injection; V

r is the inner radius of the capillary; m

C is the sample concentration; mol/m^3

t_{inj} is the time of injection; s

L is the total length of the capillary; m

3. The EOF contribution was negligible (*i.e.* wall coating efficient in the neutralization of wall charges).

4. The deposition times and durations given by the Excel sheet (see appendix A for further detail) were accurate and reliable. This would imply that the fractionations were always precise and no residues of the same analyte would be accidentally deposited on other anchors.

5. Siphoning was neglected during fractionations.

6. The temperature of the environment was considered constant, and so were the mobilities of the analytes, which are dependent on the viscosity of the buffer. The effect of Joule heating on temperature was neglected as well.

7. The pH value of the buffer was assumed constant over time, so that the charge of each analyte (and in turn its mobility) would be constant as well.

Here, the theoretical amounts of proteins that were injected in the capillary during the deposition that led to the fractionation of figure 3.3 follow. It is legitimate to assume that the whole amount of Cyt-C is distributed within the first three anchors, the whole Lys in the subsequent three, and Lact on the whole last anchor (the ninth with respect to the starting one).

Cyt-C: 113 fmol

Lys: 110 fmol

Lact: 43 fmol

After all adjustments/optimizations, fractionation showed a very good precision, as it is possible to assess from figure 3.3. This suggests that the assumptions above were legitimate.

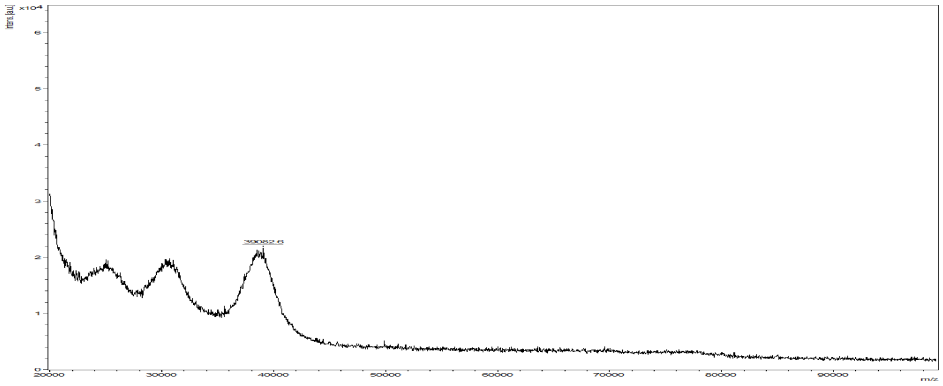

Figure 3.5: MS spectrum of intact OPN. The multiple humps are due to different isoforms. Matrix: sinapinic acid. Sum of 3000 shots.

3.2 Glycopeptides analyses

The standard OPN used in this work consists of 298 amino acid, with a total backbone weight of 33.7 kDa. A MALDI-MS spectrum of intact OPN is reported in figure 3.5.

The spectrum presents different peaks probably because of isoforms of the protein. The main peak at 39.1 kDa indicates that the protein is heavily modified, as it differs a lot from the expected backbone weight. The mass difference of approximately 5.4 kDa could be accounted for by considering the glycans themselves. Each of the core-1 structured glycans reported in figure 1.2 adds up 948.3 Da to the weight of the protein backbone. Assuming that the hyphotesis of Christensen *et al.* is right [38], glycosylations would account for a 4.7 kDa contribution, and the remainder 700 Da could be made up of phosphorylations.
Despite previous works have demonstrated OPN phosphorylations to range between 28 and 36 [37,38,61](which would mean a mass contribution ranging from 2240 Da to 2880), many of the considered proteins were originated in different regions of the human body, or sometimes even from different animal species. The real number of phosphate groups present in the standard OPN used for this work is not known.
In order to study the glycosylation pattern of osteopontin, the first approach used in the present work was enzymatic digestion. One or more peptides were targeted and their presence, mass and/or intensity were used as a clue for indirect evaluation of possible glycans. This idea was also developed by considering that MS generally has a better sensitivity for peptides than for whole proteins [47]. The final aim is to reach a very low limit of detection for the specific peptide originated from the standard OPN digestion, so that it can be possible to analyse the digested peptide directly from biological samples, after purification and concentration. This could lead to a direct assessment of the mass difference, quantity difference (internal standards would be needed in this case) and any other feature that could experience a change from healthy to diseased individuals.

3.2.1 Analysis of Trypsin Digested Glycopeptides

The first digestions were conducted with trypsin, according to the procedure described in section 2.4.3. Figure 3.6 shows a spectrum of a trypsin-digested (TD) OPN. The list of the theoretical TD peptides from OPN can be found in table B.1 in appendix B.
All glycosylations should be contained in the peptide 62-143, whose backbone weight is 9179.6 Da. If the theoretical mass of the glycans is added, the total weight adds up to 13921.1 Da. As it is clear from the spectrum, no peaks were visible with a higher m/z than 8 kDa.

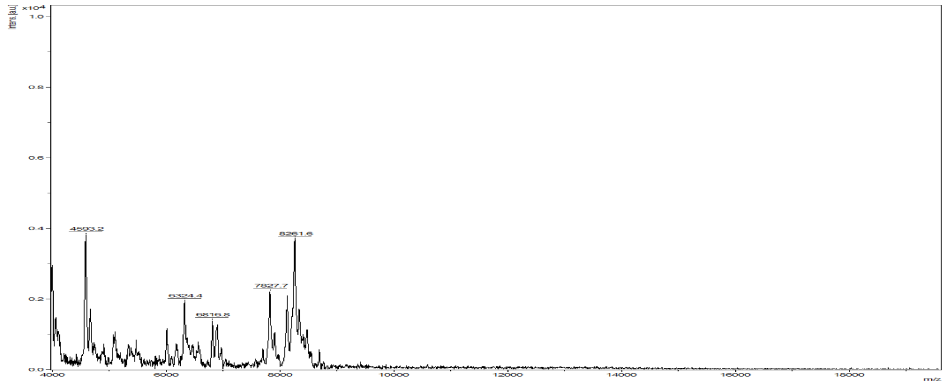

Figure 3.6: Mass spectrum of trypson digested osteopontin. Matrix: 2,5-dihydroxybenzoic acid 20 mg/ml. Sum of 3000 shots.

Analysis of Chymotrypsin Digested Glycopeptides

Since trypsin seemed not to give promising results, it was decided to focus on smaller peptides on which the glycans could be contained. After a short screening of possible proteolytic enzymes, it was noticed that chymotrypsin would be the most suitable one for the purpose, because of its selective cleavage after F, W, Y, M and L (not when L is preceded by P). This would lead to two small target peptides, one containing three glycans and one containing two. The former peptide is formed in residues 117-130 and the latter in 131-140, with respective backbone weights of 1551.8 Da and 1091.6 Da. By adding up three and two glycans respectively, the total masses become 4396.6 Da and 2988.1Da. Figure 3.7 shows a spectrum of a chymotrypsin digest (CD) of OPN.

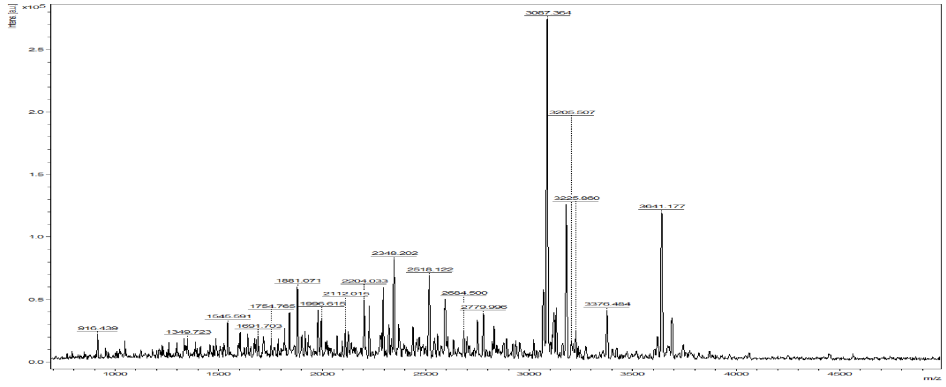

Figure 3.7: MS spectrum of chymotrypsin digested osteopontin. Matrix: 2,5-dihydroxybenzoic acid 20 mg/ml. Sum of 3000 shots.

Even though many of the peptides from table B.2 in appendix B are visible in the spectrum, none of the peaks of interest could be found. It was also considered that sialic acids tend to be easily released in DHB [49]; thereby producing peptides with potentially smaller masses. In table 3.5, the possible mass combinations are summarized for both peptides.
None of the peaks above were identified in MS analyses of OPN CD samples. A double check with the theoretical list of peptides with a single MC also gave no matches. It was concluded that the study of glycans through complex protein digest is a challenging task due to several reasons. Firstly, glycopeptides are often only a minority among the total of the peptides. Secondly, the signal intensities of glycopeptides are generally lower than those of non-glycosylated peptides, mainly due to the fact that the signal is distributed among a population of peptide species that bear heterogeneous glycan structures, and to their lower ionization efficiency. Lastly, glycopeptide signals are often suppressed in the presence of other

Table 3.4: Possible m/z of the target peptides in case of potential desialylations, because sialic acids are easyly liberated from glycans during analyses. The base structure for a single glycan is $[GalNAc-Gal-NeuNAc_2]$. Three glycans are supposed to be found on peptide 1, and two glycans on peptide 2. The subsequent mass losses are calculated removing 291.3 Da from the previous mass, that is the mass loss consequent to a sialic acid release.

Peptide	**Sugars**	**Exp. mass [Da]**
Pept 1 (1551.8)	$[GalNAc-Gal-NeuNAc_2]_3$	4396.6
	$[GalNAc-Gal-NeuNAc_2]_3-1NeuNAc$	4105.3
	$[GalNAc-Gal-NeuNAc_2]_3-2NeuNAc$	3814.0
	$[GalNAc-Gal-NeuNAc_2]_3-3NeuNAc$	3522.7
	$[GalNAc-Gal-NeuNAc_2]_3-4NeuNAc$	3231.5
	$[GalNAc-Gal-NeuNAc_2]_3-5NeuNAc$	2940.2
	$[GalNAc-Gal-NeuNAc_2]_3-6NeuNAc$	2648.9
Pept 2 (1091.6)	$[GalNAc-Gal-NeuNAc_2]_2$	2988.1
	$[GalNAc-Gal-NeuNAc_2]_2-1NeuNAc$	2696.8
	$[GalNAc-Gal-NeuNAc_2]_2-2NeuNAc$	2405.5
	$[GalNAc-Gal-NeuNAc_2]_2-3NeuNAc$	2114.2
	$[GalNAc-Gal-NeuNAc_2]_2-4NeuNAc$	1822.9

peptides, especially if glycans have sialic acid moieties. [62]

3.3 Glycan Analysis

3.3.1 Analysis of Chemically Released Glycans

As explained before in this work, analysing protein glycosylations is crucial to understand the roles of glycans in health and disease. For this purpose, many techniques and different approaches are currently used to tackle this kind of analyses. The study of glycopeptides of OPN described in the previous section was not successful for the goal; therefore, it was decided that a more appropriate way of studying the glycosylaton of OPN would be through direct release of the glycans. In contrast to N-glycans which are generally enzymatically released, the enzyme digestion of glycans for O-linked moieties is not so straightfoward. Only one enzyme is currently available for O-glycans liberation, and it cleaves solely core-1 and core-3 glycans. Therefore, it is not suitable for proteins whose glycosylation is not yet knmown. For this reason, O-glycans are generally released by chemical methods involving tedious procedures [63]. In this case, three chemical release methods were tried and compared, according to the β-elimination procedures described in the the papers of Gao *et al.* [57], Ozcan *et al.* [58], Turyan *et al.* [59], with respective purifications. No or little changes were made to the procedures, except that the volumes used for washing, conditioning and eluting were adapted to the unit volumes of the HyperSep Hypercarb™ cartridges (150 μL). After release, quenching, PGC purification and concentration of the three samples (see section 2.5.1 and 2.5.2), they were analysed in MALDI-TOF-MS. All glycan spectra were further analysed in the software Glycoworkbench (Imperial College London (London, England) for a software-assisted annotation of glycans.

Unlike peptides, glycans have low ionization efficiency. The $[M+H]^+$ ion is therefore not sufficiently abundant. However, neutral glycans can be more frequently detected as alkali metal adducts which ionize efficiently. In most cases, $[M+Na]^+$ is the predominant ion, usually followed by a weaker signal from a $[M+K]^+$ ion. [49] DHB was used as a matrix for glycans because it was the most suitable for these kind of samples according to manufacturer. On the other hand, DHB is not ideal for acidic glycans, as the ones bearing sialylations, because the detection limit is poor compared to neutral glycans [49]. Fragmentation is also most likely to happen with losses of carboxylic and/or whole sialic acid groups. In addition, sialylated glycans analyzed in the positive-ion mode yield a mixture of cation adducts producing multiple peaks. [49]. Since there is no reliable information on the glycosylation pattern of OPN, the goal of this part of the work is to test the hypothesis that the di-sialylated glycan shown in figure 1.2 is indeed the only O-glycan of OPN, or at least that it is present among its glycosylations.
Figure 3.8 shows the spectra recorded with the three methods.
The intact glycan shown in figure 1.2 would produce a peak at 966.3 m/z, *i.e.* [GalNAc-Gal-$NeuNAc_2$+H]$^+$). None of the purified spectra from the three method showed any 966 m/z peak, but both the samples released with the methods of Ozcan and Turyan showed a 697 m/z peak, which could be an indication of that same glycan with one desialylation and a sodium adduct, [GalNAc-Gal-NeuNAc+Na]$^+$. Also, the spectrum from the Turyan method shows a quite intense peak at 713.8 m/z, possible further confirmation of the same structure with a potassium adduct, *i.e.* [GalNAc-Gal-NeuNAc+K]$^+$.
In figure 3.9 a close-up of the bottom spectrum in figure 3.8 can be appreciated, where the 697.7 m/z and 713.8 m/z peaks are clearly visible.

Further experiments of samples released in this way sometimes lead to spectra with more peaks but lower S/N. In one of those, a pretty intense m/z 384 peak was recorded, which could indicate the presence of a fully desialylated unit of the intact glycan. The [GalNAc-Gal+H]$^+$ ion sums up to a mass of m/z 384.15. Its presence in the sample is not to be

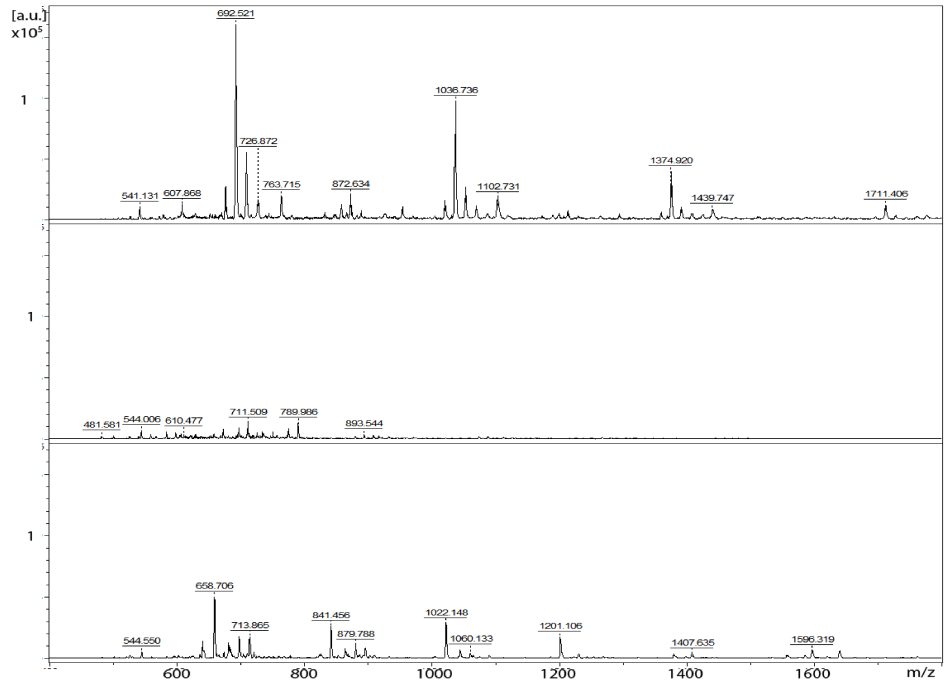

Figure 3.8: Spectra recorded of method I - (top), method II - (middle) and method III - (bottom) released glycan samples. Matrix: 2,5-dihydroxybenzoic acid 20mg/ml. Sum of 6000 shots in reflectron mode.

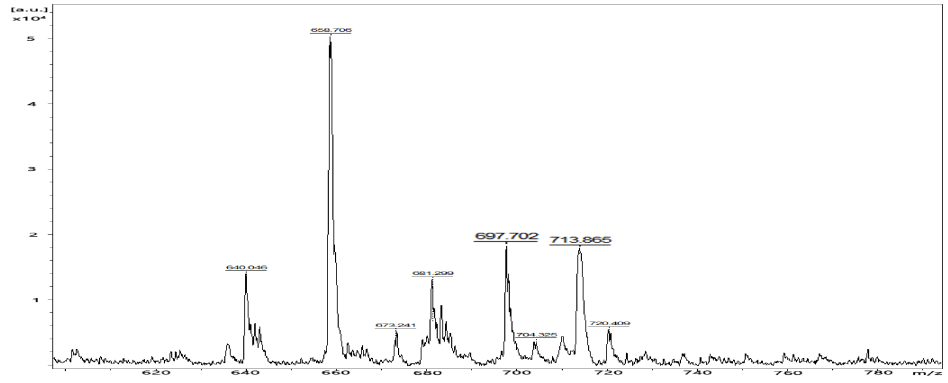

Figure 3.9: Close up of the bottom spectrum in figure 3.8, for the region: m/z 600 - m/z 800a. Matrix: 2,5-dihydroxybenzoic acid 20mg/ml. Sum of 6000 shots in reflectron mode.

excluded because of the easy release of sialic acids when included in DHB. The peak related to the GalNAc-Gal ion is shown on top of figure 3.10. The weaker peak at m/z 421 could indicate the same ion with a potassium adduct. Since in the lower mass range it is more likely to see peaks which originate from the matrix, a blank DHB spectrum was added to the bottom of figure 3.10 to show that the peak of interest is actually originating from the sample.

The peak at m/z 384 from the same spectrum was further analysed in tandem MS by means of a MALDI LIFT™ TOF-TOF MS technique intrinsic to the mass spectrometer. The fragment ion spectrum is shown in figure 3.11.
The tandem MS spectrum shows a very intense fragment ion at m/z 340.9. This indicates a neutral loss of m/z 43, which could be understood by considering the structure of GalNAc, shown in the top of figure 3.12.
The oxygen in the carbonyl group is easily ionized because of its two lone electron pairs. This creates a site of positive charge that can start an alpha cleavage, with a resulting neutral loss of m/z 43, and fragment ion of m/z 341, which is the intense peak visible in the spectrum.

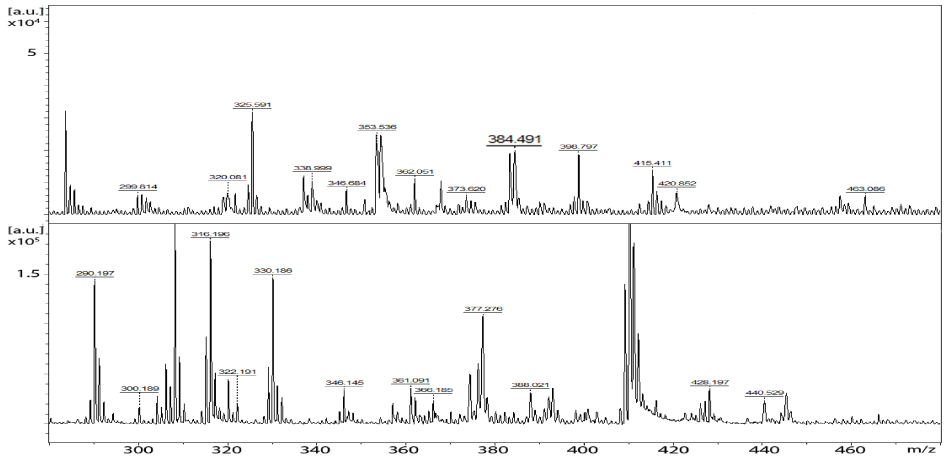

Figure 3.10: spectrum of a *top*: method I-released glycan sample zoomed in in the low mass region. The peak at m/z 384 could be a GalNAc-Gal protonated ion, and the one at m/z 421 could instead be the same molecule with a potassium adduct. Matrix: 2,5-dihydroxybenzoic acid 20mg/ml. Sum of 6000 shots in reflector mode. *Bottom*: 2,5-dihydroxybenzoic acid 20mg/ml blank matrix for comparison. Sum of 6000 shots in reflector mode.

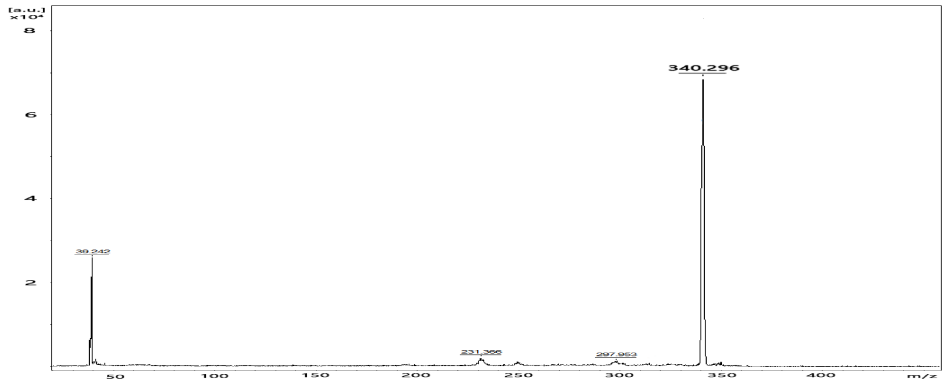

Figure 3.11: Fragment ion MS/MS spectrum of the 384 m/z peak in the Gao spectrum shown in figure 3.10. Matrix: 2,5-dihydroxybenzoic acid 20mg/ml.

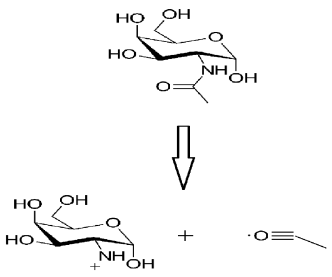

Figure 3.12: *Top*: molecular structure of GalNAc. *Bottom*: possible outcome of the α-cleavage fragmentation upon MS^2.

The bottom part of figure 3.12 shows the possible outcome of the fragmentation.

3.3.2 Evaluation of Purification Efficiency

The quality of the PGC releasing method was firstly investigated. Figure 3.13 shows a comparison of the same method III-treated OPN. The peak at m/z 699.0 in the top spectrum and 701.0 in the bottom one are from the same ion (by overlaying the spectra one can see that all peaks are simply shifted by approximately 2 m/z). The wrong m/z value of the peak as well as the mismatched signal before and after purification were due to bad calibration on the instrument previous to this analysis. However, the two spectra can still provide the reader with an idea of how PGC cartridges would not be optimal for glycan purifications: it is clear that this procedure did not contribute to any improvement for the targeted peaks. On the contrary, a lot of previously absent peaks (or anyway relatively lower than the peaks present before purification) showed up, and more noise was introduced. In the specific example, the S/N for the same ion (699.0 and 701.0) decreased from 13 to 11. This is a crucial decrease if one considers that samples were not only purified but also concentrated.

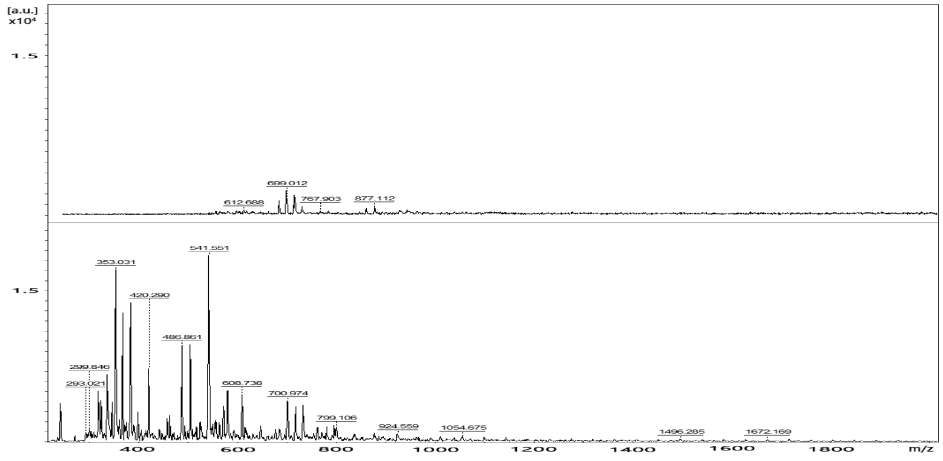

Figure 3.13: Before(top)/after(bottom) of a Turyan-purified glycan sample with PGC cartridges. Matrices: 20 mg/ml DHB for both spectra.

A Pierce® C18 purification was also tried according to the procedure described in section 2.5.3. Compared to graphitized carbon cartridges, a reversed procedure was adopted. In the case of graphitized carbon, glycans are attracted to the stationary phase because they are slightly polar. In that method they are in the eluted sample. With a C18 phase instead the process is reversed. In fact, there is low affinity between the stationary phase and glycans, hence they will not be retained, and they will instead be pipetted back into the vial after the sample is unloaded from the cartridge. An Ozcan sample was analysed in MALDI before and after Pierce® C18 purification. The comparison is shown in figure 3.14.
The improvement in the spectrum quality is evident. All extra peaks were eluted into waste after retention on the stationary phase, whereas the slightly polar glycans were not retained by the cartridge and were left behind in the vial. This resulted in an over 4-fold increase in the S/N value of the $[GalNAc\text{-}Gal\text{-}NeuNAc+Na]^+$ peak at 696.9 m/z (93.3 ± 68.6,n=3, before the treatment, 389.3 ± 185.0, n=3, after purification), and more than 7-fold increase in the S/N value of the $[GalNAc\text{-}Gal\text{-}NeuNAc+K]^+$ peak (16.7 ± 13.3, n=3, before, 125.0 ± 79.3, n=3, after). Also, the mass difference between the m/z 696.9 peak and the 534.3 one is 162.6, which indicates a potential loss of a Gal unit (mass: 162.16 Da). So the m/z 534.3 could be another glycan of the type $[GalNAc\text{-}NeuNAc+Na]^+$. Figure 3.15 shows a zoom of the same spectra in figure 3.14 in the region around m/z 700, in order to demonstrate how the resolution of weaker peaks was also improved.

Before any purification, glycans were detectable but with lower sensitivity compared to the

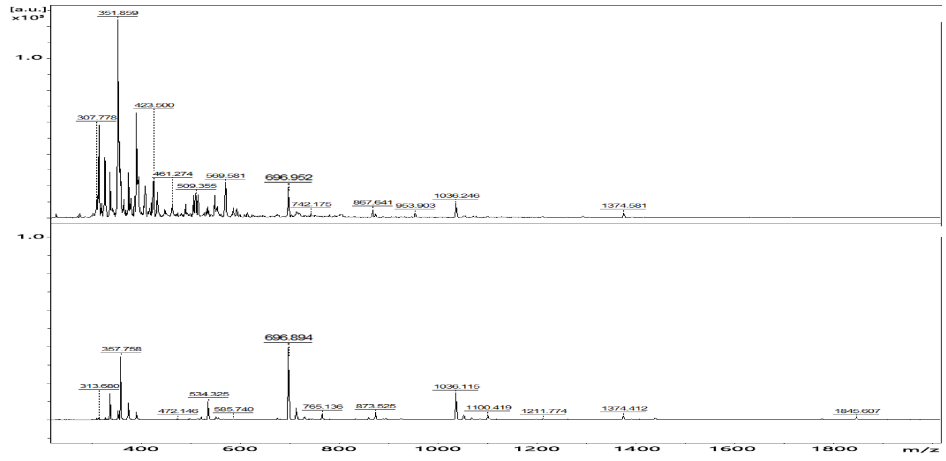

Figure 3.14: Before(top)/after(bottom) of an Ozcan-purified glycan sample with Pierce® C18 cartridges. Matrices: DHB 20 mg/ml for both spectra. Each is a sum of 6000 shots in reflector mode.

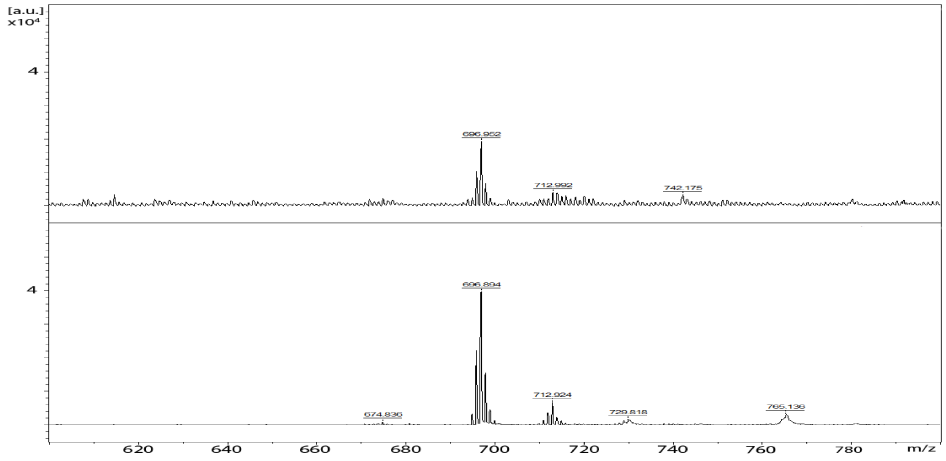

Figure 3.15: Close-up in the relevant region m/z 640 - m/z 740 for a Pierce® C18 purified Ozcan sample (Before(top)/after(bottom). Matrix: 2,5-dihydroxybenzoic acid 20 mg/ml. Sum of 6000 shots in reflectron mode.

purified sample. After purification, even weaker signals gained sufficient intensity to be considered peaks. In the specific case, the peak at m/z 674.8 indicates the same mono-desialylated glycan with no metal adducts, which is $[\text{GalNAc-Gal-NeuNAc+H}]^+$. It is not common to see the protonated glycan in MALDI-MS [49], and this is why the peaks with Na adducts and K adducts (m/z 696.9 and m/z 712.9) are more intense, whereas the one at m/z 675.0 is very weak. On the other hand, the S/N to noise for this latter peak was 19.0 ± 11.4, which is more than six times the cutoff value of three for a signal to be assigned to a peak. Before purification, this signal was confused with the noise in each of the spectra used for this assessment. Table 3.5 shows a summary of the relevant peaks found in the work that can represent a confirmation of the hypothesis of Christensen *et al.*, 2008.

Irrespective of the efficiency of purification, all methods (both PGC and C18) improved the crystallization of DHB on the MALDI plate. To assess this improvement, three MALDI spots were compared for each method: one containing the non purified released glycan sample, one containing the purified glycan sample, and one containing the wash (for example from the 600 μL of water used to wash in method II, section 2.5.2). From a comparison of the crystallization on the MALDI spot of glycan samples before and after purification, there is a big difference in the quality of crystals: purified samples with both techniques showed a neat

Table 3.5: Recap of the peaks found in the MALDI-TOF-MS analysis of glycans released from OPN. The glycan structure to be confirmed is $[GalNAc - Gal - NeuNAc_2]$.

m/z	Glycan	Adduct	Structure
384	[GalNAc-Gal]	H^+	
675	[GalNAc-Gal-NeuNAc]	H^+	
697	[GalNAc-Gal-NeuNAc]	Na^+	
713	[GalNAc-Gal-NeuNAc]	K^+	

and standard crystallization. This is because the purification step takes away salts, which tend to hinder a good shaping of the crystals. A proof of this is in the right picture of figure 3.16, *i.e.* the wash, that also crystallized badly, because the salts that were present in the non-treated sample were washed away in that step.

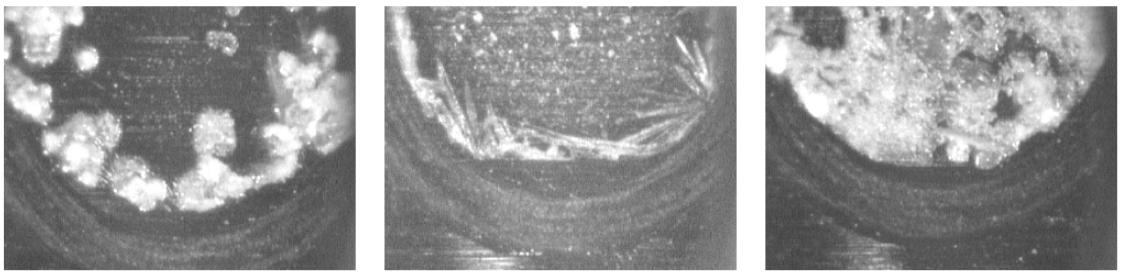

Figure 3.16: Snapshot of the bottom areas of three different anchors. From left to right: deposited sample before purification with C18, deposited sample after purification with C18, deposited elution. Matrix: 2,5-dihydroxybenzoic acid 20 mg/ml.

4. Conclusions and Future Outlooks

In the first part of the present work, a reliable method for protein fractionation has been developed and optimized, by the hyphenation of a CE system with MALDI-MS, through an offline, robot-assisted coupling of the systems. Fractionated deposition was performed with the help of an Excel deposition sheet, which was used from run to run to calculate all relevant data on single fractionations. The method proved good efficiency in the fractionation of proteins, both with very different and very similar pIs, thereby demonstrating very good versatility. In the second part of the work, the focus was shifted to OPN and its glycosylation. The investigation of glycopeptides and/or glycans directly from biological samples turned out to be too demanding, because of the extremely low concentrations of OPN in plasma, which would require strong concentration and purification methods prior to any kind of analyses. The study was therefore performed on standard OPN, which allowed to work with higher concentrations for development of suitable procedures. The first approach was to enzymatically digest the protein and try to analyse potentially glycosylated peptides in MALDI-TOF-MS. Trypsin and chymotrypsin were used to tackle the problem in different ways, but neither brought to a signal for the relevant peptides upon MS analysis. The second approach was a direct study of OPN glycans after their chemical release, for direct characterization. Multiple cues were provided that can be considered as a possible confirmation of the core-1 GalNAc-Gal-$NeuNAc_2$ being the glycan (or at least among the glycans) of OPN. Among them, the detection of a peak corresponding to the protonated core-1 structure, as well as the detection of peaks indicating mono-desialylated glycan structures with and without alkali metal adducts. An indirect purification method was also proposed, which brought to an average increase of signal to noise ratio to up to 7 times (for the mono-desialylated glycan with a potassium adduct), and it made possible the detection of other formerly hidden peaks.
The lack of a well-developed concentration method for OPN, as well as the lack of reliable information on which glycans to look for in the glycosylation study, made the analysis on biological samples an inaccessible path. Nevertheless, this is what the initial aim of the work was, and it is crucial to keep working in this direction to eventually reach the fixed goal.
The endpoint of the present work would be the union of the two main blocks that were faced in the thesis. More specifically, it would be extremely helpful if glycans could be firstly CE-separated, and then spotted on the MALDI plate through the fractionation method. If this idea worked, it would sensibly help the understanding of the glycosylation of OPN, as well as that of many other proteins: the CE would act as a purification step for glycans prior to deposition, and glycans could be isolated on single anchors, thus allowing for single-glycan analysis which would lead to enhanced detectability. The union of these two parts has, however, some major obstacles. Firstly, glycans are not UV-absorbant. Therefore extra steps are needed either in a direct way (*e.g.* derivatization, which is anyway not recommendend because it often brings to the release of sialic acids) or indirect (*e.g.* making the CE buffer UV absorbant, that is indirect UV absorption). Secondly, there is not much work done on glycans in CE; therefore all the methods should be adapted to suit best to this kind of sample. Thirdly, even supposing that a proper concentration method will be found, there is no certainty that the LOD of CE would be low enough to ensure the detection of glycans. A

sufficient detectability is needed because it is functional for their fractionation.

A complementary study to the CE-MALDI-MS analysis of released glycans could be LC-ESI-MS analysis of glycopeptides. Glycopeptides did not work because MALDI-MS is not very suitable for those species, especially if the peptides are heavy and the glycan moieties are sialylated. Many have instead worked with glycopeptides in MS by means of an electrospray ionisation (ESI) source, due to the advantages that one can obtain by online coupling with CE and especially LC. This combined approach could provide a clearer idea of all possible glycans present in OPN: in this work the focus was put on one structure only, because this structure (figure 1.2) was presented in the literature as the most likely to be the only glycan of OPN. On the other hand, there have been some works out of the choirs that disagreed with this hypothesis; further investigation is therefore needed.
It could also be interesting to investigate how low in concentration could the analyses presented go. In this work, focus was never put on concentrations, because the main goal of the work became to understand the glycosylation of OPN, from a qualitative standpoint. The target was to understand if the glycan GalNAc-Gal-$NeuNAc_2$ was actually present in OPN. A future possibility to gain further confirmation for the presence of the core-1 glycan would be the addition of some Lithium salts to the sample. $[M+Li]^+$ could in fact be further proof of glycan presence, even though they are not so common as sodium and potassium adducts. Therefore, it could be interesting to compare the same glycan sample with and without lithium salts, to assess if extra peak are showing up. The same idea could be exploited by simply adding sodium salts, which would create more adducts, and assessing the difference in the $[M+Na]^+$ signal by quantification with *e.g.* internal standards.

With this project, new information on the glycosylation pattern of OPN was provided, which can hopefully add up to the already available knowledge upon the involvement of the glycoprotein with bone formation and resorption, and possibly lead the way towards its establishment as a biomarker for bone-conditions.

5. Acknowledgments

I would like to thank the group of Analytical Chemistry in KTH for the chance it has given me to grow as a student and as a scientist, during the past six months. in particular, I thank professor Åsa Emmer, as well as Leila Josefsson and Joakim Romson for the thorough and educational supervision. I would also like to thank Yuye Zhou for her help in and out of the lab. I would like to thank the division of Organic Chemistry of KTH for providing some of the chemical that I needed for my experiments. In the same way, I express my thankfulness to Alessandro Quaranta, Stockholm University, for the supervision and the advices during the early stage of my glycan analyses. Moreover, I would like to thank my Italian supervisor, prof. Francesca Bosco, PoliTo, for the long-distance correspondence during the work in KTH. Finally, I cannot miss to express my gratitude to my closest friends, my girlfriend and my family, for the their trust, their support and their love.

Abbreviations

4-HCCA	α-cyano-4-hydroxycinnamic acid
Ala (A)	Alanine
Arg (R)	Arginine
Asn (N)	Asparagine
Asp (D)	Aspartic Acid
BGE	Background Electrolyte Solution
CE	Capillary Electrophoresis
CRPS	Complex Regional Pain Syndrome
CD	Chymotrypsin Digest
Cys (C)	Cystein
Cyt-C	Cytochrome C
CZE	Capillary Zone Electrophoresis
DHB	2,5-dihydroxybenzoic acid
EOF	Electroosmotic Flow
ESI	Electrospray Ionization
EtOH	Ethanol
Fuc	Fucose
Gal	Galactose
GalNAc	N-Acetylgalactosamine
Glc	Glucose
GlcNAc	N-Acetylglucosamine
Gln (Q)	Glutamic Acid
Glu (E)	Glutamine
Gly (G)	Glycine
His (H)	Histidine
HPC	Hydroxypropyl Cellulose
Ile (I)	Isoleucine
Lact	α-Lactalbumin
Leu (L)	Leucine
Lys (K)	Lysine
Lys	Lysozyme
LOD	Limit of Detection
m/z	Mass to Charge Ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization
Man	Mannose
MC	Miscleavages
Met (M)	Methionine
MS	Mass Spectrometry
NaOH	Sodium Hydroxide
NeuNAc	N-Acetylneuraminic acid
OPG	Osteoprotegerin
OPN	Osteopontin
PGC	Porous Graphitized Carbon
Phe (F)	Phenylalanine
pI	Isoelectric Point
Pro (P)	Proline
PTM	Post Translational Modification(s)

RSD	Relative Standard Deviation
SA	Sinapinic Acid
Ser (S)	Serine
SPE	Solid Phase Extraction
TD	Trypsin Digest
Thr (T)	Threonine
TA30	30:70 %v Acetonitrile:Trifluoroacetic Acid 0,1% in water
TFA	Trifluoroacetic Acid
TOF	Time of Flight
Tris	Tris(hydroxymethyl)aminomethane
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
UV	Ultra-Violet
Val (V)	Valine
μ	Mobility

Bibliography

[1] H. J. An, S. R. Kronewitter, M. L. de Leoz, and C. B. Lebrilla, "Glycomics and disease markers," *Curr Opin Chem Biol*, vol. 13, no. 5-6, pp. 601–7, 2009.

[2] B. Blomme, C. V. Steenkiste, N. Callewaert, and H. V. Vlierberghe, "Alteration of protein glycosylation in liver diseases," *Journal of Hepatology*, vol. 50, no. 3, pp. 592 – 603, 2009.

[3] C. Ferreira, E. Lo Turco, S. Saraiva, R. Bertolla, F. Perecin, G. Souza, M. Murgu, J. Garcia, S. Cortezzi, F. Meirelles, C. Klitzke, E. Cabral, M. Miglino, P. Porciuncula, C. Leal, E. Borges Jr., D. Martins, C. Ambrosio, F. D´Alexandri, L. Smith, and M. Eberlin, "Proteomics, metabolomics and lipidomics in reproductive biotechnologies: The ms solutions," *Acta Scientiae Veterinariae*, vol. 38, pp. s591–s603, 2010.

[4] P. R. Graves and T. A. J. Haystead, "Molecular biologist's guide to proteomics," *Microbiology and Molecular Biology Reviews*, vol. 66, no. 1, pp. 39–63, 2002.

[5] R. Beetham, "Introduction to proteomics: Tools for the new biology," *Annals of Clinical Biochemistry*, vol. 39, p. 538, 2002.

[6] J. Seo, K.-J. Lee *et al.*, "Post-translational modifications and their biological functions: proteomic analysis and systematic approaches," *Journal of biochemistry and molecular biology*, vol. 37, no. 1, pp. 35–44, 2004.

[7] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, vol. 21, 2003.

[8] N. Tavernarakis, *Protein metabolism and homeostasis in aging*. Springer Science & Business Media, 2011, vol. 694.

[9] J. MacNeil, "Better biomarkers for the diagnostics labyrinth," *Genome Technol*, vol. 15, pp. 24–33, 2004.

[10] T. E. Angel, U. K. Aryal, S. M. Hengel, E. S. Baker, R. T. Kelly, E. W. Robinson, and R. D. Smith, "Mass spectrometry-based proteomics: existing capabilities and future directions," *Chem Soc Rev*, vol. 41, no. 10, pp. 3912–28, 2012.

[11] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, no. 6928, p. 198, 2003.

[12] T. M. Devlin, *Textbook of biochemistry: with clinical correlations*. Wiley-Liss, 2006.

[13] T. F. Scientific, "Overview on post-translational modifications (ptms)," last access: 10/02/2018. [Online]. Available: https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html

[14] G. P. Moss, P. A. S. Smith, and D. Tavernier, "Glossary of class names of organic compounds and reactivity intermediates based on structure (iupac recommendations 1995)," *Pure and Applied Chemistry*, vol. 67, no. 8-9, 1995.

[15] R. J. Owens and J. E. Nettleship, *Functional and Structural Proteomics of Glycoproteins*. Springer Netherlands, 2010.

[16] A. Ceroni, K. Maass, H. Geyer, R. Geyer, A. Dell, and S. M. Haslam, "Glycoworkbench: a tool for the computer-assisted annotation of mass spectra of glycans," *Journal of proteome research*, vol. 7, no. 4, pp. 1650–1659, 2008.

[17] S. K. Niazi, *Biosimilars and Interchangeable Biologics: Tactical Elements*. CRC Press, 2016.

[18] G. Y. Wiederschain, "Essentials of glycobiology," *Biochemistry (Moscow)*, vol. 74, no. 9, pp. 1056–1056, 2009.

[19] G. Littlejohn, "Complex regional pain syndrome," *Rheumatology*, vol. 53, no. 7, pp. 1157–1158, 2014.

[20] R. N. Harden, S. Bruehl, M. Stanton-Hicks, and P. R. Wilson, "Proposed new diagnostic criteria for complex regional pain syndrome," *Pain Medicine*, vol. 8, no. 4, pp. 326–331, 2007.

[21] J. Malby, "Putting out the flames of complex regional pain syndrome - updated guide to crps/rsd," 2016, last access: 05/01/2018. [Online]. Available: http://princessinthetower.org/complex-regional-pan-syndrome-crpsrsd/

[22] R. N. Harden, S. Bruehl, R. S. Perez, F. Birklein, J. Marinus, C. Maihofner, T. Lubenow, A. Buvanendran, S. Mackey, J. Graciosa *et al.*, "Validation of proposed diagnostic criteria (the "Budapest Criteria") for complex regional pain syndrome," *Pain*, vol. 150, no. 2, pp. 268–274, 2010.

[23] S. Bruehl, "Complex regional pain syndrome," *BMJ*, vol. 351, p. h2730, 2015.

[24] H. H. Kramer, L. C. Hofbauer, G. Szalay, M. Breimhorst, T. Eberle, K. Zieschang, M. Rauner, T. Schlereth, M. Schreckenberger, and F. Birklein, "Osteoprotegerin: a new biomarker for impaired bone metabolism in complex regional pain syndrome?" *Pain*, vol. 155, no. 5, pp. 889–95, 2014.

[25] D. T. Denhardt and M. Noda, "Osteopontin expression and function: Role in bone remodeling," *Journal of Cellular Biochemistry*, vol. 72, no. S30-31, pp. 92–102.

[26] T. Standal, M. Borset, A. Sundan *et al.*, "Role of osteopontin in adhesion, migration, cell survival and bone remodeling," *Exp Oncol*, vol. 26, no. 3, pp. 179–84, 2004.

[27] A. Altintas, G. Saruhan-Direskeneli, G. Benbir, M. Demir, and S. Purisa, "The role of osteopontin: a shared pathway in the pathogenesis of multiple sclerosis and osteoporosis?" *J Neurol Sci*, vol. 276, no. 1-2, pp. 41–4, 2009.

[28] M. Cabiati, M. Gaggini, M. M. Cesare, C. Caselli, P. De Simone, F. Filipponi, G. Basta, A. Gastaldelli, and S. Del Ry, "Osteopontin in hepatocellular carcinoma: A possible biomarker for diagnosis and follow-up," *Cytokine*, vol. 99, pp. 59–65, 2017.

[29] C. M. Giachelli and S. Steitz, "Osteopontin: a versatile regulator of inflammation and biomineralization," *Matrix Biology*, vol. 19, no. 7, pp. 615 – 622, 2000.

[30] P. Kivisakk, B. C. Healy, K. Francois, R. Gandhi, T. Gholipour, S. Egorova, V. Sevdalinova, F. Quintana, T. Chitnis, H. L. Weiner, and S. J. Khoury, "Evaluation of circulating osteopontin levels in an unselected cohort of patients with multiple sclerosis: relevance for biomarker development," *Mult Scler*, vol. 20, no. 4, pp. 438–44, 2014.

[31] M. Stilund, M. C. Gjelstrup, T. Petersen, H. J. Moller, P. V. Rasmussen, and T. Christensen, "Biomarkers of inflammation and axonal degeneration/damage in patients with newly diagnosed multiple sclerosis: contributions of the soluble cd163 csf/serum ratio to a biomarker panel," *PLoS One*, vol. 10, no. 4, p. e0119681, 2015.

[32] M. Vogt, S. Floris, J. Killestein, D. Knol, M. Smits, F. Barkhof, C. Polman, and L. Nagelkerken, "Osteopontin levels and increased disease activity in relapsing-remitting multiple sclerosis patients," *Journal of neuroimmunology*, vol. 155, no. 1-2, pp. 155–60, October 2004.

[33] S. Ohshima, H. Kobayashi, N. Yamaguchi, K. Nishioka, M. Umeshita-Sasai, T. Mima, S. Nomura, S. Kon, M. Inobe, T. Uede, and Y. Saeki, "Expression of osteopontin at sites of bone erosion in a murine experimental arthritis model of collagen-induced arthritis: Possible involvement of osteopontin in bone destruction in arthritis," *Arthritis & Rheumatism*, vol. 46, no. 4, pp. 1094–1101, 2002.

[34] P. K. Petrow, K. M. Hummel, J. Schedel, J. K. Franz, C. L. Klein, U. Müller-Ladner, J. Kriegsmann, P.-L. Chang, C. W. Prince, R. E. Gay, and S. Gay, "Expression of osteopontin messenger rna and protein in rheumatoid arthritis: Effects of osteopontin on the release of collagenase 1 from articular chondrocytes and synovial fibroblasts," *Arthritis & Rheumatism*, vol. 43, no. 7, pp. 1597–1605, 2000.

[35] Y. Kariya, M. Kanno, K. Matsumoto-Morita, M. Konno, Y. Yamaguchi, and Y. Hashimoto, "Osteopontin o-glycosylation contributes to its phosphorylation and cell-adhesion properties," *Biochemical Journal*, vol. 463, no. 1, pp. 93–102, 2014.

[36] B. Ek-Rylander, M. Flores, M. Wendel, D. Heinegård, and G. Andersson, "Dephosphorylation of osteopontin and bone sialoprotein by osteoclastic tartrate-resistant acid phosphatase. modulation of osteoclast adhesion in vitro." *Journal of Biological Chemistry*, vol. 269, no. 21, pp. 14 853–14 856, 1994.

[37] B. Christensen, M. S. Nielsen, K. F. Haselmann, T. E. Petersen, and E. S. Sorensen, "Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five o-glycosylation sites and their biological implications," *Biochem J*, vol. 390, no. Pt 1, pp. 285–92, 2005.

[38] B. Christensen, T. E. Petersen, and E. S. Sørensen, "Post-translational modification and proteolytic processing of urinary osteopontin," *Biochemical Journal*, vol. 411, no. 1, pp. 53–61, 2008.

[39] M. Keykhosravani, A. Doherty-Kirby, C. Zhang, D. Brewer, H. A. Goldberg, G. K. Hunter, and G. Lajoie, "Comprehensive identification of post-translational modifications of rat bone osteopontin by mass spectrometry," *Biochemistry*, vol. 44, no. 18, pp. 6990–7003, 2005.

[40] B. Christensen, C. C. Kazanecki, T. E. Petersen, S. R. Rittling, D. T. Denhardt, and E. S. Sørensen, "Cell type-specific post-translational modifications of mouse osteopontin are associated with different adhesive properties," *Journal of Biological Chemistry*, vol. 282, no. 27, pp. 19 463–19 472, 2007.

[41] D. S. Bautista, Z. Saad, A. F. Chambers, K. S. Tonkin, F. P. O'Malley, H. Singhal, S. Tokmakejian, V. Bramwell, and J. F. Harris, "Quantification of osteopontin in human plasma with an elisa: basal levels in pre-and postmenopausal women," *Clinical biochemistry*, vol. 29, no. 3, pp. 231–239, 1996.

[42] H. H. Lauer and G. P. R., *High Perfomance Capillary Electrophoresis. A Primer.* Agilent Technologies, 2014.

[43] J. Horvath and V. Dolnik, "Polymer wall coatings for capillary electrophoresis," *Electrophoresis*, vol. 22, no. 4, pp. 644–55.

[44] M. Biacchi, R. Gahoual, N. Said, A. Beck, E. Leize-Wagner, and Y.-N. François, "Glycoform separation and characterization of cetuximab variants by middle-up off-line capillary zone electrophoresis-uv/electrospray ionization-ms," *Analytical chemistry*, vol. 87, no. 12, pp. 6240–6250, 2015.

[45] J. Gross, *Mass Spectrometry: A Textbook.* Springer International Publishing, 2017.

[46] D. Harris, *Quantitative Chemical Analysis.* W. H. Freeman, 2010. [Online]. Available: https://books.google.se/books?id=tHR4nsLDFAwC

[47] B. Syed, "Maldi-tof," chemistry LibreText® - Last access: 11/02/2018. [Online]. Available: https://chem.libretexts.org/Core/Analytical_Chemistry/Instrumental_Analysis/Mass_Spectrometry/MALDI-TOF

[48] H. W. Ressom, R. S. Varghese, L. Goldman, C. A. Loffredo, M. Abdel-Hamid, Z. Kyselova, Y. Mechref, M. NOVOTNY, and R. Goldman, "Analysis of maldi-tof mass spectrometry data for detection of glycan biomarkers," in *Biocomputing 2008.* World Scientific, 2008, pp. 216–227.

[49] I. Sigma-Aldrich, "Mass spectrometry of glycans," last access: 14/02/2018. [Online]. Available: https://www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/mass-spectrometry-of-glycans.html

[50] Y. Huang, Y. Mechref, and M. V. Novotny, "Microscale nonreductive release of o-linked glycans for subsequent analysis through maldi mass spectrometry and capillary electrophoresis," *Analytical chemistry*, vol. 73, no. 24, pp. 6063–6069, 2001.

[51] S. Romand, J. Schappler, J. L. Veuthey, P. A. Carrupt, and S. Martel, "cief for rapid pka determination of small molecules: a proof of concept," *Eur J Pharm Sci*, vol. 63, pp. 14–21, 2014.

[52] Y. Shen, F. Xiang, T. D. Veenstra, E. N. Fung, and R. D. Smith, "High-resolution capillary isoelectric focusing of complex protein mixtures from lysates of microorganisms," *Analytical Chemistry*, vol. 71, no. 23, pp. 5348–5353, 1999.

[53] D. T. Sawyer, J. M. Beebe, and W. R. Heineman, *Chemistry experiments for instrumental methods.* John Wiley & Sons,, 1984.

[54] J. Romson, J. Jacksén, and Å. Emmer, "Simple and environmentally friendly fabrication of superhydrophobic alkyl ketene dimer coated maldi concentration plates," *Journal of The American Society for Mass Spectrometry*, vol. 28, no. 8, pp. 1733–1736, Aug 2017.

[55] I. Sigma-Aldrich, "Product information - trypsin from porcine pancreas proteomics grade, bioreagent, dimethylated," last access: 29/01/2018. [Online]. Available: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/t6567bul.pdf

[56] ——, "Chymotrypsin," last access: 29/01/2018. [Online]. Available: https://www.sigmaaldrich.com/life-science/metabolomics/enzyme-explorer/analytical-enzymes/chymotrypsin.html

[57] Y. Gao, F. Ma, W. Zhang, F. Zhong, H. Tang, D. Xu, and L. Zhao, "O-glycan profiling of serum glycan for potential renal cancer biomarkers," *Sci China Life Sci*, vol. 56, no. 8, pp. 739–44, 2013.

[58] S. Ozcan, H. J. An, A. C. Vieira, G. W. Park, J. H. Kim, M. J. Mannis, and C. B. Lebrilla, "Characterization of novel o-glycans isolated from tear and saliva of ocular rosacea patients," *J Proteome Res*, vol. 12, no. 3, pp. 1090–100, 2013.

[59] I. Turyan, X. Hronowski, Z. Sosic, and Y. Lyubarskaya, "Comparison of two approaches for quantitative o-linked glycan analysis used in characterization of recombinant proteins," *Anal Biochem*, vol. 446, pp. 28–36, 2014.

[60] M. e. O'Neil, *The Merck Index - An Encyclopedia of Chemicals, Drugs, and Biologicals.* Cambridge, UK: Royal Society of Chemistry, 2013.

[61] E. S. Sørensen, T. E. Petersen, and P. Højrup, "Posttranslational modifications of bovine osteopontin: Identification of twenty-eight phosphorylation and three o-glycosylation sites," *Protein Science*, vol. 4, no. 10, pp. 2040–2049, 1995.

[62] H. Geyer and R. Geyer, "Strategies for analysis of glycoprotein glycosylation," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1764, no. 12, pp. 1853 – 1869, 2006.

[63] G. Zauner, C. A. Koeleman, A. M. Deelder, and M. Wuhrer, "Mass spectrometric o-glycan analysis after combined o-glycan release by beta-elimination and 1-phenyl-3-methyl-5-pyrazolone labeling," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1820, no. 9, pp. 1420 – 1428, 2012.

All molecular structures were taken from https://pubchem.ncbi.nlm.nih.gov/. PIs and digested peptides were computed theoretically through the "PeptideMass" tool in https://expasy.org.

A. Deposition Times

The optimized deposition method involves a pressure driven ejection of the samples from the capillary. The pressure used was 0.125 bar. The velocity of any species in the capillary was estimated experimentally trough an average of the migration times of quinine in three pressure-driven runs. The average migration time resulted to be (9.364±0.548) min over three runs, which yields a pressure velocity of $40/(9.364 * 60) = 0.0712$ cm/s. Therefore, during depositions, analytes would only be ejected from the capillary through a hydrodynamic push of 0.0712 cm/s. The electrophoretic velocities of cytochrome C (Cyt-C), lysozyme (Lys) and α-lactalbumin (Lact) were calculated starting from an average of their recorded mobilities (through equation 1.1). Velocities were in turn used to assess the extension in width of a single peak (*e.g.* a peak of an analyte that has an electrophoretic velocity of 0.1 cm/s and that is 10 seconds broad will occupy 1 cm in the capillary). The values of the averaged mobilities for the proteins are:

Cyt-C: $(2.87 \pm 0{,}12)\cdot 10^{-4}$ cm^2/(Vs) - see section 3.1.2.

Lys: $(2.80 \pm 0{,}12)\cdot 10^{-4}$ cm^2/(Vs) - see section 3.1.2.

Lact: $(1.35 \pm 0{,}02)\cdot 10^{-4}$ cm^2/(Vs) - average on three runs.

Table A.1 describes how the deposition times were calculated in the Excel sheet.
Because Cyt-C and Lys have very similar mobilities, they migrate approximately at the same time. Therefore, it was decided to let the deposition sheet calculate only the starting deposition time for Cyt-C, and then let the robot deposit on few subsequent anchors, so that both Cyt-C and Lys would be involved in the deposition. Since the expected elution of Cyt-C was approximately 0.8 min, and the one for Lys was approximately 1.3 min, it was decided to use 3 anchors for each of the two proteins, and deposit for 30 seconds each (*i.e.* 1.5 min in total per protein). It was therefore sufficient to calculate the exact starting time for the fractionation, and the first two proteins would be automatically deposited within the first 6 anchors. On the contrary, the information on Lact could not be given as an input in every run because Cyt-C and Lys always reached the outlet of the capillary before Lact would reach the detection window. The migration of Lact was therefore calculated as a sum of an electrophoretic contribution (up to the moment when deposition started) and a hydrodynamic one (from the start of deposition). The peak width was as well provided as a parameter based on the width recorded in lactalbumin from previous runs, equivalent to a time space of approximately one minute at the detection window.
Hence, in the case of Lact it was crucial to assess how long after Lys would Lact reach the end of the capillary. A waste was used to collect the BGE that came out after Lys: a parameter of 185 seconds was provided to the deposition sheet as the value for t_{w1}, as the waste collection would always start after the $6 \cdot 30$ s depositions of the two proteins (a 5 seconds extra was included to account for the time that the robot takes to lift the capillary outlet and move it to the next anchor). The waste was distributed on 2 anchors so that eventual residues of lagging Lys would never be on the same waste anchor as eventual residues of leading Lact.

After the two wastes, Lact had a fixed deposition time of 200 s.

Table A.1: Description of the main values used to program depositions through the robotic arm. The "type" column indicates "input" if the value is to be provided as a new entry for every run, "output" if it is calculated from the sheet, "param" if it is a fixed value. Peak broadnesses were assumed to grow linearly with time (*i.e.* longitudinal diffusion $\propto$ time)

	Symbol	Unit	Type	Formula	Description
	v_p	cm/s	param	-	pressure velocity
Cyt-C	v_c	cm/s	param	-	Electrophoretic velocity
	$t_{1/2c}$	min	input	-	Migration time to detection window
	$b_{1/2c}$	min	input	-	Broadness of peak at detection window
	t_c	min	output	$t_{1/2c} \cdot 2$	Migration time to capillary outlet
	b_c	min	output	$b_{1/2c} \cdot 2$	Broadness of peak at capillary outlet
	t_{el-c}	min	output	$b_c \cdot v_c / v_p$	Time for the peak to elute from the outlet
	t_{start}	min	output	$t_c - b_c/2 - 0.5$	Start of deposition, *i.e.* 30 s before Cyt-C reaches the outlet
Lys	v_{ly}	cm/s	param	-	Electrophoretic velocity of Lysozyme
	$t_{1/2ly}$	min	input	-	Migration time to detection window
	$b_{1/2ly}$	min	input	-	Broadness of peak at detection window
	t_{ly}	min	output	$t_{1/2ly} \cdot 2$	Migration time to capillary outlet
	b_{ly}	min	output	$b_{1/2ly} \cdot 2$	Broadness of peak at capillary outlet
	t_{el-ly}	min	output	$b_{ly} \cdot v_l y / v_p$	Time for the peak to elute from the outlet
Waste & Lact	v_{la}	cm/s	param	-	Electrophoretic velocity of Lysozyme
	t_{la}	min	output	$\frac{t_{start}+80-(t_{start}60v_{la})/v_p}{60}$	Migration time to capillary outlet
	b_{la}	min	param	-	Broadness of peak at capillary outlet
	t_{el-la}	min	param	$b_{la} \cdot v_l a / v_p$	Time for the peak to elute from the outlet
	t_{dist}	min	output	$t_{la} - t_c$	Time between elutions of Cyt-C and Lact
	t_{w1}	min	param	-	Start of the first waste deposition
	$\underline{d_{w1}}$	min	output	$(t_{dist} - t_{w1} - 90)/2$	Duration of the first waste deposition
	t_{w2}	min	output	$t_{w1} + d_{w1}$	Start of the second waste deposition
	d_{w2}	min	output	d_{w1}	Duration of the second waste deposition
	$t_{start-la}$	min	output	$t_{w1} + d_{w1} + d_{w2}$	Start of lactalbumin deposition

90 seconds were subtracted to the start of lactalbumin deposition (d_{w1}, underlined in A.1) to account for the parameters of lactalbumin migrations which were based on previous experiments rather than taken from the same runs. This anticipation, together with a sufficient prolonging of the deposition time for lactalbumin, provided a sufficient slack to ensure that lactalbumin would actually be deposited on the right anchor.

The volumes of sample deposited with this method were calculated using the Hagen-Poiseuille equation:

$$V = \frac{\Delta P d^4 \pi t}{128 \eta L} \tag{A.1}$$

Where:

V is the volume; L

ΔP is the applied relative pressure; Pa

d is the inner diameter; m

t is the time of deposition; s

η is the viscosity of the fluid, that was assumed to be the same as water; Pa·s

L is the total length of the capillary; m

B. List of Digested Peptides from OPN

Tables B.1 and B.2 show the list of trypsin digested (TD) and chymotrypsin digested (CD) peptides from osteopontin (OPN), respectively. The lists include OPN peptides that could be originated by a "miscleavage" (MC), that is a site where trypsin was supposed to cleave, while it did not. Since miscleavages happen quite often mainly due to steric hindrance, it is important to consider them when the peaks in spectra are analysed.

Table B.1: List of osteopontin peptides from trypsin digestion. Here are reported peptides that have no miscleavages, having a "MC" value equal to 0 - or that only have one, with a "MC" value equal to 1. The underlined peptide is the one that is supposingly glycosylated. Only the glycosylated peptide with 0 miscleavages was underlined. Only peptides with a mass bigger than 500 are shown.

mass	**position**	**MC**	**peptide sequence**
10126.0865	62-152	1	SNESHDHMDDMDDEDDDDHV DSQDSIDSNDSDDVDDTDDS HQSDESHHSDESDELVTDFP TDLPATEVFTPVVPTVDTYD GRGDSVVYGLR
9963.0119	55-143	1	QETLPSKSNESHDHMDDMDD EDDDDHVDSQDSIDSNDSDD VD-DTDDSHQSDESHHSDESD ELVTDFPTDLPATEVFTPVV PTVD-TYDGR
9179.5993	62-143	0	SNESHDHMDDMDDEDDDDHV DSQDSIDSNDSDDVDDTDDS HQSDESHHSDESDELVTDFP TDLPATEVFTPVVPTVDTYD GR
5060.3126	160-204	1	RPDIQYPDATDEDITSHMES EELNGAYKAIPVAQDLNAPS DWDSR
3887.8722	20-54	1	YPDAVATWLNPDPSQKQNLL APQNAVSSEETNDFK
3527.6019	158-187	1	FRRPDIQYPDATDEDITSHM ESEELNGAYK
3224.4324	160-187	0	RPDIQYPDATDEDITSHMES EELNGAYK
2888.4272	36-61	1	QNLLAPQNAVSSEETNDFKQ ETLPSK
2662.2466	253-274	1	VSREFHSHEFHSHEDMLVVD PK
2549.1135	207-228	1	DSYETSQLDDQSAETHSHKQ SR
2535.1721	256-276	1	EFHSHEFHSHEDMLVVDPKS K
2459.1280	234-255	1	ANDESNEHSDVIDSQELSKV SR
2448.2194	15-35	1	QLYNKYPDAVATWLNPDPSQ K
2363.0382	205-225	1	GKDSYETSQLDDQSAETHSH K
2320.0451	256-274	0	EFHSHEFHSHEDMLVVDPK
2245.0214	233-252	1	KANDESNEHSDVIDSQELSK
2177.9217	207-225	0	DSYETSQLDDQSAETHSHK
2116.9265	234-252	0	ANDESNEHSDVIDSQELSK
2105.0145	36-54	0	QNLLAPQNAVSSEETNDFK
2040.0144	188-206	1	AIPVAQDLNAPSDWDSRGK
1854.8980	188-204	0	AIPVAQDLNAPSDWDSR
1801.8755	20-35	0	YPDAVATWLNPDPSQK
1690.8031	284-298	1	FRISHELDSASSEVN
1683.7820	5-19	1	QADSGSSEEKQLYNK
1474.7383	1-14	1	IPVKQADSGSSEEK
1387.6335	286-298	0	ISHELDSASSEVN
1180.6320	144-154	1	GDSVVYGLRSK
1037.4381	5-14	0	QADSGSSEEK
965.5050	144-152	0	GDSVVYGLR
898.4628	277-283	1	EEDKHLK
802.4305	55-61	0	QETLPSK
794.4519	226-231	1	QSRLYK
735.3519	275-280	1	SKEEDK
700.4253	281-285	1	HLKFR
665.3617	15-19	0	QLYNK
579.3613	229-232	1	LYKR
520.2249	277-280	0	EEDK

Table B.2: List of osteopontin peptides from trypsin digestion. Here are reported peptides that have no miscleavages, having a "MC" value equal to 0 - or that only have one, with a "MC" value equal to 1. The underlined peptide is the one that is supposingly glycosylated. The underlined peptides are the ones that are supposingly glycosylated. Only peptides with a mass bigger than 500 are shown.

mass	position	MC	peptide sequence
6480.4765	73-130	1	DDEDDDDHVDSQDSIDSNDS DDVDDTDDSHQSDESHHSDE SDELVTDFPTDLPATEVF
5308.8310	70-116	1	DDMDDEDDDDHVDSQDSIDS NDSDDVDDTDDSHQSDESHH SDESDEL
4947.7367	73-116	0	DDEDDDDHVDSQDSIDSNDS DDVDDTDDSHQSDESHHSDE SDEL
3357.4924	40-69	1	APQNAVSSEETNDFKQETLP SKSNESHDHM
3147.5301	231-257	1	KRKANDESNEHSDVIDSQEL SKVSREF
3121.5483	152-177	1	RSKSKKFRRPDIQYPDATDE DITSHM
2847.2737	159-182	1	RRPDIQYPDATDEDITSHME SEEL
2624.3017	117-140	1	VTDFPTDLPATEVFTPVVPT VDTY
2477.1539	230-250	1	YKRKANDESNEHSDVIDSQE L
2314.0905	231-250	0	KRKANDESNEHSDVIDSQEL
2297.0752	210-229	1	ETSQLDDQSAETHSHKQSRL
2260.0298	159-177	0	RRPDIQYPDATDEDITSHM
2228.9546	54-72	1	KQETLPSKSNESHDHMDDM
2039.9920	131-149	1	TPVVPTVDTYDGRGDSVVY
1901.8736	215-230	1	DDQSAETHSHKQSRLY
1878.9443	1-17	1	IPVKQADSGSSEEKQLY
1867.8603	54-69	0	KQETLPSKSNESHDHM
1798.9697	270-284	1	VVDPKSKEEDKHLKF
1738.8103	215-229	0	DDQSAETHSHKQSRL
1715.8810	1-16	0	IPVKQADSGSSEEKQL
1636.8904	269-282	1	LVVDPKSKEEDKHL
1624.8329	187-201	1	KAIPVAQDLNAPSDW
1621.7340	39-53	1	LAPQNAVSSEETNDF
1597.6877	196-209	1	NAPSDWDSRGKDSY
1551.7577	117-130	0	VTDFPTDLPATEVF
1543.7347	285-298	1	RISHELDSASSEVN
1523.8063	270-282	0	VVDPKSKEEDKHL
1508.6499	40-53	0	APQNAVSSEETNDF
1489.7182	251-262	1	SKVSREFHSHEF
1485.6816	202-214	1	DSRGKDSYETSQL
1392.5385	258-268	1	HSHEFHSHEDM
1359.7266	183-195	1	NGAYKAIPVAQDL
1327.6317	17-27	1	YNKYPDAVATW
1277.6524	18-28	1	NKYPDAVATWL
1253.6484	28-38	1	LNPDPSQKQNL
1253.6484	29-39	1	NPDPSQKQNLL
1164.5684	18-27	0	NKYPDAVATW
1140.5644	29-38	0	NPDPSQKQNL
1137.5534	141-151	1	DGRGDSVVYGL
1091.5619	131-140	0	TPVVPTVDTY
1050.6418	150-158	1	GLRSKSKKF
1029.5840	283-290	1	KFRISHEL
1011.4265	178-186	1	ESEELNGAY
967.4479	141-149	0	DGRGDSVVY
954.5618	187-195	0	KAIPVAQDL
927.4166	202-209	0	DSRGKDSY
880.5363	152-158	0	RSKSKKF
868.3618	263-269	1	HSHEDML
852.4574	251-257	0	SKVSREF
808.3319	291-298	0	DSASSEVN
755.2777	263-268	0	HSHEDM
754.4206	285-290	0	RISHEL
689.2889	196-201	0	NAPSDW
656.2787	258-262	0	HSHEF
606.2617	178-182	0	ESEEL
577.2828	210-214	0	ETSQL