
PROTEOMICS AND BIOTECHNOLOGIES: NEW METHODS FOR GLYCOPROTEOME ANALYSIS

Chiara Giangrande

Dottorato in Scienze Biotecnologiche – XXIV ciclo
Indirizzo Biotecnologie Industriale e molecolare
Università di Napoli Federico II



Dottorato in Scienze Biotecnologiche – XXIV ciclo
Indirizzo Biotecnologie Industriale e Molecolare
Università di Napoli Federico II



PROTEOMICS AND BIOTECHNOLOGIES: NEW METHODS FOR GLYCOPROTEOME ANALYSIS

Chiara Giangrande

Dottoranda: Chiara Giangrande
Relatore: Prof. Gennaro Marino
Coordinatore: Prof. Giovanni Sannia

*"Perché non esistono scorciatoie a nulla:
non certo alla salute, non alla felicità o alla saggezza.
Niente di tutto questo può essere istantaneo.
Ognuno deve cercare a modo suo,
ognuno deve fare il proprio cammino,
perché uno stesso posto può significare cose diverse
a seconda di chi lo visita"*

Tiziano Terzani

INDEX

SUMMARY	1
RIASSUNTO	3
I. Introduction	
I.1 Glycobiology	10
I.1.1 Proteins glycosylation	10
I.1.2 Biological functions of glycoproteins	11
I.2 Glycoproteins and biotechnologies	14
I.2.1 Glycans and glycoproteins for diagnostics and therapeutics	14
I.2.2 Glycans and glycoproteins for food formulations	17
I.3 Biochemical approaches to the study of protein glycosylation in the systems biotechnology era	18
I.3.1 Sample enrichment techniques	19
I.3.2 The role of mass spectrometry in glycomics	21
I.4 Aim of the PhD thesis	25
I.5 References	26
II. Oligosaccharides characterization by dansyl labelling and integrated mass spectrometry techniques	
II.1 Introduction	33
II.2 Materials and methods	34
II.3 Results and discussion	36
II.4 Conclusions	49
II.5 References	49
III. N-glycoproteome in the study of myocarditis	
III.1 Introduction	52
III.2 Materials and methods	53
III.3 Results and discussion	57
III.4 Conclusions	76
III.5 References	77
IV. Arabidopsis thaliana glycoproteome characterization by integrated mass spectrometry techniques	
IV.1 Introduction	79
IV.2 Materials and methods	80
IV.3 Results and discussion	81
IV.4 Conclusions	89
IV.5 References	90
V. Cell surface glycoconjugates: innate immunity probed by lipopolysaccharides affinity strategy and proteomics	

V.1 Introduction	92
V.2 Materials and methods	94
V.3 Results and discussion	96
V.4 Conclusions	100
V.5 References	100
Publications, communications and visiting appointments	104

Summary

Biological systems are made up of a plethora of organic components. Since nowadays attention has been focused on two important classes of bioinformative molecules, nucleic acids and proteins, whose large scale study has led to the rise of the so called “omics” sciences, genomics and proteomics. In particular the study of the proteome implies not only the protein complement of a given cell, but even the study on a high throughput scale of proteins post translational modifications, interactions, and functions. More than 50% of mammalian proteins are glycosylated and this observation has led to the conclusion that sugars attachment broadens variability among gene products. Glycans generally cover cellular surfaces, ranging from viruses to the most complex multicellular organisms and they can be considered as a molecular code that dictates to cells how to communicate with each other. The wide range of important biological processes mediated by carbohydrates has given origin to glycomics, the large scale study of the whole set of glycans of an organism. This PhD thesis targeted the development of methodological platforms for the study of glycoproteins and glycoconjugates by the integration of affinity chromatography strategies together with high performance liquid chromatography and mass spectrometry. My first steps in the study of glycosylation were focused on the development of enrichment, derivatization, and mass spectrometry procedures, that were applied to the study of the protein content of egg. Peptides of egg glycoproteins, bearing N-glycosylation sites, were captured by Concanavalin A affinity chromatography and detected by LC-MS/MS after deglycosylation. Oligosaccharides were analyzed by MALDI-MS/MS before and after dansylhydrazine derivatization of the reducing end. This derivatization was introduced to enhance oligosaccharides fragmentation characteristics. Attempts to achieve more structural information on glycans were carried out using multi-stage mass spectrometry on MALDI-LTQ-Orbitrap.

A proteomic approach was employed to investigate the molecular bases of myocarditis, a group of diseases that have in common the inflammation of the heart. In this case, the study of glycosylation anomalies in the sera sample of myocarditis-affected patients came after a wider study including both the analysis of free peptides and a general overview on protein content by 2D gel electrophoresis. This study represents a starting point for further studies aiming at the screening of proteic biomarkers for this pathological status and involving different fields of clinical investigations, opening up new opportunities for therapeutics and early diagnostics.

Glycoproteins-based therapeutics are generally produced in mammalian host systems, but plants can be considered a valuable alternative for their ability to produce homogeneously glycosylated recombinant proteins. The possibility of exploiting plants as host systems and the lack of consistent information concerning glycosylation in the plant model system *Arabidopsis thaliana* encouraged me to investigate its glycoproteome, using the strategies previously developed.

The last part of my work was devoted to the study of innate immunity mechanisms triggered by lipopolysaccharides, bacterial glycoconjugates located on cell surfaces. The full exploitation of glycoconjugates potential for industrial biotechnological and pharmaceutical applications requires a deep knowledge of the immunological mechanisms at the base of host-pathogen recognition. The goal of the project was the development of a strategy to capture LPS-interacting proteins and glycoproteins in human serum, considering that some of these mechanisms are still unknown in non-human biological systems.

Riassunto.

I glicani possono essere annoverati tra le componenti organiche più abbondanti sulla Terra e svolgono funzioni centrali in tutti gli organismi viventi, dagli organismi più semplici agli eucarioti più complessi, i quali impiegano in media l'1% dei loro genomi in enzimi coinvolti nella sintesi e modifica degli zuccheri. Questi enzimi sono molto conservati tra le specie viventi [1].

L'acquisizione progressiva di informazioni riguardanti la centralità dei carboidrati in quanto macromolecole biologiche "bioinformative" ha dato l'impulso per la nascita della "glicomica", intesa come studio sistematico e su larga scala del "glicoma", l'intero set di carboidrati prodotto da un organismo [2]. Tuttavia la mappatura del glicoma mostra un più elevato livello di complessità rispetto al genoma e al proteoma, in quanto i glicani presentano un numero maggiore di unità costitutive, rispetto agli acidi nucleici (4 nucleotidi), e alle proteine (20 amminoacidi). Il processo di mappatura del glicoma è anche complicato dal fatto che la sintesi dei carboidrati non è guidata da uno stampo ed è quindi impossibile amplificarli per poterne facilitare l'analisi. Inoltre i glicani possono adottare delle strutture complesse, altamente ramificate, con differenti conformazioni anomeriche.

I glicani possono trovarsi legati ad altre molecole a formare glicoconjugati. In tal senso la coniugazione di oligosaccaridi alle proteine aumenta considerevolmente la variabilità tra i prodotti genici. Infatti essi esistono come complesse miscele di varianti glicosidiche, dette glicoforme.

Le glicoproteine svolgono svariate funzioni biologiche, in cui la componente glicidica è assolutamente indispensabile per garantire il raggiungimento della corretta conformazione tridimensionale, per prevenire il riconoscimento di specifici aminoacidi nello scheletro polipeptidico da parte di proteasi, per coordinare processi biologici basati su fenomeni di riconoscimento cellulare. In tutti questi casi il comune denominatore è la diversità nei profili di glicosilazione delle proteine, una diversità riscontrabile tra le differenti specie, all'interno di popolazioni, tra individui della stessa popolazione e persino tra differenti tipi cellulari dello stesso organismo [3]. I carboidrati che ricoprono le superfici cellulari costituiscono un "codice molecolare" che le cellule utilizzano per comunicare tra di loro, un sistema per controllare e coordinare innumerevoli processi biologici [4].

Gli oligosaccaridi sono responsabili del riconoscimento delle cellule ospiti da parte di organismi patogeni e tossine [5,6]: è questo il caso del virus dell'influenza [7], di *Plasmodium falciparum*, l'agente eziologico della malaria [8], di *Helicobacter pylori*, la cui adesione all'epitelio della mucosa gastrica, può provocare l'ulcera e l'adenocarcinoma gastrico [9]. Il riconoscimento di epitopi oligosaccaridici da parte di lectine localizzate sulle superfici cellulari, è generalmente alla base del riconoscimento del "self/non self", nel senso che il sistema dell'immunità innata è in grado di riconoscere i glicani endogeni come "self". Molte malattie autoimmuni presentano un'alterazione in questo sistema di riconoscimento, connessa con la modifica del profilo di glicosilazione in glicolipidi e glicoproteine [10,11]. Alterazioni nel profilo di glicosilazione sono state riscontrate in molti altri fenomeni patologici, coinvolgenti processi di adesione intercellulare, come ad esempio il cancro e malattie infiammatorie.

I glicani e le glicoproteine presentano enormi potenzialità di profitto nello sviluppo di farmaci. E' chiaro che la produzione di farmaci a base di glicoproteine prevede un'accurata formulazione che deve tenere conto anche dei livelli qualitativi e quantitativi di glicosilazione. In questo contesto le industrie farmaceutiche sono alla

continua ricerca di nuovi sistemi di espressione, al fine di produrre glicoproteine con un profilo di glicosilazione adeguato. Il sistema di espressione più utilizzato è senz'altro costituito dalle cellule di mammifero, ma nuovi sviluppi nel campo della glicoproteomica non escludono l'utilizzo di sistemi di espressione differenti, che insieme alla nascente "glicoingegneria", saranno in grado in un futuro molto prossimo, di fornire prodotti più omogenei e a costi più contenuti.

Lo studio della glicosilazione trova applicazione anche in campo alimentare, in quanto un'alimentazione integrata con "cocktail" oligosaccaridici può indurre meccanismi di difesa atti a prevenire l'adesione di lectine batteriche all'epitelio della mucosa intestinale, fornendo un notevole supporto al sistema immunitario individuale [12].

La molteplicità di processi biologici che coinvolgono glicoproteine e glicconiugati ha reso necessario lo sviluppo di metodologie analitiche e computazionali che hanno dato un grande impulso allo sviluppo della glicomica. Quest'ultima si basa sull'integrazione di procedure di arricchimento, sistemi cromatografici multidimensionali, spettrometria di massa ad alta risoluzione e strumenti bioinformatici. I risultati raccolti possono costituire uno strumento indispensabile nell'approccio di "systems biology" mediante l'integrazione di dati ottenuti da differenti tecniche analitiche complementari.

In generale tutti i metodi analitici devono tenere in conto che le sequenze di glicani e glicconiugati non possono essere amplificate, diversamente dagli acidi nucleici, in quanto la loro sintesi non è diretta da uno stampo. Per questo motivo il primo passo nell'approccio allo studio dei glicomi è sempre costituito dall'arricchimento, che può essere effettuato utilizzando diversi supporti, dalle lectine alle resine al boronato, alla cromatografia per esclusione molecolare.

La finalità del presente progetto di dottorato è stata la messa a punto di piattaforme metodologiche basate sull'integrazione di strategie di cromatografia di affinità, accoppiate a cromatografia liquida ad alta prestazione e a tecniche di spettrometria di massa multidimensionale per lo studio di glicani e glicoproteine e la loro validazione in sistemi biologici modello e non. In particolare l'attenzione è stata concentrata su sistemi modello, come l'uovo e *Arabidopsis thaliana*, ma studi sono stati effettuati anche per valutare eventuali anomalie nella glicosilazione nella miocardite.

Messa a punto di tecniche integrate di spettrometria di massa e marcatura di oligosaccaridi.

La caratterizzazione delle sequenze oligosaccaridiche e della loro funzione richiede procedure sperimentali sensibili e accurate in grado di ovviare all'assenza di cromofori. Per questo motivo la loro rivelazione viene facilitata dall'introduzione di molecole cromofore all'estremità riducente. A tal riguardo, derivati dell'idrazina vengono di solito impiegati per dar luogo ad idrazoni che, se necessario, possono essere ridotti dando luogo a strutture ad anello aperto [13]. La derivatizzazione influenza fortemente l'efficienza di ionizzazione dei glicani, introducendo cariche aggiuntive e fornendo un maggior carattere idrofobico [14].

I primi esperimenti volti alla caratterizzazione della N-glicosilazione sono stati rivolti alla messa a punto di metodologie di arricchimento e di marcatura in una matrice complessa, quale estratti proteici di uovo. Il tuorlo e l'albume sono stati trattati separatamente in maniera tale da poterne analizzare sia la componente glicidica, sia quella peptidica e riscontrare se le metodologie messe a punto fossero in grado di rivelare eventuali differenze. Infatti una cromatografia di arricchimento su resina

Concanavalina A è stata effettuata su entrambe le frazioni, dopo riduzione, alchilazione e idrolisi. Tutte le frazioni dell'arricchimento sono state trattate con endoglicosidasi F. L'identificazione delle proteine e la diretta localizzazione dei siti di N-glicosilazione è avvenuta mediante analisi LC-ESI-MS/MS, considerando che l'idrolisi del legame N-glicosidico comporta un incremento di massa di 0.98 Da a carico dell'asparagina, che viene così convertita in un residuo di aspartico. Questo incremento di massa, insieme alla presenza della sequenza consensus Asn-X-S/T ha determinato l'identificazione di un certo numero di siti sia di glicosilazione sia nei ritenuti che negli aspecifici della cromatografia di arricchimento, in maniera tale da poter identificare nell'aspecifico eventuali siti di N-glicosilazione recanti epitopi oligosaccaridici non riconosciuti dalla lectina Concanavalina A.

La componente glicidica è stata analizzata mediante spettrometria di massa MALDI-TOF/TOF, una tecnica particolarmente indicata per l'analisi dei glicani, in quanto richiede basse quantità di materiale e tollera relativamente la presenza di sali. Gli spettri relativi al profilo degli oligosaccaridi, sia nel ritenuto che nel rifiutato, hanno mostrato la stessa tipologia di glicani sia nell'albume, che nel tuorlo, con l'unica differenza che in quest'ultimo, le stesse strutture sono sialilate. Le strutture sono state determinate soprattutto grazie alla conoscenza della biosintesi delle glicoproteine, ma sono state anche confermate grazie alla frammentazione in MS/MS.

Per migliorare la frammentazione dei glicani in MS/MS è stata messa a punto una nuova metodologia di derivatizzazione dell'estremità riducente con un derivato del dansile, la dansilidrazina, soprattutto grazie alla profonda esperienza con i derivati del dansile nel gruppo di ricerca in cui ho svolto le mie attività [15,16,17]. Inoltre recenti studi hanno dimostrato che la marcatura mediante derivati del dansile migliora la ionizzazione e frammentazione dei peptidi in spettrometria di massa MALDI [18]. La procedura di derivatizzazione è stata messa a punto su un oligosaccaride sintetico, e l'analisi di spettrometria di massa MALDI-MS/MS ha evidenziato un incremento nell'intensità dei segnali corrispondenti agli ioni di tipo B e C della nomenclatura di Domon e Costello [19]. Le stesse analisi sono state effettuate mediante ESI-MS ed anche in questo caso l'analisi ha rivelato un deciso miglioramento nella qualità dello spettro di frammentazione. La metodologia è stata applicata ai campioni di oligosaccaridi estratti dal tuorlo e dall'albume dell'uovo ed è stata possibile una più semplice caratterizzazione delle sequenze oligosaccaridiche, grazie al miglioramento dell'efficienza di ionizzazione degli analiti dopo marcatura con dansile, con un più alto rapporto segnale/rumore.

Un manoscritto riportante i risultati ottenuti è stato sottomesso a Glycoconjugate Journal

- Giangrande C, Melchiorre M, Marino G, Amoresano A. Oligosaccharides characterization by dansyl labelling and integrated mass spectrometry techniques.

Caratterizzazione di oligosaccaridi mediante spettrometria di massa multi-stadio

In letteratura è riportato l'utilizzo della cromatografia multi-stadio nella caratterizzazione strutturale di oligosaccaridi, sfruttando la rapidità di analisi e la sensibilità della sorgente MALDI, associate alla capacità delle trappole ioniche quadrupolari di effettuare la frammentazione sequenziale di ioni prodotto [20,21].

Durante il mio dottorato ho svolto un periodo di permanenza presso l'École supérieure de physique et de chimie industrielles della città di Parigi, dove ho avuto la possibilità di condurre esperimenti di caratterizzazione strutturale di oligosaccaridi

mediante uno spettrometro di massa MALDI LTQ Orbitrap. Un recentissimo lavoro di Karas e collaboratori ha esplorato le potenzialità di questo strumento nell'analisi di oligosaccaridi mediante tre diverse tipologie di frammentazione, CID, PQD e HCD, mostrando vantaggi e svantaggi di ognuna di esse e concludendo che solo l'integrazione dei dati ottenuti può fornire una caratterizzazione esauriente sulla struttura degli oligosaccaridi esaminati [22]. Nel nostro caso lo spettrometro di massa MALDI LTQ Orbitrap è stato adoperato per l'analisi di glicani estratti da proteine standard N-glicosilate allo scopo di sfruttare le potenzialità di questo strumento negli esperimenti di spettrometria di massa tandem multidimensionale. Gli spettri MS/MS registrati in modalità CID hanno richiesto la messa a punto dei parametri di energia di collisione e di intensità del laser per l'ottimizzazione dello spettro. Tuttavia durante la frammentazione in CID, non è possibile intrappolare frammenti aventi un rapporto m/z inferiore al 30% del rapporto m/z del precursore. Questo motivo ci ha convinti ad esplorare le potenzialità della frammentazione multi-stadio, eseguendo esperimenti MS³ ed MS⁴. Purtroppo nonostante gli ulteriori eventi di frammentazione abbiano aumentato il numero di tagli all'interno dell'anello, i "cross-ring cleavages", altamente informativi nella delucidazione strutturale, essi non sono stati sufficienti a fornire una completa caratterizzazione strutturale, a causa di problemi di sensibilità, dovuti probabilmente all'elevato consumo di campione da parte della sorgente MALDI. Questa sorgente, diversamente dalle altre sorgenti MALDI che sono state utilizzate durante il mio percorso di studi, è dotata di un sistema automatico di riconoscimento dei cristalli. Durante l'esperimento, l'utilizzo della matrice DHB insieme alla necessità di un'alta intensità del laser, ha determinato lunghi tempi di analisi, provocando un alto consumo di campione.

- Questo progetto è stato svolto in collaborazione con i dottori Joelle Vinh e Giovanni Chiappetta presso l'École supérieure de physique et de chimie industrielles.

Il glicoproteoma nello studio della miocardite.

Il termine "miocardite" indica un gruppo di patologie caratterizzate dall'infiammazione del muscolo cardiaco, che può avere origini infettive o può essere causata da meccanismi autoimmuni o ancora può essere provocata da ipersensibilità a farmaci [23,24,25,26,27].

Durante questo lavoro di tesi si è cercato di investigare le basi molecolari di questa patologia, utilizzando un approccio proteomico per lo studio delle proteine del siero. Gli esperimenti non sono stati ristretti alla sola componente del glicoproteoma, ma hanno previsto degli studi preliminari sul peptidoma, inteso come l'insieme dei peptidi endogeni del siero, e sulla componente proteica in generale, mediante un classico approccio di elettroforesi bidimensionale. Lo studio è proseguito con l'analisi della N-glicosilazione impiegando due diverse metodologie di arricchimento.

Lo studio dei peptidi liberi per LC-MS/MS ha evidenziato differenze qualitative e semi-quantitative tra i campioni provenienti da pazienti affetti da miocardite e pazienti sani.

L'elettroforesi bidimensionale è stata effettuata successivamente alla deplezione delle proteine sieriche più abbondanti. Anche le mappe bidimensionali hanno evidenziato una maggiore espressione nel siero affetto da miocardite di proteine indubbiamente correlate con l'insorgenza di eventi infiammatori, come l'emopessina, la proteina del complemento C3 e la callicreina.

Il vero e proprio esperimento per l'analisi della N-glicosilazione è avvenuto in seguito a procedure di arricchimento su resina derivatizzata con Concanavalina A e resina al

boronato. In entrambi i casi sono state analizzate per lo studio della glicosilazione sia la componente del ritenuto della cromatografia, sia quella del rifiutato, dopo trattamento con endoglicosidasi F. I peptidi contenenti il sito di N-glicosilazione sono stati identificati per LC-MS/MS, tenendo conto del fatto che l'idrolisi del legame N-glicosidico introduce un incremento di massa di 0.98 Da nel peptide, incremento che insieme alla presenza della sequenza consenso Asn-X-Ser/Thr è diagnostico per l'identificazione del sito di N-glicosilazione. I risultati ottenuti da queste analisi hanno rivelato un profilo più o meno simile per i campioni di miocarditi rispetto a quelli sani. Tuttavia alcuni specifici peptidi sono stati identificati in una sola delle frazioni della Concanavalina A, suggerendo che differenze possono essere attribuibili al differente profilo di glicosilazione di queste glicoproteine nei due campioni (bisogna considerare che la Concanavalina A è in grado di legare strutture cosiddette "high mannose" e biantennarie). Per questo motivo è stato effettuato un profilo della componente glicidica ottenuta dall'idrolisi con endoglicosidasi F e separata dalla componente peptidica per cromatografia a fase inversa su resina C-18. Questi oligosaccaridi sono stati analizzati per spettrometria di massa MALDI-TOF e alcune specifiche strutture sono state frammentate per PSD. Se il profilo degli oligosaccaridi derivanti dal ritenuto della Concanavalina A presentano un profilo più o meno simile, quelli del rifiutato mostrano la presenza nei campioni patologici di strutture completamente assenti nei campioni da individui sani. Si può concludere che è stata riscontrata la correlazione di questa patologia con strutture peculiari, incomplete o tronche, insieme a strutture iperfucosilate.

In definitiva questo approccio proteomico e glicobiologico rappresenta la messa a punto di metodologie per lo screening di marcatori proteici nella miocardite e pone le basi per lo sviluppo di nuove opportunità diagnostiche e terapeutiche.

Questo lavoro è confluito in alcune recenti pubblicazioni:

- Carpentieri A, Giangrande C, Pucci P, Amoresano A. *Glycoproteome study in myocardial lesions serum by integrated mass spectrometry approach: preliminary insights*. Eur J Mass Spectrom (Chichester, Eng). 2010;16(1):123-49.
- Andrea Carpentieri, Chiara Giangrande, Piero Pucci and Angela Amoresano (2011). *A Proteomic Approach to Investigate Myocarditis*, Myocarditis, Daniela Cihakova (Ed.), ISBN: 978-953-307-289-0, InTech
- Amoresano A, Carpentieri A, Giangrande C, Palmese A, Chiappetta G, Marino G, Pucci P. *Technical advances in proteomics mass spectrometry: identification of post-translational modifications*. Clin Chem Lab Med. 2009;47(6):647-65.

Caratterizzazione del glicoproteoma in *Arabidopsis thaliana*

La crescente domanda di proteine ricombinanti, soprattutto nell'industria farmaceutica, ha incoraggiato la ricerca di nuovi sistemi di espressione, in maniera tale da superare le basse capacità di espressione delle cellule di mammifero. A questo proposito le piante possono costituire una valida alternativa per i loro bassi costi di produzione e per la capacità di produrre proteine di mammifero perfettamente strutturate. Molte proteine di interesse biotecnologico sono state recentemente prodotte, utilizzando piattaforme vegetali, quali piante intere, sospensioni cellulari, radici e semi. Tuttavia, nonostante le piante siano in grado di fornire prodotti abbastanza omogenei per ciò che riguarda la N-glicosilazione, quest'ultima differisce dalla glicosilazione nei mammiferi.

Per questo motivo è molto importante provvedere alla messa a punto di metodologie per la caratterizzazione di glicoproteomi di interesse vegetale, scegliendo come

oggetto di studio il glicoproteoma di *Arabidopsis thaliana*, specie vegetale il cui genoma è disponibile.

L'estratto proteico di *A.thaliana* è stato dapprima frazionato per elettroforesi su gel di poliacrilammide e il gel ottenuto è stato colorato con un reattivo specifico per le glicoproteine. Alcune bande particolarmente visibili dopo colorazione sono state escisse dal gel e sottoposte ad idrolisi *in situ* e le proteine identificate per LC-MS/MS. Le proteine identificate sono state ricercate in banca dati e ne è risultato che la maggior parte di esse era effettivamente glicosilata. Purtroppo questa procedura non è stata in grado di individuare se una singola proteina fosse effettivamente glicosilata e quale fosse il sito di N-glicosilazione. Per rispondere a queste domande è stato applicato ad *A.thaliana* il consueto protocollo di arricchimento con Concanavalina A. In questo caso però è stata utilizzata la endoglicosidasi A, in grado di effettuare l'idrolisi anche in presenza di α -1,3-fucosio. La procedura ha portato all'identificazione di una serie di siti di N-glicosilazione da differenti proteine. A queste analisi si affiancano le analisi della componente glicidica per MALDI-TOF/TOF. Gli spettri di frammentazione sono stato interpretati con l'ausilio di software bioinformatici, confermando i dati già ottenuti precedentemente.

-Questo progetto è stato svolto in collaborazione con il Prof. Edgardo Filippone e il Dott. Pasquale Chiaiese del Dipartimento di scienze del suolo e della pianta dell'Università degli studi di Napoli Federico II

È in corso la stesura di un manoscritto che riporta i risultati ottenuti.

- Giangrande C, Melchiorre M, Chiaiese P, Palomba F, Filippone E, Marino G, Amoresano A. *Arabidopsis thaliana* glycoproteome characterization by integrated mass spectrometry techniques

Glicoconiugati di superficie: i lipopolisaccaridi nello studio dell'immunità innata.

I glicani e i glicoconiugati che ricoprono le superfici cellulari svolgono un ruolo cruciale nell'interazione patogeno-cellula ospite. I batteri sintetizzano numerosi glicoconiugati, come i lipopolisaccaridi. La sintesi di lipopolisaccaridi e glicoproteine coinvolge una serie di proteine omologhe, caratterizzate da un'ampia specificità di substrato, che rende queste vie biosintetiche particolarmente interessanti nell'ingegnerizzazione di batteri per la produzione industriale di glicoconiugati, generando potenziali vaccini e farmaci [28]. Tuttavia l'utilizzo di queste molecole per scopi biotecnologici richiede una profonda conoscenza dei meccanismi immunologici alla base del riconoscimento patogeno-cellula ospite. Per quanto riguarda i lipopolisaccaridi questi meccanismi coinvolgono in primo luogo l'immunità innata, intesa come una serie di risposte non specifiche che hanno luogo in caso di infezione da parte di una minaccia esterna, e che non generano una memoria immunologica. Il mio lavoro specifico è consistito nella messa a punto di una metodologia per l'individuazione di proteine del siero che interagiscono con l'LPS. Questa procedura ha previsto la marcatura con biotina della porzione glicidica del lipopolisaccaride estratto da *Salmonella typhimurium*, e la sua successiva immobilizzazione su una resina di avidina, in maniera tale da poter essere utilizzato come esca per la cattura di putativi interattori nel siero umano. Il siero è stato preventivamente cimentato con una resina di avidina, alla quale non era stato aggiunto l'LPS. A valle del sistema di affinità è stato eseguito un gel di poliacrilammide monodimensionale su cui sono stati caricati il controllo (proteine che interagiscono specificamente con la resina derivatizzata con avidina) e il campione (contenente proteine che interagiscono con il lipopolisaccaride). Dall'identificazione per LC-MS/MS delle proteine escisse dal gel,

sono stati considerati putativi interattori soltanto le proteine esclusivamente presenti nel campione e non nel controllo.

Le proteine identificate possono essere divise in due gruppi funzionali: il primo gruppo comprendente proteine che regolano il sistema del complemento (il fattore H, C4b, C4BP e α -2-macroglobulina), il secondo gruppo comprendente proteine coinvolte nell'inibizione dell'infiammazione indotta dall'LPS (HRG e le apolipoproteine).

- Questo progetto è stato svolto in collaborazione con i Prof. Rosa Lanzetta e Antonio Molinaro del Dipartimento di Chimica Organica e Biochimica dell'Università degli studi di Napoli Federico II

Un manoscritto riportante i risultati ottenuti è in preparazione

- Giangrande C, Colarusso L, Lanzetta R, Molinaro A, Pucci P, Amoresano A. Innate immunity probed by lipopolysaccharides affinity strategy and proteomics

- [1] Kiessling LL, Splain RA. Chemical approaches to glycobiology. *Annu Rev Biochem.* 2010;79:619-53.
- [2] Hirabayashi J, Arata Y, Kasai K. Glycome project: concept, strategy and preliminary application to *Caenorhabditis elegans*. *Proteomics.* 2001 Feb;1(2):295-303.
- [3] Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology.* 1999 Aug;9(8):747-55.
- [4] Kulkarni AA, Weiss AA, Iyer SS. Glycan-based high-affinity ligands for toxins and pathogen receptors. *Med Res Rev.* 2010 Mar;30(2):327-93
- [5] Karlsson KA. Microbial recognition of target-cell glycoconjugates. *Curr Opin Struct Biol.* 1995 Oct;5(5):622-35.
- [6] Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* 1993 Apr;3(2):97-130.
- [7] Wiley, D.C., Skehel, J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* 56, 365–394 (1987)
- [8] Brown A, Higgins MK. Carbohydrate binding molecules in malaria pathology. *Curr Opin Struct Biol.* 2010 Oct;20(5):560-6
- [9] Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol.* 2007 Jun;5(6):441-52.
- [10] van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol.* 2008 Jun;9(6):593-601
- [11] Alavi A, Axford JS. Sweet and sour: the impact of sugars on disease. *Rheumatology (Oxford).* 2008 Jun;47(6):760-70
- [12] Lane JA, Mehra RK, Carrington SD, Hickey RM. The food glycome: a source of protection against pathogen colonization in the gastrointestinal tract. *Int J Food Microbiol.* 2010 Aug 15;142(1-2):1-13
- [13] Lattová E, Perreault H. Method for investigation of oligosaccharides using phenylhydrazine derivatization. *Methods Mol Biol.* 2009;534:65-77.
- [14] Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wuhrer M. Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem.* 2010 Aug;397(8):3457-81.
- [15] Amoresano A, Chiappetta G, Pucci P, D'Ischia M, Marino G. *Anal Chem.* 2007, 79, 2109-17.
- [16] Amoresano A, Monti G, Cirulli C, Marino G. *Rapid Commun Mass Spectrom.* 2006, 20, 1400-4.

- [17] Cirulli C, Marino G, Amoresano A. Membrane proteome in Escherichia coli probed by MS3 mass spectrometry: a preliminary report. *Rapid Commun Mass Spectrom.* 2007;21(14):2389-97.
- [18] Chiappetta G, Ndiaye S, Demey E, Haddad I, Marino G, Amoresano A, Vinh J. *Rapid Commun Mass Spectrom.* 2010;24, 3021-32.
- [19] Domon B, Costello CE. A Systematic Nomenclature for Carbohydrate Fragmentations in FABMS/MS of Glycoconjugates., *Glycoconjugate J.* 1988, 5, 397-409.
- [20] Creaser CS, Reynolds JC, Harvey DJ. Structural analysis of oligosaccharides by atmospheric pressure matrix-assisted laser desorption/ionisation quadrupole ion trap mass spectrometry. *Rapid Commun Mass Spectrom.* 2002;16(3):176-84.
- [21] Ojima N, Masuda K, Tanaka K, Nishimura O. Analysis of neutral oligosaccharides for structural characterization by matrix-assisted laser desorption/ionization quadrupole ion trap time-of-flight mass spectrometry. *J Mass Spectrom.* 2005 Mar;40(3):380-8.
- [22] Rohmer M, Baeumlisberger D, Stahl B, Bahr U, Karas M. Fragmentation of neutral oligosaccharides using the MALDI LTQ Orbitrap. *Int J Mass Spectrom* 305, (2-3), 15 August 2011:199-208
- [23] Rezkalla SH, Kloner RA. Influenza-related viral myocarditis. *WMJ.* 2010 Aug;109(4):209-13
- [24] Blauwet LA, Cooper LT. Myocarditis. *Prog Cardiovasc Dis.* 2010 Jan-Feb;52(4):274-88.
- [25] Cihakova D, Rose NR. Pathogenesis of myocarditis and dilated cardiomyopathy. *Adv Immunol.* 2010 8;99:95-114.
- [26] Leuschner F, Katus HA, Kaya Z. Autoimmune myocarditis: past, present and future. *J Autoimmun.* 2009 Nov-Dec;33(3-4):282-9
- [27] Kühl U, Schultheiss HP. Viral myocarditis: diagnosis, aetiology and management. *Drugs.* 2009 Jul 9;69(10):1287-302
- [28] Hug I, Feldman MF. Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology.* 2011 Feb;21(2):138-51.

I. Introduction

I.1. Glycobiology

Glycans are among the most abundant organic components on the Earth and are critical constituents of all living organisms. Approximately 1% of each genome, from eubacteria to archaea and eukaryotes, is devoted to sugar processing enzymes and these genes are highly conserved [1].

"Glycome" is a term meaning the whole set of glycans produced by individual organisms, and it can be considered the third class of bioinformative macromolecules to be elucidated as well as the genome and proteome [2]. 20th century research was essentially devoted to the study and comprehension of nucleic acids and proteins biology, while glycans represent an intriguing challenge for 21st century investigations.

Glycome mapping shows a bigger level of complexity than genomics or proteomics, as glycans have a greater number of building blocks than nucleic acids (4 nucleotides) and proteins (20 aminoacids). To make the process more complex, oligosaccharide biosynthesis is not template-driven. Furthermore glycans can adopt complex branched structures of monomeric units linked together, and these structures have different conformations. They can be covalently attached to other compounds forming glycoconjugates. Among them glycoproteins are particularly important for the huge number of biological functions they are involved in. They are ubiquitous proteins, essentially found in all living organisms, ranging from eubacteria to eukaryotes [3]. Conjugation of sugars to proteins considerably increases variation among gene products, with a total of 13 monosaccharides and 8 aminoacids forming at least 41 types of different glycosidic linkages [4]. Therefore glycoproteins exist as complex mixtures of glycosylated variants (glycoforms). The microheterogeneity is due to the incomplete nature of the enzymes involved in the glycosylation and deglycosylation reactions at specific sites and depends on the cell type and on the physiological status. Anomalous pattern of glycosylation may be caused by developmental events or pathological conditions [5].

I.1.1. Proteins glycosylation

Four major types of proteins glycosylation have been essentially discovered: N-linked, O-linked glycosylation, GPI anchors and C-glycosylation. N-glycosylation consists in the linkage at asparagines residues and needs a rather strict consensus sequence *Asn-X-Ser/Thr*, where X is any aminoacid except proline [6]. Together with this sequence, conformational factors influence oligosaccharides attachment during glycoproteins biosynthesis [7]. N-linked oligosaccharides presents a common trimannosyl-chitobiose core, deriving from a biosynthetic precursor $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$. Biosynthesis starts in the endoplasmic reticulum and its first steps have been conserved during evolution, since the same precursor is synthesized in plants, mammals and fungi. The diversity which characterizes oligosaccharides linked to Asn residues in mature glycoproteins originates by differential processing reactions in Golgi apparatus [8]. Trimming by α -mannosidases without any subsequent addition leads to "high mannose" oligosaccharides whose composition is $\text{Man}_x\text{GlcNAc}_2$. In the case of glycosyl addition after α -mannosidases action "complex type" structures originates, characterized by the presence of a variable number of antennae, whose biosynthesis is initiated by the addition of $\text{Gal}\beta 1-4\text{GlcNAc}$ residues to the core structures ($\text{Man}_3\text{GlcNAc}_2$). Also "hybrid" between these two model

structures are common. Glycans can be further modified by the addition of sialic acid, fucose and xylose [5].

O-glycosylation starts in the Golgi apparatus and consist of a single sugar addition on serine or threonine: a residue of N-acetylgalactosamine is transferred to the protein from a UDP-GalNAc donor and this reaction is catalyzed by (UDP)-GalNAc-polypeptide N-acetylgalactosaminyltransferases (pp-GalNAc-Ts). Alternatively, it is reported that some other monosaccharides can initiate O-glycosylation, such as GlcNAc and Man-linked oligosaccharides [9,10] Then the protein moves to the trans-Golgi vesicles, where the addition of galactose or N-acetylglucosamine residues elongates the glycan forming different O-glycans core structures, of which eight are reported in literature. Structural diversity is amplified by the transfer of fucose and sialic acid residues [11].

Some specific enzymes are responsible for further modification of oligosaccharides by the addition of chemical groups, such as methylation, phosphorylation and sulfation [12].

Glycosylphosphatidylinositol anchor is a specific post-translational modification that attaches the protein to the outer leaflet of the cell membrane. The protein C-terminus is linked to a highly conserved core glycan (mannose(α 1-2)mannose(α 1-6)mannose(α 1-4)glucosamine(α 1-6)myo-inositol) by a phosphoethanolamine bridge. Anchorage to cell membrane is mediated by a phospholipid tail linked to the sugar. These structures vary not only for the lipid species but even for the oligosaccharidic side chains attached to the core [13].

C-glycosylation consists in the covalent attachment of an α -mannopyranosyl residue to the C2 of the tryptophan via a C-C covalent bond in a consensus sequence of Trp-X-X-Trp or Trp-Ser/Thr-X-Cys [14].

I.1.2. Biological functions of glycoproteins

Considering that 50% of mammalian proteins are glycosylated and that glycosylation presents a high level of complexity and microheterogeneity, it is now clear that glycoproteins play a wide range of roles in biological processes.

One of the reasons for glycosylation at specific sites is the coating of protein backbone which is intended to protect it from recognition by proteases and antibodies, thus affecting its turnover and half life.

However this function does not explain one of the most important features of protein glycosylation: diversity. Diversity is discovered between species, within populations, among individuals and even among different cell types of the same organism [15].

Glycans covering cell surfaces can be considered a molecular code for cells to communicate with each other, a “glycocode” which controls several biological functions [16].

Glycans on the surface of host cells can act as specific attachment sites or receptors for pathogens and toxins, in a mechanism defined “traitorous” by Varki: oligosaccharides are responsible for the first steps of infections. Noxious agents like bacteria, viruses, parasites and plant or microbial toxins gain entry into the cell thanks to these recognition phenomena [17,18]. Generally glycans carrying sialic acid at the terminal side are the most probable pathogens targets. This is the case for influenza virus: it is well known that hemagglutinin, the viral coat glycoprotein, binds to sialylated glycans receptors present on host cells [19] but only recently conformational analyses have revealed that only strains endowed with the ability to bind specifically Neu5 α 2-6 glycans present an efficient respiratory transmission

between humans, whereas avian and swine viruses are characterized by a specificity towards Neu5 α 2-3 glycans [20]. Another example is represented by *Plasmodium falciparum*, the parasites causing malaria: even in this case, it is a pathogen protein EBA-175 which binds to Glycophorin A, a sialylated O-glycosylated protein present on the surface of erythrocytes cells. Sialic acid presence is necessary but not sufficient for recognition, because optimal binding is achieved only with the correct protein chain [21]. *Helicobacter pylori* is a Gram negative bacterium, a very common pathogen of human stomachs, whose infection in some cases leads to very dangerous consequences such as ulcer disease and gastric adenocarcinoma. This bacterium adhere to fucosylated glycoproteins and sialylated glycolipids on human epithelium. Adhesion is mediated by a number of bacterial outer membrane proteins, the most studied of which are BabA and SabA [22] and depends on the blood group antigen exposed on the mucosa epithelium. In particular BabA has been shown to bind H1 and Lewis b antigens, characteristic of the blood group O phenotype, but babA gene is subjected to a high variation which explains the wider spectrum of binding of many strains that are able to recognize antigens on cells of individuals of blood group A and B [23]. Thus it is clear that both glycan binding pathogens and their host cells have developed biological strategies to perturb this equilibrium: the first ones are interested to broaden the spectrum of potential individual host cells, the second ones to expose on the surface different epitopes to escape pathogens recognition systems. The difference consists in the fact that microbial pathogens can evolve much more rapidly thanks to their short generation times and their high mutation rates. For this reason complex multicellular organisms have developed systems for changing their glycan profiles in order to evade glycan binding pathogens, even if a complete alteration of the glycosylation patterns usually leads to severe abnormalities or death [24].

Multicellular organisms are also endowed with surface lectins, carbohydrate binding proteins able to bind specific sugar epitopes. Lectins covering mammalian cells of the innate immune system recognize microbial glycans (for example lipopolysaccharides of Gram negative bacteria) and activate an immune response against infections, as in the case of Toll-like receptors [25]. It can be inferred that lectins are involved in the mechanism of self/nonself recognition, in the sense that innate immunity system is able to recognize endogenous glycans as self. Autoimmune diseases present an alteration in self antigens recognition: some glycoproteins/glycolipids can trigger an immunogenic response as a consequence of a modification of the glycosylation pattern. In several inflammatory autoimmune diseases, such as *lupus erythematosus*, effector T cell populations bearing GalNAc and Gal- β (1-4)-GlcNAc as terminal structures are highly upregulated. There is a strong evidence that galectins treatment can play a role in chronic inflammation suppression. Galectins constitute a class of lectins containing a carbohydrate-recognition domain towards N-acetyllactosamine. In particular galectin-1 is able to induce T cell apoptosis in order to recover a balance in the cytokine profile [26]. In rheumatoid arthritis some studies have shown an aberrant glycosylation affecting both innate and adaptive immunity responses. N-glycosylation of Fc regions of IgGs seems to be particularly relevant, with healthy individuals presenting little fluctuations in Fc glycoforms. On the contrary rheumatoid arthritis-affected individuals reveal a high level of agalactosylation on biantennary structures, and in some cases a higher level of outer arm fucosylation, leading to conformational changes in both the glycan moiety and polypeptide chain. These changes have as a consequence the uptake by mannose receptors on macrophages and dendritic cells and generating epitopes and the subsequent

activation of T cells. The interaction of the epitopes with certain lectins leads to an inappropriate activation of the complement cascade, giving rise to inflammatory molecules and cellular destructive activities [27].

The central role of glycosylation in the coordination of biological processes at the interface between cells is underlined by its involvement in many aspects of reproductive biology. It is now clear that the egg-spermatozoa recognition process relies essentially on the zona pellucida, an extracellular matrix surrounding the oocyte. The use of lectins with oocytes deriving from different mammalian species has shown that different species have different lectin-binding patterns, explaining the fact that fertilization is strictly species-specific [28]. In particular human zona pellucida exposes mannose, N-acetylglucosamine and galactose, diagnostic of N-glycosylated proteins. Sialylation has the function to mask β -Gal-(1-3)-GalNAc. Some O-glycosylated sites are also present. Recent studies have revealed that glycosylation is not critical for the oocyte recognition by the spermatozoa, but is essential for the acrosome exocytosis: the acrosome is a cap-like organelle at the head of the spermatozoon, that can be stimulated by external signals, such as hormones, in order to induce its fusion with the oocyte plasma membrane during fertilization [29].

Glycoproteins are key regulators of numerous events in other stages of reproduction: communication between the embryo and the endometrium is based on a complex network of signalling molecules that mediate cell-to-cell and cell-to-extracellular matrix interactions [30]. The fetomaternal interface is heavily glycosylated in all placental species and speciation is often accompanied by subtle changes in this structure. Differences concern changes in terminal glycans, especially in fucosylation and sialylation. In hemochorial placenta mammals most species are characterized by abundant bisecting bi/triantennary complex N-glycans and all species bear NeuAc α 2-3 [31].

Structural variations in the glycome of cell surfaces affect embryogenesis and morphogenesis: the expression of different glycosyltransferases in various combination, during different stages of development or during the same step but in different region of the embryo, could explain the extreme variety of ligands for cell-cell interactions [19]. These differences consist in the loss of some glycans, or the expression of unusual ones, or in variations concerning the degree of branching, all leading to alterations in cell-cell adhesion, modulating embryonic stem cells proliferation and differentiation. In some cases anomalies in the glycosylation pattern can disrupt fertilization, following mechanisms not very well understood [32].

Therefore different cell types express characteristic glycans that in several cases are useful to identify different cellular populations. Cell cycle perturbation is often associated with changes in glycosylation pattern including both the under- and overexpression of naturally occurring glycans, as well as neoexpression of glycans normally restricted to embryonic tissues. Altered expression levels of Golgi glycosyltransferases are at the origin of these glycosylation modifications. The size and branching of N-linked glycans increase considerably in cancerous cells: in particular N-acetylglucosaminyltransferase V (GnT-V), responsible of β 1,6GlcNAc branching is upregulated during malignancy, as well as sialyltransferases and fucosyltransferases, involved in linking terminal residues [33]. Adhesion between tumour cells can depend on N-linked glycosylation of E-cadherin, in the sense that the increase in branching by GnT-V results in a reduction in cell-cell adhesion and a consequent increase of cells motility and invasiveness, characteristic of metastasis. On the contrary GnT-III, whose activity is characterized by the formation of bisecting

β 1,6GlcNAc structures, results in an enhancement of cell-cell adhesion and tumour progression suppression in highly metastatic melanoma cells [34]. Some glycan epitopes are commonly found in transformed cells, such as Lewis^a, sialyl- Lewis^a, and their isomers Lewis^x and sialyl- Lewis^x. Sialyl- Lewis^a antigen is now used as a tumour marker for pancreatic cancer, colorectal cancer and other tumors [35] and is responsible for cancer cells adhesion to E-selectin present on the surface of endothelial cells [36], whereas sialyl- Lewis^x antigen plays a central role in cancer cells adhesion to E-selectin present on the surface of vascular endothelium, participating to tumour angiogenesis and metastasis [37].

Inflammation is based on the same processes as tumour metastasis. Inflammatory diseases, such as arthritis, psoriasis, asthma and diabetes, are characterized by leukocyte homing into the affected tissues. Circulating lymphocytes need adhesion to enter peripheral lymph nodes and also leukocyte emigrating into inflamed tissues have to adhere to activated endothelial cells lining blood vessel walls. Selectins (E-, P- and L-selectin), mediating the transient initial interaction responsible of the rolling of the leukocytes along the endothelial surface, bind sialylated and fucosylated epitopes which comprise a terminal component of glycans on most leukocytes, endothelial cells in the lymph node and the endothelium of inflamed tissues [38].

I.2. Glycoproteins and biotechnology

The wide range of important biological processes where glycoproteins are involved in has placed them among the most useful molecules in the biotechnological market. However, due to their non-template driven biosynthesis, oligosaccharides cannot be readily produced and their manufacture cannot be easily controlled [39].

I.2.1. Glycans and glycoproteins for diagnostics and therapeutics

Glycans and glycoproteins present large potentials for benefits in drug development. Since the 1980s proteins have cornered the market with about 200 products, most of them represented by therapeutics with diagnostics and vaccines being a small share of the market [40].

It is clear that protein pharmaceuticals should be properly glycosylated to exhibit a therapeutic efficacy. Physiological glycosylation presents a high degree of structural variability, which sometimes constitutes a problem for the biopharmaceutical market, requiring structural homogeneous products with humanized glycosylation patterns. In this context engineered expression systems together with *in vitro* remodelling methods are usually explored by pharmaceutical industries.

The demand for new therapeutic glycoproteins by the biopharmaceutical market is leading to the search of alternative expression systems.

The preferred mammalian host system for biopharmaceuticals production is CHO cells, as they have not been implicated in adverse effects and they have a long history of regulatory approval for recombinant proteins to use as drugs. Many epoetin alpha products have been developed in CHO cells (Epogen, Procrit, Aranesp). Only a few recombinant proteins have been produced in BHK21 cells (Baby Hamster Kidney), a system generally used for the production of vaccines. An example is Kogenate, the recombinant Factor VIII by Bayer, an antihemophilic factor containing 25 potential N-glycosylation sites and many O-glycosylation sites. Nevertheless the expression of therapeutics in murine systems might lead to the synthesis of potentially immunogenic recombinant products, exposing epitopes containing the

residue Neu5Gc. Furthermore BHK and CHO cells lack some sugar-transferring, such as α 2-6 sialyltransferase, α 1-3/4 fucosyltransferase, and bisecting N-acetylglucosamine transferase). So several human expression systems are subject of study. The cell line Human Embryo Kidney 293 (HEK293) is widely used for the production of recombinant proteins for research purposes and only one product, Xigris by Eli Lilly and Company is already on the market. Instead Dynepo, epoetin delta by Shire pharmaceuticals is produced in human fibrosarcoma cell line (HT1080) [41].

Although mammalian cells are the preferred host systems for recombinant expression of glycoproteins, they present several disadvantages such as long fermentation times, elevated risks of contamination, and higher costs for fermentation media. The use of different expression systems implies the introduction of an artificial glycosylation pathway where necessary.

Yeasts and other fungi are very advantageous because their fermentation is simple and cost-effective. They are able to synthesize only high mannose structures which can contain up to 100 mannose units which are highly immunogenic. Recently a glycoengineering approach has been used for the production of recombinant proteins with human-type glycosylation in *Pichia pastoris* [42]. This implies the introduction of a new enzymatic machinery via genetic manipulations, and products result to be more homogeneous, possessing a controlled glycosylation pattern [43].

Pharmaceutical or diagnostic proteins have also been expressed in filamentous fungi due to their secretor qualities, to their capacity to grow on very inexpensive media and to their property of being GRAS ("generally regarded as safe") [44].

Insect cells might constitute another choice for the expression of recombinant glycoproteins for their good qualities in protein folding and post-translational modifications. [45]. Insect cells are able to perform high mannose and paucimannosidic glycosylation, so that they are not routinely used as pharmaceutical host systems even though efforts have been made to the engineering of insect N-glycosylation pathways for assembly of mammalian-style glycans on baculovirus-expressed glycoproteins[46].

Recombinant proteins for human therapeutics can be produced also using plants as factories. Plants high mannose type glycoproteins have structures quite similar to mammals whereas complex type plant N-glycans are characterized by the presence of α (1,3)-fucose instead of an α (1,6)-fucose, and/or a β (1,2)-xylose residues on the core. Additional α (1,4)-fucose and β (1,3)-galactose residues have been identified on the terminal N-acetyl glucosamine (Lewis a antigens). Plants N-linked glycans are recognized as IgE determinants and are responsible for some human allergies [47]. Two strategies have been developed to prevent the formation of highly immunogenic plant N-glycans on recombinant glycoproteins: recombinant glycoproteins can be retained in the ER so that the immunogenic oligosaccharide, generally achieved in the Golgi compartment, cannot be added. On the other hand, also in plants, the enzymatic machinery of the Golgi apparatus can be knocked out by antisense strategy and subsequently it can be replaced by adding new enzymes by transformation [48]. In Europe there are only a few plant-derived glycoprotein therapeutics, and most of them are still under clinical trials. Examples are Locteron by Biotech, the interferon alpha produced in *Lemna minor*, and glucocerebrosidase prCGD by Protalix, a candidate for the Gaucher disease, produced in carrot cells [49].

Planning successful protein-based therapeutics means to optimize molecular stability as well as to improve pharmacokinetics and pharmacodynamics. These characteristics can be simultaneously improved by means of glycoengineering.

In general, glycosylation has been found to protect protein products from proteolytic degradation and several products have been engineered in order to improve protein stability by preventing contact between proteins backbone and cleaving proteases, as in the case of thyroid stimulating hormone (ThyrogeonOne, Genzyme) and granulocyte colony stimulating factor (G-CSF) (Granocyte, Chugai Biopharmaceuticals). In many cases glycosylation seems to impact proteins bioactivity also influencing solubility and conformational stability against pH, thermal and chemical denaturation, as in the case of interferon beta (REBIF, Pfizer/Serono and AVONEX, Biogen). Moreover improperly glycosylated proteins are recognized by specific receptor-based mechanisms and removed from circulation. The formulation of a glycoproteins-based therapeutic should consider even the fact that the content of sialic acid (and of negatively charged monosaccharides in general) is directly connected to longer circulating lifetimes, probably decreasing the adhesive properties of the protein and increasing solubility. Planning a biopharmaceutical it should be taken into account also that the glycosylation pattern strongly affects the targeting to tissues and organs.

In the last years several glycoengineering companies have arisen with the aim of the rational design of targeted glycoprotein structures (Neose Technologies, USA, GlycoFi, USA, GlycoArt, Switzerland, Glycoform, UK) [50].

Glycoengineering approaches rely on modifications of the glycosylation pattern with the aim of influencing functional activities, namely binding to cognate ligands, resistance to proteolytic cleavages, stability and solubility.

As concerning biotechnological production of drugs, it is very difficult to obtain highly homogeneous products, that often contain subtle differences as a consequence of the fermentation procedure. These differences are particularly significant from an economical point of view, especially when the patent for the production of a biological molecule expires. Other manufacturers can produce the same product, simply termed “generic” or similar products, generally referred to as “follow-on biologics” (FOBs) in United States, and “biosimilars” in European Union [51]. Biosimilars are similar but not identical to the reference products, because their characteristics depend on the manufacturing process, which is impossible to replicate exactly. In the European Community (EC) they need a marketing authorization (MA), different from both originators and generics. The main advantage for consumers is the reduced cost of biosimilars compared to their respective innovator products, although savings are not comparable to generics, because their manufacturing procedures, the non-clinical and clinical studies and the need for a pharmacovigilance program require many investments [52].

The most significant example for this strategy is suggested by erythropoietin, for the treatment of anaemia to stimulate the production of red blood cells in the bone marrow. rHuEPO (recombinant human erythropoietin) is produced by different manufacturers and the various products are not identical to each other. Epoetin alfa was the first produced rHuEPO (Epogen®, Procrit®, Eprex®) in Chinese hamster ovary (CHO) cells. Its biosimilars have been indicated with different Greek names, and their differences consist in the different cell lines and manufacturing methods. The second approved rHuEPO was epoetin beta (Recormon®, Neorecormon®), manufactured in CHO cells but presenting some minor differences in the microheterogeneity in the attached carbohydrate. Epoetin delta used human

fibrosarcoma (HT1080) cells in its manufacture, but since December 2008 is no longer commercialized. Epoetin omega, produced in baby hamster kidney cells (BHK), showed a different glycosylation profile from epoetin alfa and beta. The novel erythropoiesis-stimulating protein darbepoetin alfa (Aranesp®) is not a biosimilar but it can be considered an innovator pharmaceutical, containing two additional N-linked carbohydrate chains at positions 30 and 88 as a result of five amino acid substitutions (Ala30/Asn, His32/Thr, Pro87/Val, Trp88/Asn, Pro90/Thr) in the protein backbone. The new sites were introduced with the aim of increasing the number of sialic acid residues from 14 to 22, thus increasing serum half life without interfering with at Epo:EpoR interaction [53].

I.2.2. Glycans and glycoproteins for food formulations

Homeopathic and traditional medicine, nutraceuticals and the study of bioactive components and probiotic bacteria are becoming matter of interest in recent years.

One of the most promising application in food industry is constituted by bacterial anti-adhesive. The principle of their activity lies on the fact that bacterial infections rely on adhesion mechanisms based on the recognition of surface carbohydrate structures. Nevertheless rational design of anti-adhesives requires a complete knowledge of human as well as pathogens' glycome. [54].

Therefore glycans play an important role even in the food industry, providing a valuable alternative to conventional antibiotics. In several cases antibiotics are responsible for the microflora disruption in the gastrointestinal tract and the consequent growth of opportunistic pathogens. Defence mechanisms can be induced by cocktails of oligosaccharides to prevent adhesion of bacterial lectins to the gastrointestinal epithelium. Oligosaccharides possessing anti-adhesive properties can be found in numerous natural sources, and among them the most studied one is milk. In particular human milk oligosaccharides have been endowed with several biological functions, acting as prebiotics, anti-inflammatory agent, and it is thought to influence brain development and to drive the growth of intestinal cells [55, 56, 57]. The principle on which human milk anti-adhesive properties are based, lies on its high content in oligosaccharides and glycoproteins, that can function as soluble receptor analogs of epithelial cells carbohydrates [58]. Commercially available infant formula do not possess oligosaccharides with such anti-adhesive properties. So, other food sources might provide this important supplement, such as bovine, goat or buffalo milk. Recent studies have emphasized therapeutic and/or prophylactic effects of pig milk in *Helicobacter pylori* infections. These effects are essentially based on the presence in porcine milk of glycoproteins bearing sialyl Lewis^b and sialyl-Lewis^x structures that are the principal epitopes responsible for *H.pylori* binding and colonization. These milk components may act as decoys simulating the microbial receptors [59]. Bioactivity against *H.pylori* infections has been shown even in butter milk, whose milk fat globule membrane contains various sialylated glycoproteins [60]. Human milk has shown therapeutic properties even towards *Campylobacter jejuni* infections. This activity is associated with fucosylated oligosaccharides and glycoproteins [61].

Beyond milk another relevant food source for anti-adhesive agents is egg yolk contains high concentrations of oligosaccharides and since 1995 Koketsu et al. reported significant inhibitory effects of sialylated oligosaccharides from egg yolk towards rotavirus [62]. Moreover sialylated glycopeptides have been shown to prevent *Salmonella enteritidis* adhesion to epithelial cells [63].

To sum up bacterial adhesion can be seen as a multifactorial event and in several cases a cocktail of oligosaccharides with anti-adhesion capabilities may be the best solution not only from an economical point of view but even for anti-adhesive properties against a range of pathogens and toxins. The great advantages in such therapies reside in the easiness of supplementation with food. Furthermore these compounds are quite safe because they do not kill pathogens and do not arrest their growth so that resistance is rarely developed and they can be used instead of antibiotics, also thanks to the absence of any toxicity or immunogenicity [64].

I.3. Biochemical approaches to the study of protein glycosylation in the systems biotechnology era

The recent advances in analytical and computational tools in proteomics research have stimulated the origin of glycomics as a new scientific challenge for the elucidation of important biological processes. Glycomics relies on the integration of sample enrichment methodologies, coupled with multidimensional chromatographic separations, high resolution mass spectrometry and bioinformatics, providing relevant resources for the scientific community. The results collected can constitute an even more important tool when combined with data coming from complementary analytical techniques in a systems biology approach, aiming at the integration of such data and the formulation of modelling systems. Together with single laboratories working on glycomics around the world, international consortia are starting to coordinate and manage the large amount of data obtained. Among them, the “Consortium for Functional Glycomics” (CFG; <http://www.functionalglycomics.org>) has arisen with the aim of investigating cell communication mediated by proteins-carbohydrates interactions. It is structured in seven different scientific cores, one of which is the “Analytical Glycotechnology core”, whose main goal has been the mass spectrometry infrastructure implementation [65].

In general a typical workflow can be designed as in Fig. 1: the first steps for structural elucidation of glycoproteins are based on enrichment procedures, in order to select glycoproteins or glycopeptides to be analysed by mass spectrometry. MS procedures are able to detect entire glycopeptides, but often glycan pools are released from the peptide component using glycosidases or chemical methods since the presence of the glycan moiety on proteins can prevent their sequence coverage in peptide mapping and can hamper the action of proteases by shielding specific residues. Both the glycan and peptidic components can be analysed and characterized by tandem mass spectrometry. Carbohydrates do not contain any chromophores, so a common practice is to derivatize them to increase UV absorption, fluorescence or mass spectrometric ionization. The data produced from these experiments can be interpreted by glycoinformatics tools and compared with data deposited in available databases.

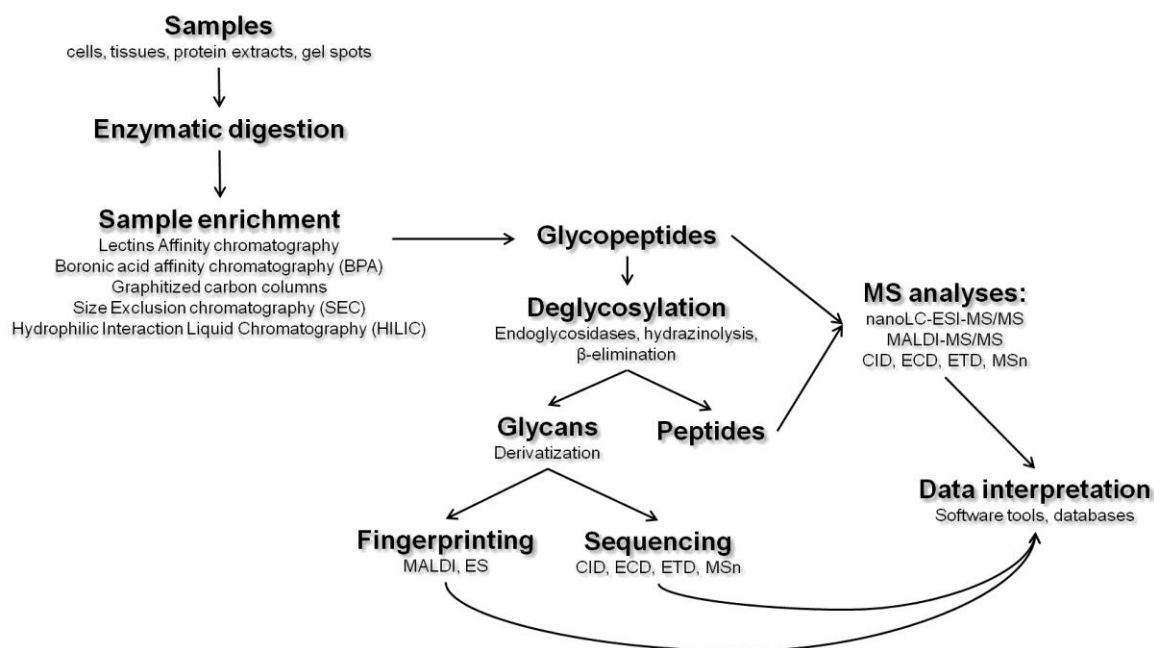


Fig. 1: a workflow for a glycoproteomic experiment. Samples deriving from different sources are treated to characterize glycosylation profiles in detail. A typical experiment can begin with the chemical or enzymatic digestion of the sample in a mixture of peptides and glycopeptides, followed by enrichment procedures in order to select glycosylated peptides. Mapping experiments are generally carried out either by online nano-LC-ES-MS/MS or by MALDI-TOF/TOF. Alternative fragmentation techniques can be used to characterize peptide backbone of glycopeptides. Glycomic analyses typically begin with the chemical or enzymatic release of specific pools of glycans. These can be derivatised and subjected to fingerprinting or sequencing. The data produced can then be analyzed and interpreted by softwares and compared to available databases.

I.3.1. Sample enrichment techniques

Understanding the structure and function of glycans and glycoproteins is the principal goal for the comprehension of central biological processes. Exploring glycans sequences and functions needs experimental procedures that could overcome the problem of the impossibility to amplify glycan structures, due to the non-template driven synthesis of these important biomolecules. In the complex mixtures of protein isoforms and variants within the cells, enrichment procedures are necessary for a successful analysis of post-translational modifications. Lectins are ubiquitous proteins of non-immune origin, specifically binding monosaccharides or oligosaccharides. They are diffuse among the living organisms, ranging from microorganisms to humans. Their carbohydrates-binding capability is widely used for the structural characterization of unknown glycoconjugates, identification of lectin-reactive components in biological materials (body fluids, cells, tissues), affinity chromatography for the fractionation and purification of glycoconjugates [66]. Lectins have been extensively exploited as an analytical tool in glycoproteomics. Lectins are able to detect subtle differences in glycosylation of proteins, even if they do not possess an “absolute” specificity. The affinity constants (K_a) for lectin-carbohydrate binding is in the low micromolar range but can reach the millimolar range. This leads to the multivalency of naturally occurring carbohydrates and glycoconjugates, resulting in an increased avidity for lectins, thus allowing their use for chromatographic separations [67].

The choice of the specific lectin to use is central in the set up of experiments for glycoproteins enrichments, because some lectins can target a broad class of glycans (for example *Concanavalin A* is able to enrich for high mannose and biantennary N-glycans) [68], or glycans with specific structures (for example L-PHA, L-phytohemagglutinin, binding $\beta(1,6)$ -branched glycans, overexpressed in malignancy) [69].

In order to extend the range of glycoproteins to be targeted, multilectin affinity chromatography was set up by Hancock and co-workers combining ConA (*Concanavalin A*), WGA (wheat germ agglutinin) and Jacalin for the analysis of serum N-glycans, sialylated glycans and O-glycans, respectively [70]. In general glycoproteins enrichment is performed by agarose-bound lectins but Novotny and co-workers developed silica-bound lectin microcolumns coupling aldehyde-modified silica with Con A or SNA in a reductive amination reaction. Their selectivity, binding capacity, reproducibility and sensitivity were tested and used for preconcentration of glycoproteins and glycopeptides from complex biological materials at trace levels. The system offered the advantage of online coupling with HPLC and MS equipments [71, 72].

In addition to this, alternatives to lectins affinity chromatography have been developed. Boronic acids are able to bind molecules carrying 1,2- or 1,3-cis-diols in a reversible esterification reaction leading to the formation of negatively charged cyclic boronate esters at alkaline pH [73]. This reaction has been exploited for glycoproteins and glycopeptides enrichments, with the development of different materials covered with boronic acid derivatives. An on-plate-selective method for MALDI mass spectrometry was developed using gold nanoparticles, allowing a rapid capture of glycosylated peptides. Elution was achieved after a washing step to remove unbound peptides and the deposition of a TFA-containing solution, capable to desorb bound peptides from the MALDI plate [74].

A frequently used alternative to lectins involves the use of activated graphitized carbon, generally employed for the separation of hydrophilic species, but often utilized for the absorption of glycans samples. Low amounts of glycoproteins, previously separated by gel electrophoresis, were digested in a double step enzymatic treatment in order to achieve small glycopeptides to purify and characterize after selection on graphite powder microcolumns in combination with mass spectrometry [75]. Recently Agilent Technologies introduced a new separation systems based on a microfluidic polymer chip for nanoflow liquid chromatography gradients [76]. Chips were packed with graphitized carbon both in the trapping column and separating one, allowing the detection of all the ribonuclease B glycoforms, undetected with C-18 stationary phase. However, larger and more hydrophobic glycopeptides remained undetected on the carbon chip, showing that only the use of both chips achieved a full comprehension of a more complex system [77].

Size exclusion chromatography (SEC) proved to be advantageous for the enrichment of glycopeptides prior to LC-MS/MS. The principle is based on the difference in size between glycosylated and non-glycosylated peptides, providing, unlike lectin, an enrichment method not dependent on the specific glycoform to be detected [78].

HILIC chromatography (Hydrophilic Interaction Liquid Chromatography) is based on zwitterionic resins, that are able to enrich for glycans and glycopeptides, by forming hydrogen bonding, ionic interactions and dipole-dipole interactions [79]. Several studies have been carried out on this methodology, which has been coupled to MALDI-MS [80] as well as to ESI-MS [81], allowing the detection of milk glycoproteins [82] and human plasma glycoproteins [83].

Another approach for glycopeptides solid phase extraction is based on the hydrazide chemistry: glycoproteins or glycopeptides are subjected to periodate oxidation in order to allow the conversion of the cis-diols to aldehydes, that are known to react with hydrazine groups forming covalent stable hydrazones. In this system, hydrazide groups are immobilized on specific beads. After oxidation, protein extracts are incubated with hydrazine-beads. The digestion with proteases allows the removal of non glycosylated peptides, whereas specific N- or O-glycosylated peptides are further released by the treatment with specific endoglycosydases. This method can be easily utilized for quantitative analyses, by stable isotope labelling of the immobilized glycopeptides [84]. Unfortunately one of the major drawbacks of this procedure is the fact that hydrazine reaction destroys the attached glycans, making them not available for further studies, even if it allows easier recovery of peptides for MS analyses and then a higher number of glycosylation sites identified [85].

1.3.2. The role of mass spectrometry in glycomics

Mass spectrometry, thanks to its high sensitivity and selectivity, is a powerful tool in glycomics and glycoproteomics, allowing characterization of glycoproteins either at the glycopeptides level or after deglycosylation, characterizing separately the peptides containing the glycosylation sites and the glycans.

Mass spectrometry analytical power lies on its capability to measure mass to charge (m/z) ratio, providing structural information with high sensitivity and accuracy.

The presence of the glycan moiety tends to increase the acidic characteristics of glycoproteins and glycopeptides, simultaneously increasing hydrophilicity and surface activity. This strongly affects ionization efficiency in the sense that unmodified peptides are more susceptible to form abundant positive ions, whereas glycopeptides are subjected to ion suppression [86]. This is the reason why glycopeptides are very difficult to detect in complex mixtures, generally requiring enrichment procedures prior to MS analyses.

In a complex mixture of glycans, more hydrophobic compounds generally suppress less hydrophobic ones, and the same behaviour is shown by basic compounds, that in positive ion mode will suppress acidic ones. Thus it is strongly recommended to equalizing glycans chemical properties with derivatization procedures or separation methodologies, in order to infuse analytes with quite similar ionization efficiencies in the mass spectrometer. Glycomics takes advantage of mass spectrometers and techniques previously developed for proteomics research.

Nowadays the vast majority of the published work on glycoconjugates relies on the matrix assisted laser desorption ionization (MALDI) and electrospray (ESI) ionization techniques. MALDI mass spectrometry is based on the irradiation of sample molecules with a laser light (generally a pulsed nitrogen laser at 337 nm) to induce sample ionisation. The sample is pre-mixed on a metal target with a highly absorbing matrix, a small aromatic molecule, that upon drying, co-crystallises with the sample. This process takes place in the ion source under a high vacuum and a strong electrical field between the target and an extraction plate. Energy deposition into the matrix molecules induces sublimation of the matrix crystals and subsequent expansion into the gas phase. In positive ion mode, protonated molecular ions $(M+H)^+$ are usually the dominant species, although they can be accompanied by salt adducts. In negative ionisation mode the deprotonated molecular ions $(M-H)^-$ are usually the most abundant species [87]. Carbohydrates generally ionize as MNa^+ species but this ionization is not as sensitive as ionization of peptides, so a higher

laser power is generally required and sometimes it is difficult to maintain a stable signal. Signal intensity even depends on the nature of the matrix and the method of sample preparation. Numerous matrices have been studied for the analysis of glycans and glycoconjugates, and a matrix very effective with a class of compounds can be ineffective with others. One of the drawbacks of MALDI-MS for oligosaccharides analysis is the high frequency of in-source decay fragmentation in the delayed extraction process with long delay times, often leading to extensive losses of acidic residues, such as sialic acid, sulphate and phosphate [88]. However MALDI-MS is rather sensitive and tolerant to contaminants, allowing rapid analyses without much effort.

Electrospray mass spectrometry (ES-MS) for biological sciences was developed by Fenn and is based on the application of a high voltage to the tip of a needle, from which the sample emerges forming a dispersed spray of highly charged droplets, with a co-axially introduced as nebulising gas. This gas is usually nitrogen, and its function is to aid the spray emerging from the capillary tip to direct towards the mass spectrometer. Solvent evaporation causes charged droplets to diminish in size, with the help of a drying. Multiply charged sample ions are extracted into the analyzer. An important feature of ESI-MS is that it confers very little extra energy to the sample (it is known as a "soft" ionization method) generally preventing in source fragmentation of the sample ions [89]. As concerning glycans and glycoconjugates analysis, electrospray tends to generate ions at different charge states and from different adducts, providing spectra very difficult to interpret. Samples should be free of salts, preventing the formation of multiple cation adduction and for this reason carbohydrates analysis by ESI-MS is generally preceded by derivatization and liquid chromatography, which can be carried out off-line or on-line.

After ionization, analytes enter the second region of the mass spectrometer, the analyser, whose main function is the separation of the ions according to their m/z values. There are four main type of mass analysers currently used for analysis of biomolecules: quadrupole (Q), ion trap (IT), time-of-flight (TOF), and Fourier transform ion cyclotron resonance (FT-ICR) analysers. They can be self-consistent or they can be combined together in tandem.

The quadrupole mass analyser is generally coupled to ESI ion source and consists of four parallel rods that have fixed DC and alternating RF potentials applied to them. The application of these voltages forces in a plane normal to the direction of ions drift allows ions oscillations which can be stable or unstable, according the m/z value. Only ions with stable trajectories will reach the detector, otherwise they will collide with the quadrupole rods.

The ion-trap mass spectrometer uses three electrodes to trap ions in a small volume. The mass analyzer consists of a ring electrode separating two end-cap electrodes. The basic principle of quadrupole ion traps is the dynamic storage of ions in a three-dimensional quadrupole device by the applied RF potentials. The RF and DC scanning can be used to eject successive m/z ratios to the detector. The compact size, the high sensitivity, the possibility to perform fragmentation experiments and the relative affordable cost of ion traps have encouraged their diffusion on the market. The main disadvantage is their relative low mass resolution due to their small volumes affecting the trapping efficiencies. These problems have been overcome thanks to the introduction of linear ion traps, characterized by a larger ion storage capacity and a higher trapping efficiency with higher scan ranges, larger electronic trap fields, increased sensitivity, higher resolution and mass accuracy [90, 91].

MALDI ion source is generally coupled to Time-Of-Flight (TOF) analyser, consisting in a tube under a high vacuum. All ions entering the TOF tube are endowed with a fixed kinetic energy, proportional to the applied voltage and the charge. Separation is based on the principle that the higher is the mass of the ion, the lower its velocity. Thus ions can be separated measuring the time each ion takes to travel through the field free region of the tube. With delayed extraction (DE), the extraction voltage is applied after a defined interval of time (expressed in ns) from the laser pulse. During this “delay time”, the extraction voltage, applied as a potential gradient, compensates for the distribution of initial kinetic energies, so that ions will not be influenced by differences in initial kinetic energies in their path. Another point for kinetic energy correction concerns the introduction of an electrostatic mirror. Thereby the flight times of ions with identical m/z values, but different kinetic energy values will be corrected when the ions arrive to the detector.

A Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer measures the cyclotron frequencies of ions oscillating around a constant magnetic field. This frequency is inversely related to the m/z ratio. The last released FT-ICR spectrometers are able of measuring frequencies with extremely high precision and a very high mass resolution, in the hundreds of thousands for instruments with magnetic field strength higher than 7 Tesla, as well as an accuracy under the ppm [92, 93]. Notwithstanding these enormous advantages, FT-ICR requires a superconducting magnet and it is very expensive.

A new type of mass analyzer, capable of resolutions comparable to that of FT-ICR, has recently entered the proteomics research: the orbitrap mass analyzer, developed by Makarov, is a non conventional ion trap, utilizing a static electrostatic field for ion trapping. A central spindle electrode operates by radially trapping ions that are injected into the Orbitrap with a velocity perpendicular to it. The electrostatic field leads ions to move in complex spiral patterns. Mass/charge ratios are obtained by measuring the frequency of harmonic ions oscillations along the axis of the electric field. Mass spectra are obtained by acquisition of time-domain image current transients and fast Fourier transforms. This equipment results in peerless characteristics in terms of resolution, mass accuracy and dynamic range [94].

Combination of mass analyzers in series can be a way to broaden the possibilities of achieving structural information by the means of tandem mass spectrometry [95], in which two analyzers are separated by a cell where the collision with an inert gas causes the fragmentation of sample ions. The first analyzer is used to select an ion to be fragmented in the collision cell. After that the second analyzer records the m/z values of the dissociation products. This experiment is called MS/MS experiment and provides structural information on the sequence of peptides or oligosaccharides. Both peptides and glycans fragmentation in CID are well known. As concerning peptides (Fig.2) they can fragment along the NH-CH, CH-CO, and CO-NH bonds, giving rise to six different fragment ions, namely a, b, and c ions (with the charge retained on the N-terminal fragment), and the x, y and z ions (with the charge on the C-terminal fragment). The most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y ions [96].

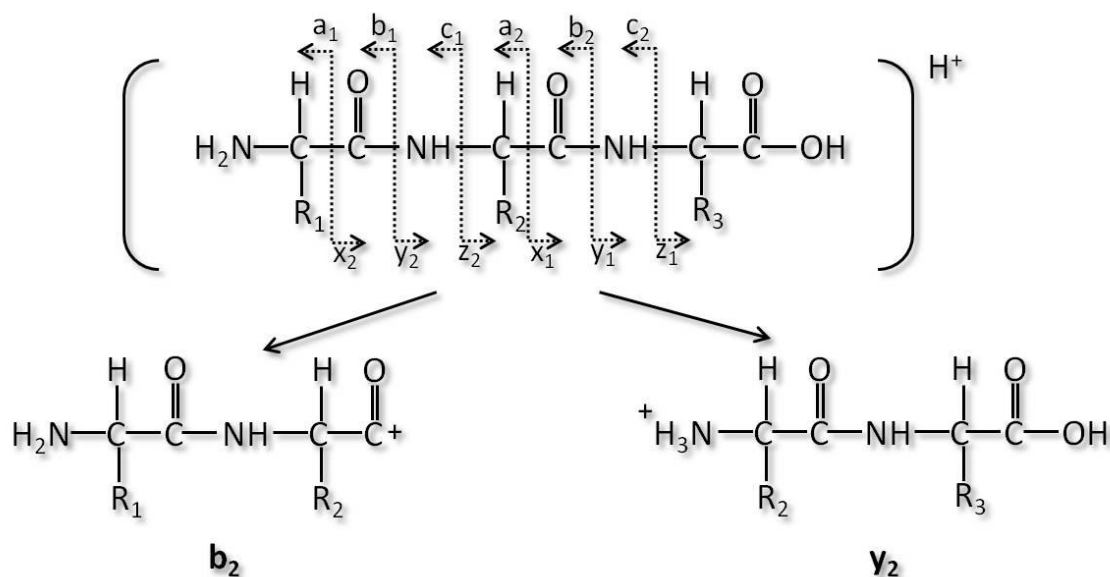


Fig. 2: Schematic representation of nomenclature for fragmentation of peptide ions

Even glycans fragmentation presents some main cleavage sites. A systematic nomenclature for labelling the fragment ions was established by Domon and Costello [97]. As shown in Fig. 3, the main cleavages, termed glycosidic cleavages, involve breaking a single bond between adjacent sugar rings. They provide much information on sequence and branching but little on linkage. Other cleavages, named cross-ring cleavages and involving two bonds cleavages are useful in the linkage determination. It is worth considering that ions retaining charge on the reducing terminus are labelled X (cross-ring), Y (C1→O-glycosidic) and Z (O→Cx glycosidic), while if the charge is retained at the non-reducing end, the produced ions are called A, B and C. The B and Y ions are usually the most abundant species. The cross-ring cleavages present additionally superscript prefixes, denoting the atoms of the two cleaved bonds. In case of branched oligosaccharides the nomenclature is based on the carbohydrate component of glycopeptides and glycoproteins, which consist of a “core” unit, and the branches, generally called “antennae”. Each antenna is identified by a Greek letter, of which the largest is given the “α” symbol; the following symbols are assigned considering the decrease in molecular weight.

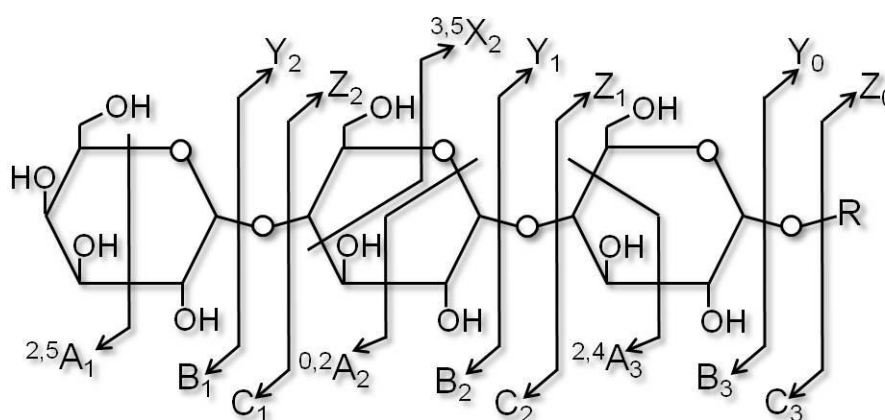


Fig. 3: Schematic representation of nomenclature for fragmentation of oligosaccharide ions

Further fragmentation events can be achieved in multi-stage mass spectrometry, allowing re-fragmentation of product ions. MSⁿ experiments can be carried out only

on ion traps, FT-ICR and Orbitrap instruments. In literature multi-stage tandem mass spectrometry experiments are reported: Harvey reports the use of MS³ and MS⁴ to achieve structural information on underivatized sugars using MALDI-QIT [98] whereas another work demonstrates the use of a hybrid MALDI-QIT-TOF mass spectrometer for the MSⁿ analysis of complex branched oligosaccharides [99].

An alternative to CID was developed by by McLafferty and co-workers [100] and is known as Electron Capture Dissociation (ECD) whose mechanism consist in the reaction of peptide cations with low energy electrons leading to peptide dissociation in FT-ICR. The interesting aspect of ECD is the possibility of maintaining intact labile modifications, allowing the determination of glycosylation sites fragmenting peptide backbone of glycopeptides [101].

Electron Transfer Dissociation (ETD) induces peptides fragmentation relying on the transfer of an electron from a radical anion to a protonated peptide. In this case the peptide backbone fragments in correspondence of the C α -N bond, creating complementary c and z-type ions instead of the typical b and y-type ions. As for ECD, ETD preserves labile modifications, but the real difference between ETD and ECD lies in the possibility to perform analyses in RF quadrupole ion traps for the first [102]. Even in this case analyses of glycopeptides have been performed [103].

The interpretation of peptides and oligosaccharides fragmentation spectra is facilitated by computer software, with the possibility to query databases with sequence and taxonomy data.

I.4. Aim of the PhD thesis

The issue of proteins correct glycosylation for the biotechnological products necessitates the development of sensitive analytical methodologies, as well as the need for sensitive and reliable diagnostic tools for glycosylation-associated diseases. This project aims at the development and setting up of new methodological platforms based on the integration of affinity chromatography strategies together with high performance liquid chromatography and mass spectrometry for the study of glycoproteins and glycoconjugates.

Chapter 2 introduces the development of a derivatization procedure to improve mass spectrometric detection and fragmentation of oligosaccharides released from egg glycoproteins. Furthermore ConA enrichment procedure to capture glycosylated proteins is described. The last part of the chapter is devoted to use of multi-stage mass spectrometry for structural characterization of oligosaccharides. This part has been carried out using MALDI LTQ Orbitrap in collaboration with Dr. Joelle Vinh and Dr. Giovanni Chiappetta of the École supérieure de physique et de chimie industrielles de la ville de Paris.

Chapter 3 describes a proteomic approach to investigate the molecular bases of myocarditis, a group of diseases characterized by the inflammation of the cardiac muscle. The study of the glycosylation profile in order to detect anomalies occurring as consequence of the onset of the pathological state is embodied in a wider study including even the analysis of free peptides and a general overview on protein content of human sera of myocarditis-affected individuals compared to healthy ones.

In **Chapter 4** the attention is focused on a plant model system, *Arabidopsis thaliana*, as plants are an intriguing alternative to mammalian host systems for the production of recombinant proteins. These protein products should be correctly homogeneous

as regards the glycosylation pattern, they should be properly glycosylated, depending on the target of the product. *A. thaliana* glycosylation is analyzed with the aim of a deep characterization of oligosaccharides structures as well as N-glycosylation sites assessment, considering that most of these sites are still signed as “potential” in proteins databases.

The work in **Chapter 5** deals with lipopolysaccharides, bacterial glycoconjugates, in order to detect LPS-interacting proteins potentially related to innate immunity response. Since some of these mechanisms are still unknown, especially in non-human biological systems, this work is aimed at the development of a “fishing for partners” strategy [104], able to capture proteins and maybe glycoproteins, involved in innate immunity reactions using an LPS bait.

I.5. References

- [1] Kiessling LL, Splain RA. Chemical approaches to glycobiology. *Annu Rev Biochem.* 2010;79:619-53.
- [2] Hirabayashi J, Arata Y, Kasai K. Glycome project: concept, strategy and preliminary application to *Caenorhabditis elegans*. *Proteomics.* 2001 Feb;1(2):295-303.
- [3] Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bond. *Glycobiology.* 2002 Apr;12(4):43R-56R.
- [4] Gabius HJ, André S, Jiménez-Barbero J, Romero A, Solís D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci.* 2011 Jun;36(6):298-313.
- [5] Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science.* 2001 Mar 23;291(5512):2351-6.
- [6] Marshall RD. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp.* 1974;(40):17-26
- [7] Bause E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probe. *Biochem J.* 1983 Feb 1;209(2):331-6
- [8] Parodi AJ. Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J.* 2000 May 15;348 Pt 1:1-13
- [9] Chai W, Yuen CT, Kogelberg H, Carruthers RA, Margolis RU, Feizi T, Lawson AM. High prevalence of 2-mono- and 2,6-di-substituted manol-terminating sequences among O-glycans released from brain glycopeptides by reductive alkaline hydrolysis. *Eur J Biochem.* 1999 Aug;263(3):879-88
- [10] Smalheiser NR, Haslam SM, Sutton-Smith M, Morris HR, Dell A. Structural analysis of sequences O-linked to mannose reveals a novel Lewis X structure in cranin (dystroglycan) purified from sheep brain. *J Biol Chem.* 1998 Sep 11;273(37):23698-703
- [11] Morelle W, Canis K, Chirat F, Faid V, Michalski JC. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics.* 2006 Jul;6(14):3993-4015

- [12] Lazar IM, Lazar AC, Cortes DF, Kabulski JL. Recent advances in the MS analysis of glycoproteins : Theoretical considerations. *Electrophoresis*. 2011 Jan;32(1):3-13.
- [13] Paulick MG, Bertozzi CR. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*. 2008 Jul 8;47(27):6991-7000.
- [14] Furmanek A, Hofsteenge J. Protein C-mannosylation: facts and questions. *Acta Biochim Pol*. 2000;47(3):781-9.
- [15] Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology*. 1999 Aug;9(8):747-55.
- [16] Kulkarni AA, Weiss AA, Iyer SS. Glycan-based high-affinity ligands for toxins and pathogen receptors. *Med Res Rev*. 2010 Mar;30(2):327-93
- [17] Karlsson KA. Microbial recognition of target-cell glycoconjugates. *Curr Opin Struct Biol*. 1995 Oct;5(5):622-35.
- [18] Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 1993 Apr;3(2):97-130.
- [19] Wiley, D.C., Skehel, J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* 56, 365–394 (1987)
- [20] Viswanathan K, Chandrasekaran A, Srinivasan A, Raman R, Sasisekharan V, Sasisekharan R. Glycans as receptors for influenza pathogenesis. *Glycoconj J*. 2010 Aug;27(6):561-70
- [21] Brown A, Higgins MK. Carbohydrate binding molecules in malaria pathology. *Curr Opin Struct Biol*. 2010 Oct;20(5):560-6
- [22] Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol*. 2007 Jun;5(6):441-52.
- [23] Aspholm-Hurtig M, Dailide G, Lahmann M, Kalia A, Ilver D, Roche N, Vikström S, Sjöström R, Lindén S, Bäckström A, Lundberg C, Arnqvist A, Mahdavi J, Nilsson UJ, Velapatiño B, Gilman RH, Gerhard M, Alarcon T, López-Brea M, Nakazawa T, Fox JG, Correa P, Dominguez-Bello MG, Perez-Perez GI, Blaser MJ, Normark S, Carlstedt I, Oscarson S, Teneberg S, Berg DE, Borén T. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science*. 2004 Jul 23;305(5683):519-22
- [24] Varki A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell*. 2006 Sep 8;126(5):841-5
- [25] Barton GM, Medzhitov R. Toll-like receptor signaling pathways. *Science*. 2003 Jun 6;300(5625):1524-5
- [26] van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol*. 2008 Jun;9(6):593-601
- [27] Alavi A, Axford JS. Sweet and sour: the impact of sugars on disease. *Rheumatology (Oxford)*. 2008 Jun;47(6):760-70
- [28] Maymon BB, Maymon R, Ben-Nun I, Ghetler Y, Shalgi R, Skutelsky E. Distribution of carbohydrates in the zona pellucida of human oocytes. *J Reprod Fertil*. 1994 Sep;102(1):81-6
- [29] Gupta SK, Bansal P, Ganguly A, Bhandari B, Chakrabarti K. Human zona pellucida glycoproteins: functional relevance during fertilization. *J Reprod Immunol*. 2009 Dec;83(1-2):50-5. Epub 2009 Oct 21
- [30] Nimbkar-Joshi S, Rosario G, Katkam RR, Manjramkar DD, Metkari SM, Puri CP, Sachdeva G. Embryo-induced alterations in the molecular phenotype of primate endometrium. *J Reprod Immunol*. 2009 Dec;83(1-2):65-71

- [31] Jones CJ, Aplin JD. Glycosylation at the fetomaternal interface: does the glycode play a critical role in implantation? *Glycoconj J.* 2009 Apr;26(3):359-66.
- [32] Amano M, Yamaguchi M, Takegawa Y, Yamashita T, Terashima M, Furukawa J, Miura Y, Shinohara Y, Iwasaki N, Minami A, Nishimura S. Threshold in stage-specific embryonic glycotypes uncovered by a full portrait of dynamic N-glycan expression during cell differentiation. *Mol Cell Proteomics.* 2010 Mar;9(3):523-37
- [33] Dube DH, Bertozzi CR. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov.* 2005 Jun;4(6):477-88
- [34] Gu J, Sato Y, Kariya Y, Isaji T, Taniguchi N, Fukuda T. A mutual regulation between cell-cell adhesion and N-glycosylation: Implication of the bisecting GlcNAc for biological functions *J Proteome Res.* 2009 Feb;8(2):431-5
- [35] Nagao K, Itoh Y, Fujita K, Fujime M. Evaluation of urinary CA19-9 levels in bladder cancer patients classified according to the combinations of Lewis and Secretor blood group genotypes. *Int J Urol.* 2007 Sep;14(9):795-9
- [36] Ugorski M, Laskowska A. Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim Pol.* 2002;49(2):303-11.
- [37] Yusa A, Miyazaki K, Kimura N, Izawa M, Kannagi R. Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. *Cancer Res.* 2010 May 15;70(10):4064-73
- [38] Kansas GS. Selectins and their ligands: current concepts and controversies. *Blood.* 1996 Nov 1;88(9):3259-87
- [39] Sinclair AS, Elliot S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci.* 2005 Aug;94(8):1626-35.
- [40] Walsh G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol.* 2010 Sep;28(9):917-24
- [41] Durocher Y, Butler M. Expression systems for therapeutic glycoprotein production. *Curr Opin Biotechnol.* 2009 Dec;20(6):700-7
- [42] Bobrowicz P. et al. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology.* 2004 Sep;14(9):757-66. Epub 2004 Jun 9.
- [43] Hamilton SR, Gerngross TU. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol.* 2007 Oct;18(5):387-92
- [44] Maras M. et al. Filamentous fungi as production organisms for glycoproteins of bio-medical interest. *Glycoconj J.* 1999 Feb;16(2):99-107.
- [45] Altmann F. et al. Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconj J.* 1999 Feb;16(2):109-23.
- [46] Harrison RL, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Adv Virus Res.* 2006;68:159-91
- [47] Gomord V, Chamberlain P, Jefferis R, Faye L. Biopharmaceutical production in plants: problems, solutions and opportunities. *Trends Biotechnol.* 2005 Nov;23(11):559-65
- [48] Lerouge P. et al. N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol Biol.* 1998 Sep;38(1-2):31-48.

- [49] Aviezer D, Brill-Almon E, Shaaltiel Y, Hashmueli S, Bartfeld D, Mizrachi S, Liberman Y, Freeman A, Zimran A, Galun E. A plant-derived recombinant human glucocerebrosidase enzyme--a preclinical and phase I investigation. *PLoS One*. 2009;4(3):e4792
- [50] Solá RJ, Griebenow K. Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci*. 2009 Apr;98(4):1223-45
- [51] Kiss Z, Elliott S, Jedynasty K, Tesar V, Szegedi J. Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. *Eur J Clin Pharmacol*. 2010 Apr;66(4):331-40
- [52] Minghetti P, Rocco P, Del Vecchio L, Locatelli F. Biosimilars and regulatory authorities. *Nephron Clin Pract*. 2011;117(1):c1-7
- [53] Kiss Z, Elliott S, Jedynasty K, Tesar V, Szegedi J. Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. *Eur J Clin Pharmacol*. 2010 Apr;66(4):331-40
- [54] Bavington C, Page C. Stopping bacterial adhesion: a novel approach to treating infections. *Respiration*. 2005 Jul-Aug;72(4):335-44.
- [55] Bode L., 2006. Recent advances on structure, metabolism, and function of human milk oligosaccharides. *Journal of Nutrition* 136, 2127–2130
- [56] Kunz, C., Rudloff, S., 2006. Health promoting aspects of milk oligosaccharides. *International Dairy Journal* 16, 1341–1346.
- [57] Hickey, R.M., 2009. Harnessing milk oligosaccharides for nutraceutical applications. In: Corredig, M. (Ed.), *Dairy-derived Ingredients Food and Nutraceutical Uses*. University of Guelph, Canada, pp. 308–343.
- [58] Kunz, C., Rudloff, S., Baier, W., Klein, N., Strobel, S., 2000. Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annual Review of Nutrition* 20, 699–722
- [59] Gustafsson A, Hultberg A, Sjöström R, Kacs Kovics I, Breimer ME, Borén T, Hammarström L, Holgersson J. Carbohydrate-dependent inhibition of *Helicobacter pylori* colonization using porcine milk. *Glycobiology*. 2006 Jan;16(1):1-10
- [60] Matsumoto M, Hara K, Kimata H, Benno Y, Shimamoto C. Exfoliation of *Helicobacter pylori* from gastric mucin by glycopolypeptides from buttermilk. *J Dairy Sci*. 2005 Jan;88(1):49-54
- [61] Ruiz-Palacios, G.M., Cervantes, L.E., Ramos, P., Chavez-Munguia, B., Newburg, D.S., 2003. *Campylobacter jejuni* binds intestinal H(O) antigen (Fuc α 1, 2Gal β 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *Journal of Biological Chemistry* 278, 14,112–14,120
- [62] Koketsu, M., Nitoda, T., Juneja, L.R., Mujo, K., Kashimura, N., Yamamoto, T., 1995. Sialyloligosaccharides from egg yolk as an inhibitor of rotaviral infection. *Journal of Agricultural Food Chemistry* 43, 858–861
- [63] Sugita-Konishi Y, Sakanaka S, Sasaki K, Juneja LR, Noda T, Amano F. Inhibition of bacterial adhesion and salmonella infection in BALB/c mice by sialyloligosaccharides and their derivatives from chicken egg yolk. *J Agric Food Chem*. 2002 Jun 5;50(12):3607-13
- [64] Lane JA, Mehra RK, Carrington SD, Hickey RM. The food glycome: a source of protection against pathogen colonization in the gastrointestinal tract. *Int J Food Microbiol*. 2010 Aug 15;142(1-2):1-13

- [65] Haslam SM, North SJ, Dell A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struct Biol.* 2006 Oct;16(5):584-91.
- [66] Wu AM, Lisowska E, Duk M, Yang Z. Lectins as tools in glycoconjugate research. *Glycoconj J.* 2009 Nov;26(8):899-913
- [67] Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J. Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. *Electrophoresis.* 2009 Sep;30(17):2957-66
- [68] Ogata S, Muramatsu T, Kobata A. Fractionation of glycopeptides by affinity column chromatography on concanavalin A-sepharose. *J Biochem.* 1975 Oct;78(4):687-96
- [69] Abbott KL, Aoki K, Lim JM, Porterfield M, Johnson R, O'Regan RM, Wells L, Tiemeyer M, Pierce M. Targeted glycoproteomic identification of biomarkers for human breast carcinoma. *J Proteome Res.* 2008 Apr;7(4):1470-80
- [70] Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A.* 2004 Oct 22;1053(1-2):79-88
- [71] Madera M, Mechref Y, Novotny MV. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal Chem.* 2005 Jul 1;77(13):4081-90
- [72] Madera M, Mechref Y, Klouckova I, Novotny MV. High-sensitivity profiling of glycoproteins from human blood serum through multiple-lectin affinity chromatography and liquid chromatography/tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2007 Jan 1;845(1):121-37
- [73] Preinerstorfer B, Lämmerhofer M, Lindner W. Synthesis and application of novel phenylboronate affinity materials based on organic polymer particles for selective trapping of glycoproteins. *J Sep Sci.* 2009 May;32(10):1673-85
- [74] Tang J, Liu Y, Qi D, Yao G, Deng C, Zhang X. On-plate-selective enrichment of glycopeptides using boronic acid-modified gold nanoparticles for direct MALDI-QIT-TOF MS analysis. *Proteomics.* 2009 Nov;9(22):5046-55
- [75] Larsen MR, Højrup P, Roepstorff P. Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Mol Cell Proteomics.* 2005 Feb;4(2):107-19
- [76] Brennen RA, Yin H, Killeen KP. Microfluidic gradient formation for nanoflow chip LC. *Anal Chem.* 2007 Dec 15;79(24):9302-9
- [77] Alley WR Jr, Mechref Y, Novotny MV. Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. *Rapid Commun Mass Spectrom.* 2009 Feb;23(4):495-505
- [78] Alvarez-Manilla G, Atwood J 3rd, Guo Y, Warren NL, Orlando R, Pierce M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res.* 2006 Mar;5(3):701-8
- [79] Hägglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res.* 2004 May-Jun;3(3):556-66

- [80] Yu YQ, Gilar M, Kaska J, Gebler JC. A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. *Rapid Commun Mass Spectrom.* 2005;19(16):2331-6
- [81] Takegawa Y, Ito H, Keira T, Deguchi K, Nakagawa H, Nishimura S. Profiling of N- and O-glycopeptides of erythropoietin by capillary zwitterionic type of hydrophilic interaction chromatography/electrospray ionization mass spectrometry. *J Sep Sci.* 2008 May;31(9):1585-93
- [82] Picariello G, Ferranti P, Mamone G, Roepstorff P, Addeo F. Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics.* 2008 Sep;8(18):3833-47
- [83] Calvano CD, Zambonin CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J Proteomics.* 2008 Aug 21;71(3):304-17
- [84] Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol.* 2003 Jun;21(6):660-6
- [85] Lee A, Kolarich D, Haynes PA, Jensen PH, Baker MS, Packer NH. Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *J Proteome Res.* 2009 Feb;8(2):770-81
- [86] Zaia J. Mass spectrometry and glycomics. *OMICS.* 2010 Aug;14(4):401-18.
- [87] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular mass exceeding 10000 daltons. *Anal. Chem.* (1988) 60:2299-2301.
- [88] Harvey DJ. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev.* 1999 Nov-Dec;18(6):349-450.
- [89] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989 Oct 6;246(4926):64-71.
- [90] Hager JW. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* (2002);16:512-526.
- [91] Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom.* (2002);13(6):659-69.
- [92] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev.* (1998);17(1):1-35.
- [93] Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* (2000);72:4266-74.
- [94] Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. 18. The Orbitrap: a new mass spectrometer. *J Mass Spectrom.* 2005 Apr;40(4):430-43.
- [95] McLafferty FW. Tandem mass spectrometry. *Science.* (1981);214(4518):280-7
- [96] Roepstorff P, Fohlmann J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom.* 1984 Nov;11(11):601.
- [97] Domon B, Costello CE. A Systematic Nomenclature for Carbohydrate Fragmentations in FABMS/MS of Glycoconjugates., *Glycoconjugate J.* 1988, 5, 397-409.
- [98] Creaser CS, Reynolds JC, Harvey DJ. Structural analysis of oligosaccharides by atmospheric pressure matrix-assisted laser desorption/ionisation quadrupole ion trap mass spectrometry. *Rapid Commun Mass Spectrom.* 2002;16(3):176-84.

- [99] Ojima N, Masuda K, Tanaka K, Nishimura O. Analysis of neutral oligosaccharides for structural characterization by matrix-assisted laser desorption/ionization quadrupole ion trap time-of-flight mass spectrometry. *J Mass Spectrom.* 2005 Mar;40(3):380-8.
- [100] Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem.* 2000 Feb 1;72(3):563-73.
- [101] Håkansson K, Cooper HJ, Emmett MR, Costello CE, Marshall AG, Nilsson CL. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal Chem.* 2001 Sep 15;73(18):4530-6.
- [102] Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta.* 2006 Dec;1764(12):1811-22.
- [103] Hogan JM, Pitteri SJ, Chrisman PA, McLuckey SA. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res.* 2005 Mar-Apr;4(2):628-32.
- [104] Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep.* 2005 Feb-Apr;25(1-2):45-56.

II. Oligosaccharides characterization by dansyl labelling and integrated mass spectrometry techniques

II.1. Introduction

Glycomics data represent a very important resource for the scientific community. This information includes results obtained by complementary analytical procedures to be integrated in a systems biology approach [1].

This project aims at the development and setting up of new methodological platforms based on the integration of affinity chromatography strategies together with high performance liquid chromatography and multi-stage mass spectrometry (MS^n) for the study of glycoproteins and glycoconjugates using a systems biotechnology approach. Analysis of glycoproteome can be performed by treating entire protein extracts with endoglycosidases so that oligosaccharides and peptides can be analysed separately. Experimental procedures are needed to explore glycan sequences and their function. However carbohydrates detection is not an easy task since no chromophores are available so far. In order to facilitate their detection, chromophores are generally introduced at the reducing end of oligosaccharides. Hydrazine derivatives are generally used to derivatize carbohydrates and their reaction leads up to hydrazone derivatives that, if needed, can be reduced to open ring structures [2]. Derivatized sugars can be analysed by mass spectrometry techniques in order to produce structural information. Both ESI-MS and MALDI-MS are generally used for oligosaccharides analysis. Derivatization strongly affects ionization efficiency by introducing additional charges and giving the sugar a more hydrophobic character. Moreover, labels influence glycans fragmentation behaviour and help the assignment of fragmentation spectra by tagging the reducing end [3].

A novel approach involving dansyl derivatives has been reported to selectively label specific PTMs [4, 5, 6, 7, 8, 9]. The dansyl-derivatization introduces a hydrophobic and aromatic moiety and basic secondary nitrogen into the molecule thus enhancing ionization efficiency. Furthermore the dansyl moiety shows a distinctive fragmentation pattern with product ions at m/z 170 and 234 in MS^2 and the diagnostic m/z 234 \rightarrow 170 fragmentation in the MS^3 mode [10]. Recently it has been demonstrated that peptide dansylation resulted to be a rapid and convenient method to improve the performances in MALDI-MS/MS analysis at proteomic scale [11]. Here, a new derivatisation protocol of glycans reducing ends by dansylhydrazide is presented. Once set up on synthetic oligosaccharides, the procedure has been successfully applied to the characterisation of oligosaccharides from a complex matrix such as egg yolk and white.

The above mentioned strategies are based on the release of the oligosaccharide component from the carrier glycoprotein, so that no information about glycosylation sites occupancy is provided. Nowadays most of the glycoproteomics procedures rely on the selective capture of glycosylated peptides by lectin affinity chromatography. Here the lectin *Concanavalin A* was used to simplify the identification of N-glycosylation sites in the mixtures deriving from egg yolk and white.

Structural characterization of glycans can be enhanced by MS^n analyses of $[M+Na]^+$ ions using instruments coupling the high sensitivity of MALDI-MS and its higher throughput and requirement for small amounts of sample. Moreover it is more tolerant to salts contaminants that in most cases co-elute with oligosaccharides in purification procedures. Here the use of MALDI LTQ Orbitrap for the same purpose was discussed.

II.2. Materials and methods

Protein extraction from egg white and yolk.

Eggs were purchased at the local market. White and yolk were separately diluted in 10 mL of MilliQ water and centrifuged at 13000 rpm for 30 min, in order to separate the soluble fraction from the other components. This fraction was then subjected to ammonium sulphate precipitation. 76 g of ammonium sulphate (Sigma Aldrich) were added to 1 L of solution. After 12-16 h at 4°C the solution was centrifuged for 30 min at 13000 rpm in order to induce proteins precipitation. The excess of salts was removed using Centricon (Millipore).

Protein concentration determination.

Egg protein concentration was determined by Bradford assay method, using BSA as standard. Known amounts of BSA were diluted in 800 µL of H₂O and then mixed to 200 µL of Coomassie Brilliant Blue. 5 different BSA concentrations were determined by measuring absorbance at 595 nm. Determined absorbance data were interpolated on the calibration curve, allowing the determination of protein concentration in the different samples. Proteins were separated by SDS-PAGE, loading on a 1.5 mm, 12.5 % gel both the white and yolk samples. The gel was run at constant 25 mA for 1h.

Reduction and alkylation.

Samples were dissolved in 200 µL of denaturation buffer (guanidine 6 M, TRIS 0.3 M, EDTA 10 mM, pH 8). Reduction was carried out by using a 10:1 DTT : cysteins molar ratio. DTT was purchased from Fluka. After incubation at 37°C for 2 h, iodoacetamide (Fluka) was added to perform carboxyamidomethylation using an excess of alkylating agent of 5:1 with respect to the moles of thiolic groups. The mixture was then incubated in the dark at room temperature for 30 min. The alkylation reaction was stopped by addition of formic acid, in order to achieve an acidic pH. The excess of salts and reagents was then removed by gel filtration on PD-10 columns (GE Healthcare). Elution was performed using ammonium bicarbonate (Fluka) 10 mM buffer. The collected fractions were analyzed by a spectrophotometer Beckman DU 7500, measuring the absorbance at 220 and 280 nm. Protein containing fractions were collected and concentrated in a centrifuge Speed-Vac.

Trypsin digestion.

Digestion was carried out in AMBIC 10 mM buffer using trypsin (Sigma Aldrich) at a 50:1 trypsin : protein mass ratio (w/w). The sample was incubated at 37°C for 12-16 h.

Concanavalin A affinity chromatography.

The peptide mixtures were incubated for 2 h with 500 µL of ConA Sepharose resin (Amersham Biosciences), which had been activated and washed in a 20mM Tris-HCl buffer; 0,5M NaCl; 1mM CaCl₂ at 4°C and pH 7.4. After 2 h the ConA containing Eppendorf was centrifuged at 3000 rpm at 4°C. The unbound peptides were removed in the supernatant. The resin was then washed six times with the same buffer described before. The elution phase was based on methyl- α -D-mannopyranoside (Sigma Aldrich) 0,25 M competition with glycopeptides and was performed for 30 min at 4°C. This procedure was repeated twice in order to be sure of having eluted all the glycopeptides bound to ConA.

RP-HPLC.

Both the peptides and the glycopeptides needed a desalting step in order to remove salts and methyl- α -D-mannopyranoside.

Both the unbound and eluted peptides were desalted by HPLC on a C-18 resin. HPLC was performed on an Agilent chromatograph, at a flow rate of 0,2 ml/min and a linear gradient from 5% B (95% acetonitrile, 5% H₂O e 0,07% trifluoroacetic acid) to 95% B in 1 min, allowing the fast elution of the unfractionated peptides.

Deglycosylation.

Egg glycopeptides were treated with PNGase F in 50 mM ammonium bicarbonate. Reactions were performed for 12-16 h at 37°C. Deglycosylation was carried out also on unbound peptides, in order to release the glycans that had not been recognized by ConA.

Oligosaccharide purification.

For glycan purification, the sample was loaded on a C-18 silica based bonded phase (Sep-pak, Waters). The column was activated with 10 mL methanol and equilibrated with 10 mL isopropanol, 10 mL H₂O and 10 mL acetic acid 5%. The oligosaccharide component was eluted in 3 mL of 5% acetic acid as eluent whereas peptides were collected in 3 mL of 40% isopropanol and 5% acetic acid. Alternatively an RP-HPLC step was performed by using a C-18 column and a gradient profile as follows: 7 min at 5% B, 5 to 95% B in 2 min and 10 min at 95% B.

LC-MS/MS.

Peptides were analyzed by a HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The capillary column works at a flow of 4 μ L/min, concentrating and washing the sample in a 40 nL enrichment column. The sample was then fractionated on a C18 reverse-phase capillary column (75 μ m~43 mm in the Agilent Technologies chip) at flow rate of 300 nl/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min. Data were acquired through MassHunter software (Agilent Technologies). Proteins identification was achieved by using Mascot software (Matrixscience), with a tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and choosing cysteine carbamidomethylation as fixed modification. Variable modifications were methionine oxidation, glutamine conversion in pyro-glutamic acid, and asparagine deamidation.

Oligosaccharides derivatization.

The reaction was set up on maltoheptaose (Sigma Aldrich).

Glycans were subjected to chemical derivatization with dansyl-hydrazine (Sigma Aldrich). Liophilized sugars were dissolved in 3-6 μ L of water, to which a solution containing 200 mM dansyl-hydrazine in 90% acetonitrile (Romil) and 10% TFA (Carlo Erba) was added. The reaction was carried out in the dark at 37°C for 3 h.

MALDI-TOF/TOF analysis of glycans.

Oligosaccharides were analysed on a 4800 Plus MALDI-TOF/TOF (Applied Biosystems). Samples were mixed on MALDI plate to the matrix consisting of a solution of 20 mg/mL DHB (Fluka), whose preparation consisted in the resuspension of DHB in water and acetonitrile 10:1 (v/v). The instrument was calibrated using a mixture of standard peptides (Applied Biosystems). Spectra were register even in

reflector positive or negative mode. MS/MS spectra were performed with CID using air as collision gas. Spectra were manually interpreted and checked against GlycoWorkbench software (14).

Q-TRAP analysis of glycans.

Supplementary MS experiments were performed on a 4000 Q-Trap system (Applied Biosystems). A micro-ionspray source was used at 2 kV with liquid coupling, with a declustering potential of 50 V, using an uncoated silica tip (o.d. 150 μm , i.d. 20 μm , tip diameter 10 μm) from New Objectives. In product ion scanning the scan speed was set to 4000 Da/s, the resolution of Q1 was set to “low” and the best collision energy was calculated as 35.

MALDI-LTQ-Orbitrap analysis of glycans.

N-glycans extracted from fetuin (Sigma Aldrich) were analyzed on MALDI-LTQ-Orbitrap (Thermo Fisher scientific), using DHB as matrix. All the spectra were recorded in positive ion mode. Full scan mass spectra were carried out using a the Orbitrap mass analyzer, due to its high mass accuracy. Resolution was set to 30,000 and 3 microscans were recorded for one position. The filling stats of the linear ion trap was regulated by automatic gain control(AGC). A system for the automatic recognition of crystals, the survey crystal positioning system(survey CPS) allowed the random choice of shot. Laser intensity was generally set at 60 μJ . CID fragmentation spectra were performed using the ion trap as mass analyzer and collision energy (CE) was optimized to 75%; the activation Q value was set to 0.250 and the activation time was 30 ms.

II.3. Results and discussion

In this study we provided a new methodology for the full characterization of N-glycome in a complex sample probed by a proteomics approach. The preliminary step for N-glycosylation sites determination was the selection of specific glycoforms to simplify the initial peptidic mixture by concanavalin A affinity enrichment on egg proteins.

White and yolk fractions were separated and proteins were precipitated in ammonium sulphate as described. SDS-PAGE was carried out on 50 μg of proteins deriving from both the white and the yolk to verify the correctness of separation procedure (Fig.1).

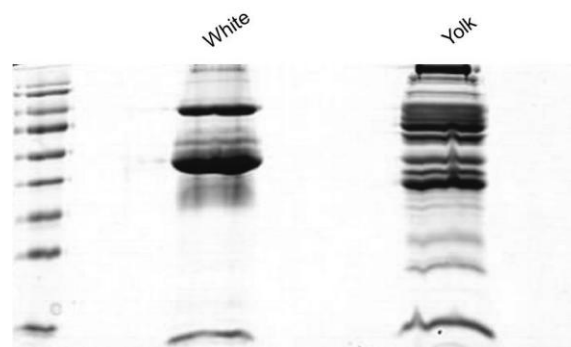


Fig. 1: SDS-PAGE of 50 μg of egg white and egg yolk

Aliquots of the two samples were reduced, carbamidomethylated and trypsin digested. Concanavalin A affinity chromatography was performed on both the digested fractions [12]. A key step in the procedure was endoglycosidase F treatment. The deglycosylation reaction leads to asparagine deamidation in the

consensus sequence. Therefore the m/z increment of 0.98 Da due to the hydrolysis of the N-glycosidic bond was used to detect N-glycosylated peptides.

Protein identification.

The identification of proteins and the direct localization of N-glycosylation sites was achieved by LC-MS/MS analysis. In ConA bound fractions, proteins identification was obtained exclusively by taking into account the peptides containing the N-glycosylation consensus sequence. As described before [12], in order to identify not glycosylated peptides and glycopeptides bearing a different glycosylation pattern (not recognized by concanavalin A) an LC-MS/MS experiment was performed on the unbound fraction.

Proteins identification was carried out by using a bioinformatics tool, namely Mascot MS/MS ion search program through the NCBI database, specifying *Gallus gallus* as taxonomy. The NCBI identified proteins were then searched against Swiss-prot database, in order to univocally recognize each protein. For the identification, we considered asparagine deamidation as a possible variable modification that took place in the consensus sequence N-X-S/T after glycosidase treatment. The data obtained are summarized in Table 1. It should be noted that several glycopeptides were detected in the unbound fraction from egg samples and identified by the diagnostic increment of 0.98 Da respect to the theoretical mass. Some proteins, as the vitellogenin-2, ovalbumin, ovotransferrin and clusterin, are indicated to be glycosylated on the basis of similarity with the counterpart from other species. For several egg proteins, the glycosylation sites are indicated in the SwissProt database as putative N-glycosylation sites.

The MS analyses performed in this work, led to the direct identification of N-glycosylation sites for several egg proteins. Some of them were exclusively expressed in one fraction, such as Cathepsin D, only present in the yolk, whereas others were expressed in both yolk and white, such as Apolipoprotein B. To the best of our knowledge, these results represent the first data on the fine localisation of N-glycosylation sites thus complementing the data already available on the characterisation of egg proteome.

Table 1: List of the proteins identified by LC-MS/MS analyses. N-glycosylation sites are marked with an asterisk after the asparagines residue. BY and BW indicate the bound fraction of egg yolk and white samples, respectively; UY and UW indicate the unbound fraction of egg yolk and white, respectively.

Protein	Sequence
Vitellogenin-1 (P87498)	MSFTCSFN [*] K ^{BY, UY, UW} EVHINTSSAN [*] ITICPAADSSLLVTCNK ^{BY}
Ovomucoid (P01005)	FPN [*] ATDK ^{BY, BW, UY, UW} PMN [*] CSSYAN [*] TTSEDGK ^{BY, BW, UY, UW} CNFCNAVVESN [*] GTLTSLSHFGK ^{BY, BW, UY, UW} LAAVSVDCSEYKPDCTAEDRPLCGSDN [*] K ^{BY, BW, UY, UW}
Vitellogenin-2 (P02845)	VQVFVTN [*] LTDSSK ^{BY} LLVDGAESPTAN [*] ISLISAGASLWIHNENQGFALAAPGHGIDK ^{BY, UY}
Apovitellenin-1 (P02659)	NFLIN [*] ETAR ^{BY, UY, UW}
Serum albumin (P19121)	KQETTPINDN [*] VSQCCSQLYANR ^{BY}
Apolipoprotein B (Q197X2)	VASPLYN [*] VTWR ^{BY, BW} SLHN [*] STLGSVLPK ^{BY, BW} NQYGMTEVN [*] QTLK ^{BY, BW} INSPSNQILIPAMGN [*] ITYDFSFK ^{BY, BW} LNIEADFSSQADLVNN [*] MTTEVEAGR ^{BY, BW} VASPFRTLSTQAEVHN [*] TTASANSPEFDTLSAQATSK ^{BY, BW}

Complement C3 (Q90633)	FEIDHALSN ^R ^{BY} AAN ^L LSDIVPNTSE ^T ETK ^{BY}
Ig mu chain C region (P01875)	VNVSGTDWR ^{BY, UY} VTHN ^G T ^S ITK ^{UW}
Cathepsin D (Q057441)	DTLQISN ^I SIK ^{BY}
Arylsulfatase K (Q5ZK90)	NDFLN ^V SAPR ^{BY}
Complement factor B-like protease (P81475)	TGN ^W STKPSCK ^{BY}
Ovotransferrin (P02789)	TAGWV ^I PMGLIHN ^R ^{BW} FGVN ^G SEK ^{BW, UY, UW}
Ovalbumin-related protein Y (P10114)	SAN ^L TGISSVDNLMISDAVHG ^V FMEVNEEGTEATG ^S TGAIGNIK ^{BW, UW}
Ovoglycoprotein (Q8JIG5)	LN ^E TCVVK ^{BW, UY, UW} HN ^S TLTHEDGQVVSMAELTHSDK ^{BW}
Clusterin (P14018)	RYDDL ^L SAFQAEMLN ^T SSLLDQLNR ^{UY, UW}
Ovalbumin (P01012)	YN ^L TSVLMAMGITDVFSSANLSG ^I SSAESLK ^{UW}

Characterization of egg oligosaccharides.

Oligosaccharides characterization is arousing a particular interest even being a challenging analytical problem. Glycan molecules present extremely various structures as concerning the type of monosaccharides they are built of, the branching, the anomericity of the glycosydic linkage and the global structure that the carbohydrate moieties confer to proteins. Nowadays mass spectrometry methodologies resolve most of the challenges connected to the study of glycans such as the glycoforms heterogeneity achieving qualitative and quantitative information on oligosaccharides structures present in the analyte, with enormous sensitivity [13]. MALDI-TOF mass spectrometry technique is able to detect any difference between the analytes in terms of speed and sensitivity in the analyses.

The bound or the unbound ConA fractions both from yolk and white were treated with PNGase F. Oligosaccharides were purified from peptides by RP-HPLC and analysed by MALDI-MS in reflectron positive and negative mode. Oligosaccharides characterization by MALDI-MS was achievable on the basis of the m/z value together with the knowledge of oligosaccharides biosynthesis. Hypothetical structures were then confirmed through tandem mass spectrometry. MS/MS spectra were interpreted using the software Glycoworkbench [14], which simulates glycans fragmentation and compares the theoretical fragments with the experimental ones.

MALDI-TOF spectra of oligosaccharides, released from protein backbone as described, were recorded in reflector mode. Peaks were manually attributed and then confirmed by MALDI-MS/MS.

In positive ion mode oligosaccharides usually ionize as MNa^+ but sometimes we noticed the presence of sodium adducts, especially when the structure contains sialic acid residues. It is quite useful to register negative spectra in order to detect oligosaccharides bearing sialic acid residues due to ionize as MH^- of sugars. Fig. 2 shows the mass spectra in positive ion mode of the ConA bound and unbound fraction from egg white. The attribution of each peak is reported in Table 2.

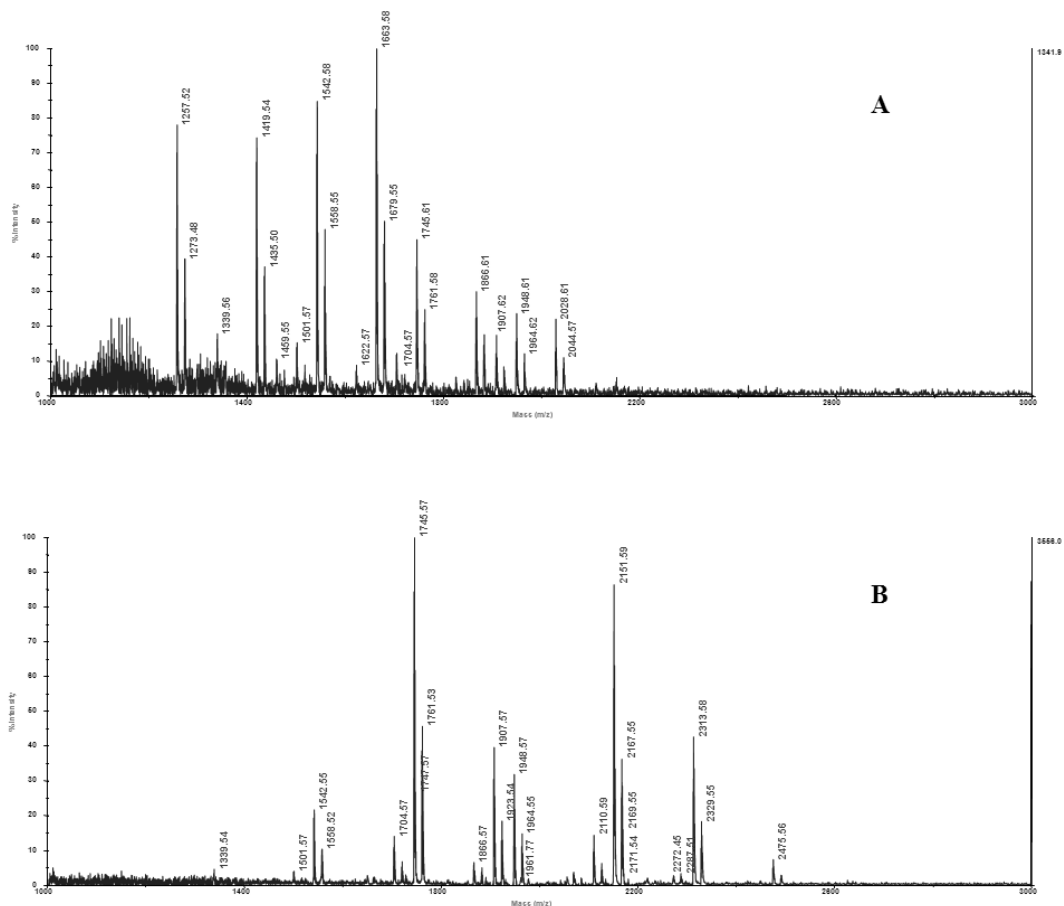











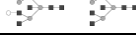











Fig. 2: positive ion mode MALDI-MS spectrum of ConA bound (panel A) and unbound fraction (panel B) oligosaccharides released from egg white glycoproteins.

Table 2: list of the oligosaccharides structures detected by MALDI-MS, where BY=ConA bound fraction of yolk, UY=ConA Unbound fraction of yolk, BW=ConA bound fraction of white, UW=ConA unbound fraction of white

m/z	Composition	Structure
1095.37 ^{BY}	HexNac ₂ Hex ₄	
1136.39 ^{BY}	HexNac ₃ Hex ₃	
1257.42 ^{BY,BW}	HexNac ₂ Hex ₅	
1298.42 ^{BY}	HexNac ₃ Hex ₄	
1339.44 ^{BY,BW,UY,UW}	HexNac ₄ Hex ₃	
1419.47 ^{BY,BW}	HexNac ₂ Hex ₆	
1460.50 ^{BY,BW,UY}	HexNac ₃ Hex ₅	
1501.53 ^{BY,BW,UY,UW}	HexNac ₄ Hex ₄	
1542.55 ^{BY,BW,UY,UW}	HexNac ₅ Hex ₃	

1581.52 ^{BY}	HexNAc ₂ Hex ₇	
1622.55 ^{BW}	HexNAc ₃ Hex ₆	
1663.58 ^{BY,BW}	HexNAc ₄ Hex ₅	
1704.61 ^{BY,BW,UY,UW}	HexNAc ₅ Hex ₄	
1745.63 ^{BY,BW,UY,UW}	HexNAc ₆ Hex ₃	
1850.66 ^{BY,UY}	FucHexNAc ₅ Hex ₄	
1866.66 ^{BW,UY,UW}	HexNAc ₅ Hex ₅	
1905.63 ^{BY,UY}	HexNAc ₂ Hex ₉	
1907.68 ^{BW,UW}	HexNAc ₆ Hex ₄	
1948.71 ^{BW,UY,UW}	HexNAc ₇ Hex ₃	
1954.67 ^{BY,UY}	HexNAc ₄ Hex ₅ Sia	
2028.71 ^{BW,UY}	HexNAc ₅ Hex ₆	
2067.69 ^{BY,UY}	HexNAc ₂ Hex ₁₀	
2110.76 ^{UY,UW}	HexNAc ₇ Hex ₄	
2151.79 ^{UY,UW}	HexNAc ₈ Hex ₃	
2245.77 ^{BY,UY}	HexNAc ₄ Hex ₅ Sia ₂	
2272.82 ^{UY,UW}	HexNAc ₇ Hex ₅	
2303.81 ^{BY,UY}	FucHexNAc ₅ Hex ₅ Sia	
2313.84 ^{UY,UW}	HexNAc ₈ Hex ₄	
2319.81 ^{BY,UY}	HexNAc ₅ Hex ₆ Sia	
2475.90 ^{UY,UW}	HexNAc ₈ Hex ₅	

The glycans detected belonged to the complex type as confirmed by MS/MS data. However, it should be noted that some peaks could be attributed to more than one structure. The MS/MS analysis let us to confirm the presence of isobaric structures or to discriminate between the different structure only for the most abundant species. As

an example, the peak occurring at m/z 1542,5 was identified as a biantennary glycan carrying a bisecting β -1,4-GlcNAc rather than as a triantennary complex structure. The heterogeneity in the peak attribution still remained for the less abundant ones. The main differences between the fractions consisted in the higher branching of unbound oligosaccharides.

Fig. 3 displays the positive ion mode mass spectra of ConA bound and unbound fractions of yolk. In both fraction the presence of sialylated structures, which were not detected in the white sample can be observed. These structures were identified in positive ion mode due to the presence of a negative charge for each sialic acid, which was neutralized by a sodium ion. For this reason we performed the analyses in negative ion mode in order to confirm the presence of sialic acid residues. A high heterogeneity was found in the MALDI spectra of the glycan moieties essentially belonging to the complex type. Moreover, differences between egg white and yolk could be appreciated and were essentially ascribed to the total absence of sialic acid in the oligosaccharides from white egg fraction.

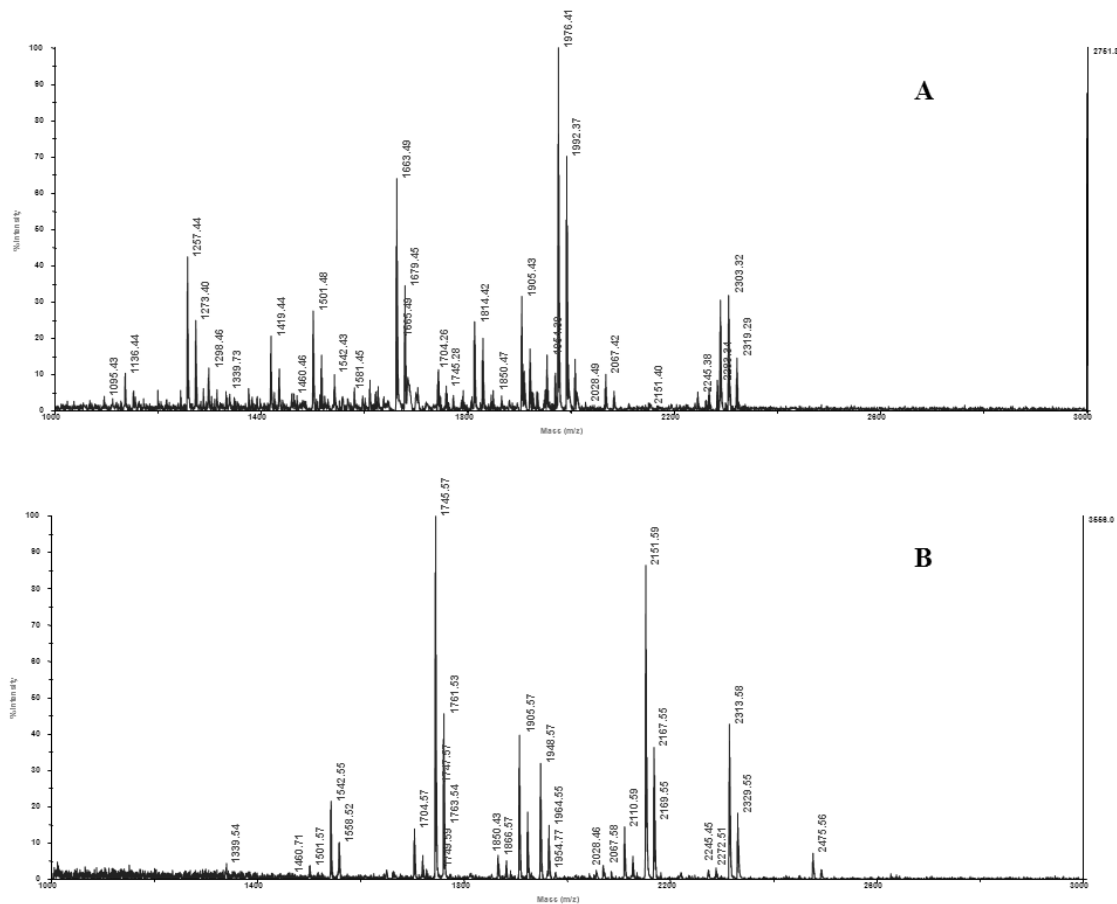


Fig. 3: positive ion mode MALDI-MS spectrum of ConA bound (panel A) and unbound fraction (panel B) oligosaccharides released from egg yolk glycoproteins.

Set up of a dansyl-labeling procedure of glycans reducing ends.

In order to improve MS/MS glycan fragmentation a novel derivatisation method was proposed. Despite the fact that the MALDI ionization mechanism is not yet completely understood, it is widely acknowledged that the ionization efficiency depends on the chemical physical properties of the analyte. In particular, the presence of basic and/or hydrophobic groups [15] enhances the yields of the ionization process. Dansyl chloride (DNS-Cl) is a fluorescent reagent which has

extensively used in biochemistry during the eighties [16] and used to selectively label specific post-translational modification in a selective proteomics approach [9, 10, 17]. In a recent paper it was demonstrated that dansylation labeling improves the MALDI ionization and fragmentation of peptides [11]; moreover this derivatisation was applied to fragmentation of glycosylated peptides whose analysis is often problematic [18]. On the basis of the experience recently made with dansyl chloride and other dansyl derivatives, we used a commercially available dansylated reagent, namely dansyl-hydrazine, to selectively label free oligosaccharides (Fig.4).

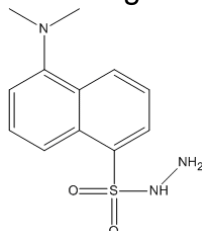


Fig. 4: Dansylhydrazine

Our experimental procedure was set up on a synthetic oligosaccharide, maltoheptaose, by using dansyl-hydrazine as label. The reaction was carried out as described in the experimental section. To monitor the effects of glycan dansylation, we diluted the unlabeled sample with a volume of ACN and we used it as control for the further experiments. The analyses were carried out by MALDI-TOF mass spectrometry where the labelled oligosaccharides showed an increment of 247.3 Da compared to the unmodified signal. Tandem mass spectrometry was performed on both the labelled and unlabelled oligosaccharide. Signals were assigned according to the nomenclature proposed by Domon and Costello [19].

Fig.5 shows the MS/MS spectra of maltoheptaose registered on a MALDI-TOF/TOF mass spectrometer where the DNS-sugar undergoes a loss of 234 Da, corresponding to the dansyl unit. We noticed the presence of Y and B ions, as well as cross-ring cleavages. As shown in the figure, the intensity strongly increased for the labelled specie. The difference in ionization efficiency could be appreciated in the difference in quality of the MS/MS data between dansylated and native oligosaccharides because the fragmentation profiles are modified. The fragmentation pattern of dansylated oligosaccharide exhibited enhanced B-ion and C-ion series in MS/MS experiments (Fig. 5B) compared to the native ones (Fig. 5A). This improvement in fragmentation could be attributed to the presence of the dansyl fluorescent moiety at the reducing end. This behavior was in agreement with previous data where dansyl moiety was added at the N-terminus of different peptides [11].

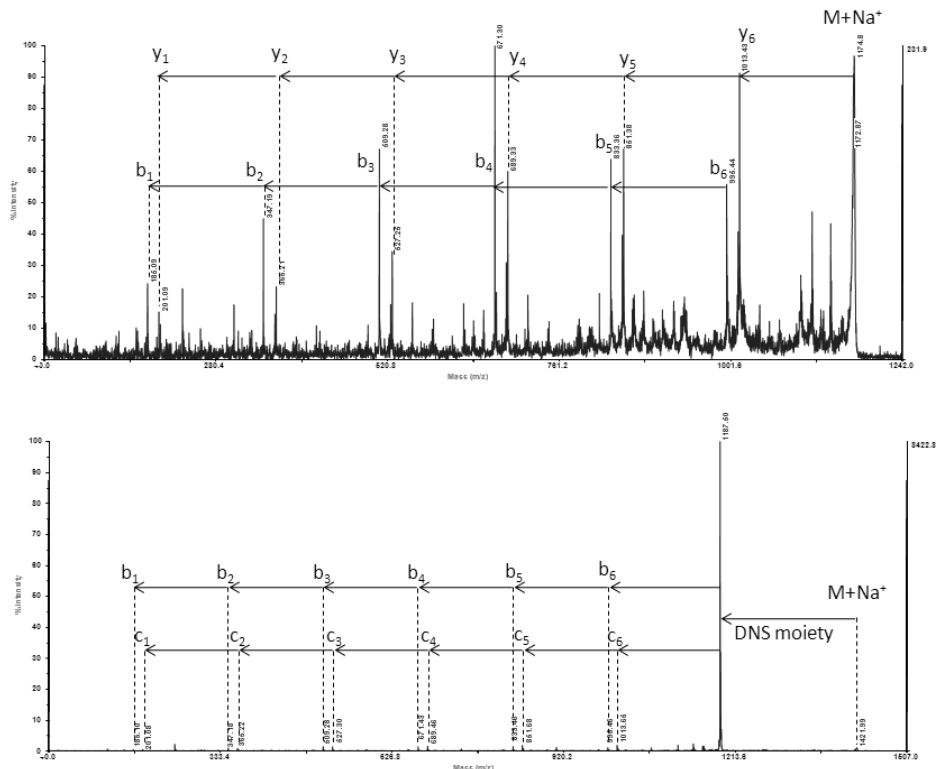


Fig. 5: positive ion mode MALDI-MS/MS spectrum of DNS-labelled maltose (B) compared to the unmodified one (A).

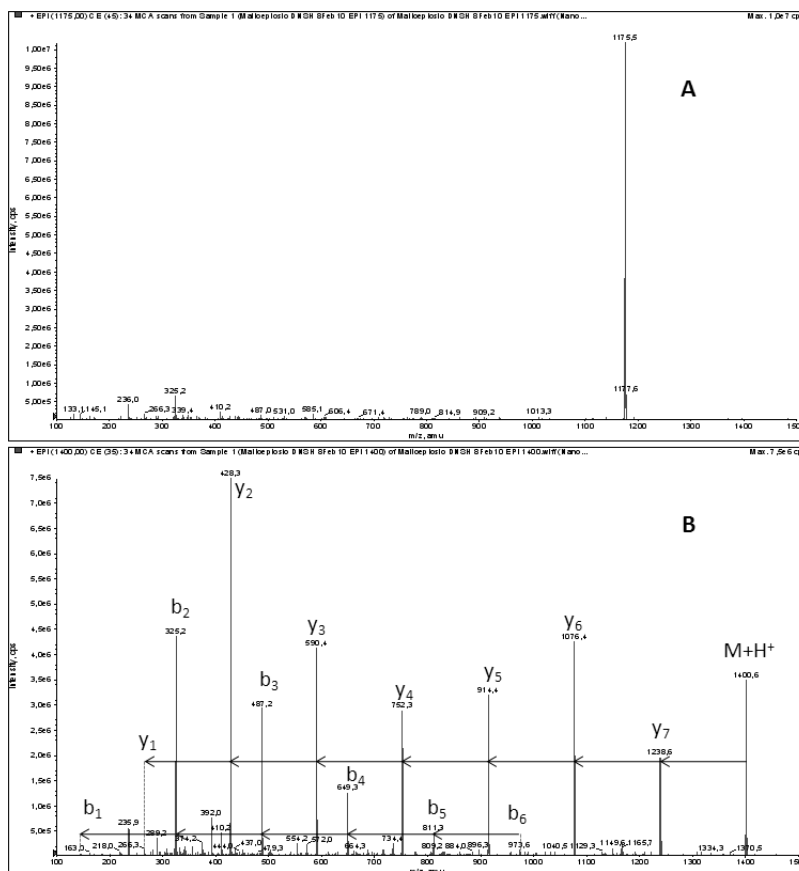


Fig. 6: ESI-MS/MS spectrum of DNS-labelled maltose (B) compared to the unmodified one (A) performed on a 4000 Q-TRAP.

The same analysis was performed by using an hybrid Q-Trap instruments and the results are reported in Fig.6. As shown, the ESI-MS/MS spectrum of the DNS-oligosaccharide reveals a better fragmentation, when compared to the unmodified one. Even in this case both Y and B ions were present, but cross-ring cleavages were less frequent than for MALDI-MS.

Once set up, the dansyl labelling procedure was applied to the characterisation of glycans in a real sample, namely the oligosaccharides released from egg glycoproteins. The N-glycans enzymatically released from glycoproteins were derivatised and subjected to MS/MS analysis in order to assign the correct structure. As an example, Fig 7 showed the MALDI-MS/MS spectrum of the precursor ion at m/z 1419.39, compared to the dansyl-labelled form. The oligosaccharide was identified as a high mannose structure on the basis of the interpretation of the fragmentation spectrum. Similar analyses were carried out on the other glycans thus confirming previous results. As a whole the data obtained by MALDI-MS/MS analyses of glycans are summarised in Table 1. It should be worth noting that DNS-derivatives are characterized by higher ionization efficiency than the unlabelled ones, producing better spectra with a higher S/N ratio with respect to the unmodified ones (Fig. 7B).

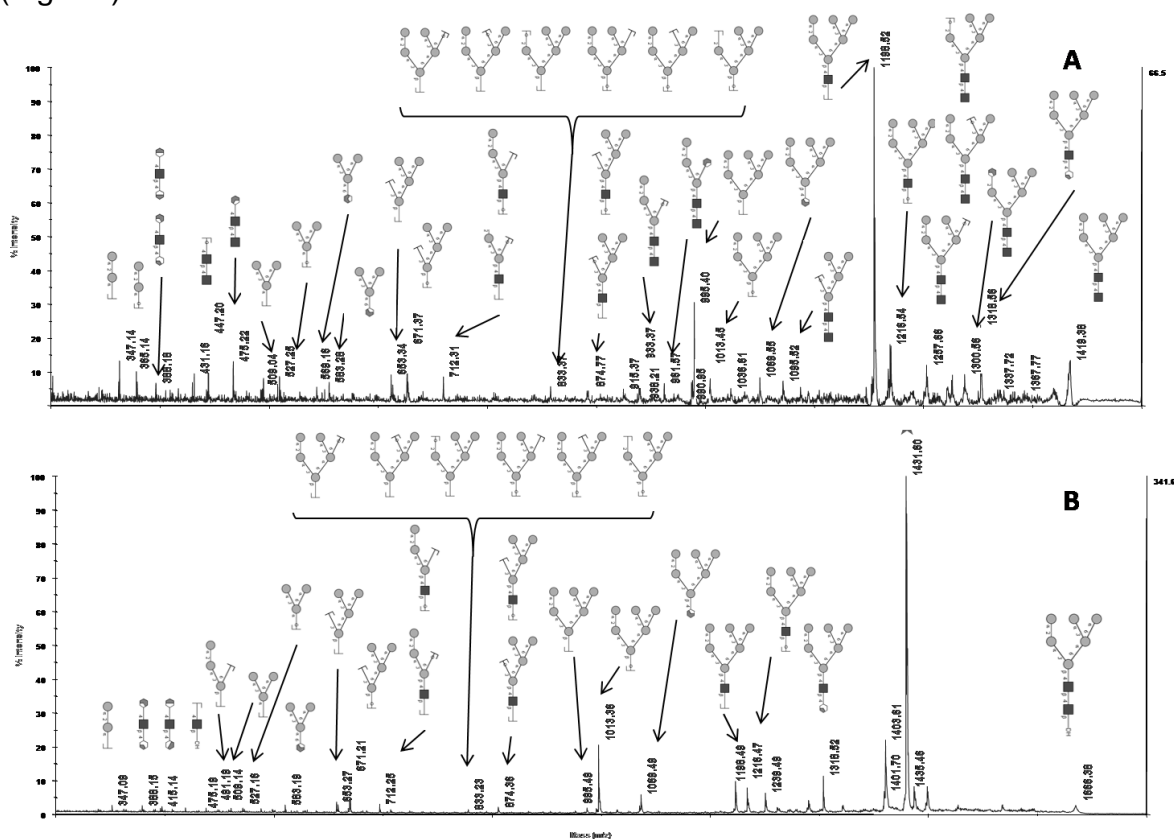


Fig. 7: positive ion mode MALDI-MS/MS spectrum of DNS-labelled oligosaccharide at m/z 1666.38 released from egg white ConA eluted fraction (B) compared to the unmodified one(A)

Thus, the derivatization protocol by dansyl-hydrazine was effective in the improvement of ionization efficiency, introducing a strong basic group. This labelling strategy not only increased peaks resolution in MS/MS spectra, but also introduced the possibility for further analysis of sugars by exploiting the potential of precursor ion scan analysis as already performed for the selective identification of other PTMs [10].

Fragmentation of oligosaccharides using MALDI LTQ Orbitrap for multi-stage MS

The use of MALDI LTQ Orbitrap for oligosaccharides fragmentation has been very recently reported by Karas and coworkers [20]. The authors explored the potential of MALDI source coupled to Orbitrap analyzer in three fragmentation techniques, namely CID, PQD and HCD. Fragmentation is highly influenced by the activation Q-values, depending on the radio frequency amplitude. The higher is this value, the greater is the dissociation efficiency, but this improvement is obtained at the expense of trapping fragments at low m/z [21]. During CID fragmentation Q-value remains constant and it is not possible to trap fragments at m/z lower than the 30% of the precursor. For this reason PQD (Pulsed-Q Dissociation) overcomes this limitation with a higher activation Q-value for precursor fragmentation followed by a rapid pulse to a lower Q-value, in order to capture fragments with lower m/z. Another attempt to solve the low mass cutoff problem is the use of HCD (Higher energy C-trap dissociation) for instruments endowed with a "C-trap", functioning as a quadrupole-like ion storage device between the LTQ and the Orbitrap analyzer. Fragments are detected with high resolution in the Orbitrap [22]. Unfortunately both PQD and HCD suffer from a very low signal intensity so that structural information to cover the entire mass range can be achieved only combining the three fragmentation procedures [20]. In addition to this a strong electronic noise sometimes suppresses ion detection. Here the use of MALDI LTQ Orbitrap is exploited for oligosaccharide fragmentation using a multi-stage mass spectrometry approach. The use of MSⁿ procedures in glycans characterization was already reported [23]. Experiments were carried out on a MALDI-LTQ Orbitrap at the École supérieure de physique et de chimie industrielles de la ville de Paris.

N-glycans from fetuin were analyzed by full scan MS, using DHB matrix as reported in Fig. 8, with resolution set at 60000 and a very high laser intensity at 60 µJ.

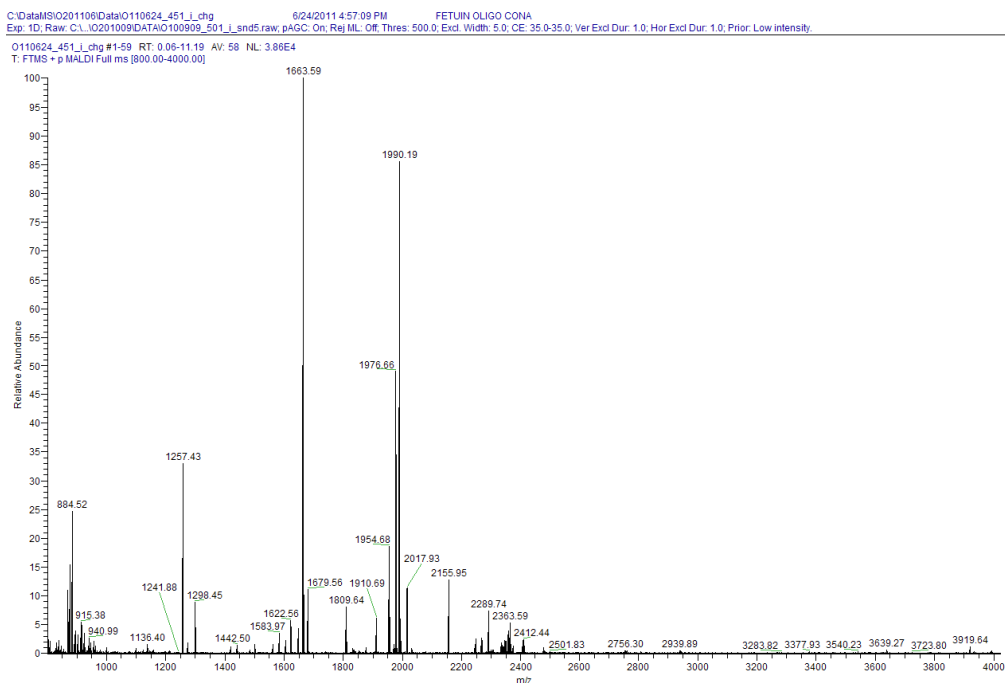


Fig. 8: MALDI-LTQ-Orbitrap full scan mass spectrum of oligosaccharides extracted from fetuin

The spectrum revealed a high mass accuracy due to orbitrap analysis. This characteristics offered the advantage to perform a reliable glycan profiling based on full scan spectra.

However structural elucidation of glycan molecules can be achieved only by fragmentation techniques. An important parameter for a good fragmentation is normalized collision energy. The first experiments were devoted to the optimization of this parameter for sugars, as reported in Fig.9 and 10, showing the MS/MS spectrum of the ion at m/z 1663.59. Fragmentation was achieved in CID using a constant Q-value of 0.25 and confirming the so-called “one third rule” stating the low mass cutoff during the trapping of fragments. Fig.9 reports the spectrum obtained with a collision energy of 35, whereas in Fig.10 spectrum this parameter was set up at 60, revealing an improvement in fragmentation efficiency.

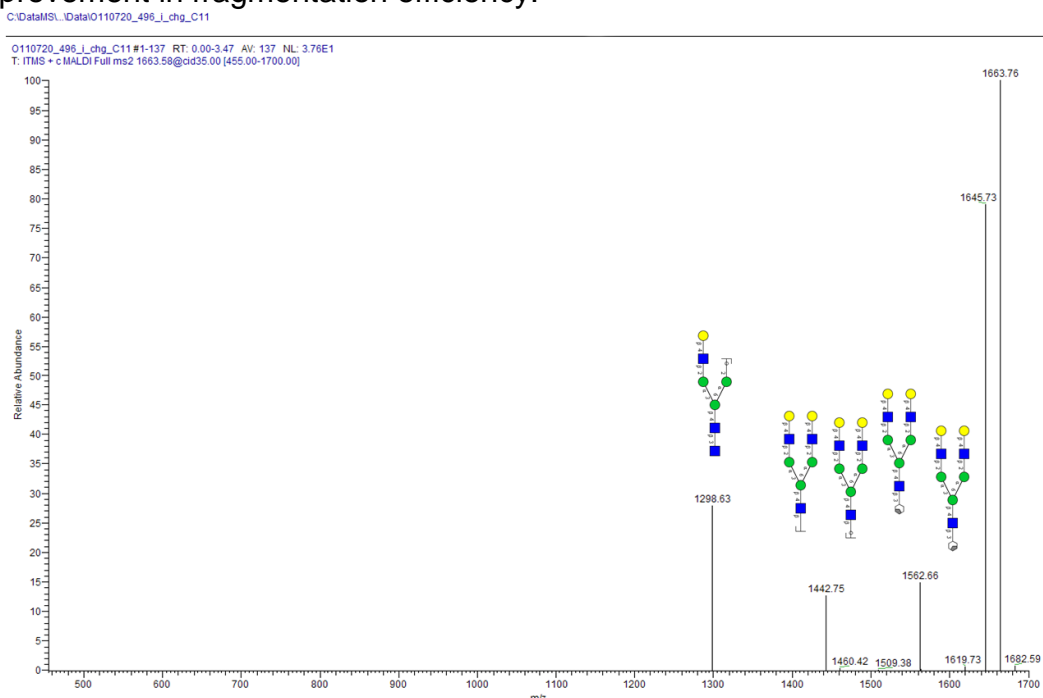


Fig. 9: MS/MS spectrum of ion at m/z 1663.59 with CE=35

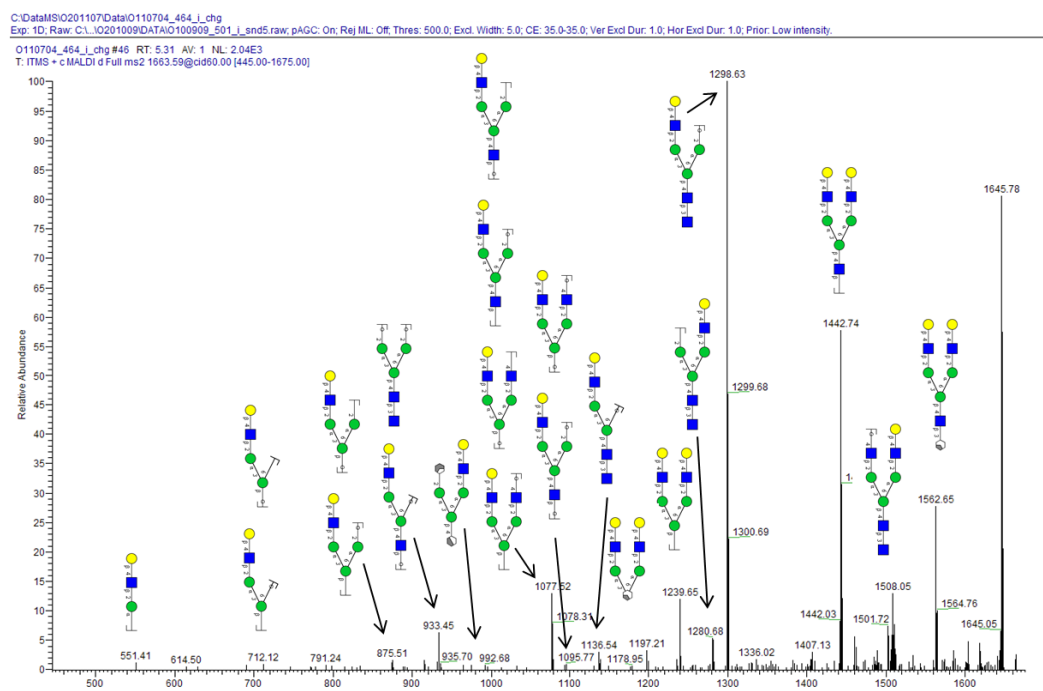


Fig. 10: MS/MS spectrum of ion at m/z 1663.59 with CE=60

The absence of significant signals at low masses and the need for this information in structural elucidation of oligosaccharides sequences convinced us to explore the potential of multi-stage fragmentation.

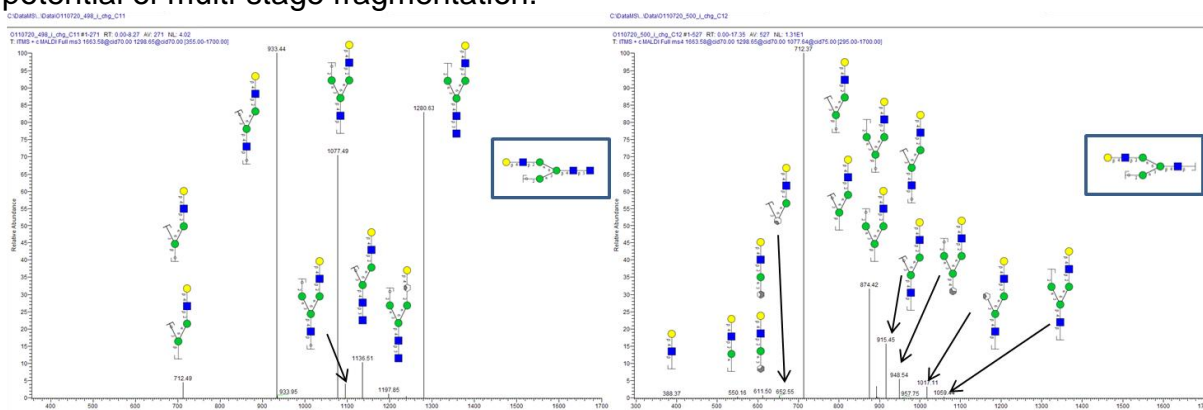


Fig. 11: MS³ and MS⁴ experiments for fragment ions at m/z 1298.65 and 1077.49, respectively

The ion at m/z 1298.65 was chosen for the further fragmentation event in MS³. The spectrum revealed the presence of essentially b and y-type ions, unfortunately covering a narrow mass range. A wider sequence coverage was achieved performing an MS⁴ experiment. The lower detected mass was at m/z 388.37, corresponding to the b₂ fragment (Fig.11).

Permethylated derivatization was carried out in order to improve structural characterization, trying to obtain data about branching via cross ring cleavages.

Fig. 12 represents the MS/MS spectrum of ion at 2070.47, corresponding to the biantennary structure at m/z 1663.59, after permethylation.

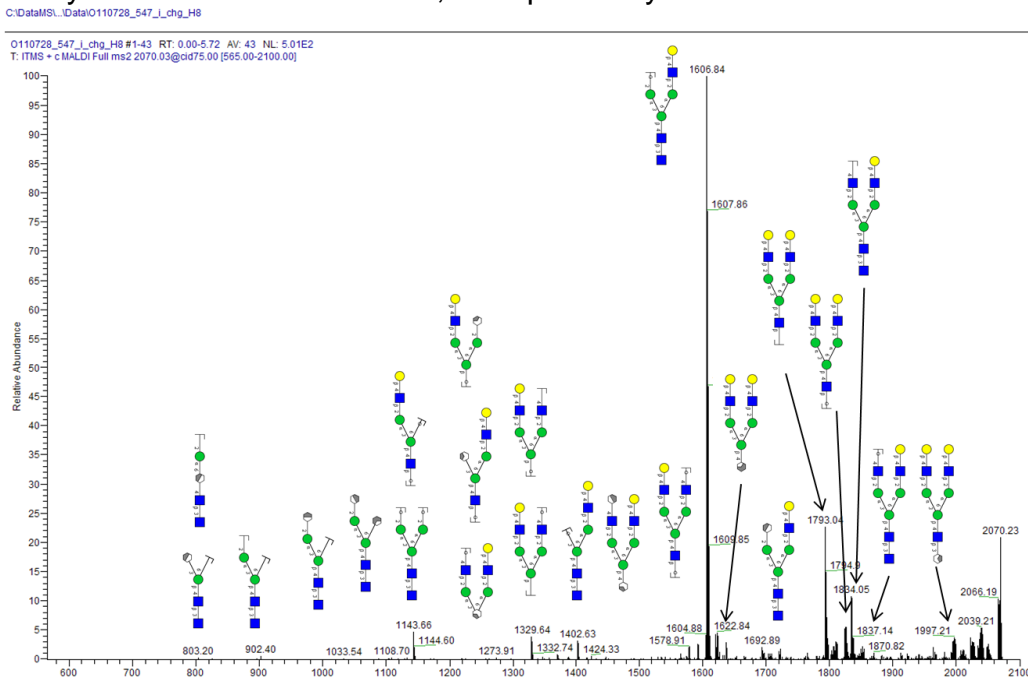


Fig. 12: MS/MS spectrum of ion at m/z 2070.47, corresponding to the biantennary structure after permethylation

Even in this case we assist to a prominence of b and y-type ions but z-type and cross ring cleavages started to be detected.

Further fragmentation events were carried out on this ion as illustrated in Fig. 13.

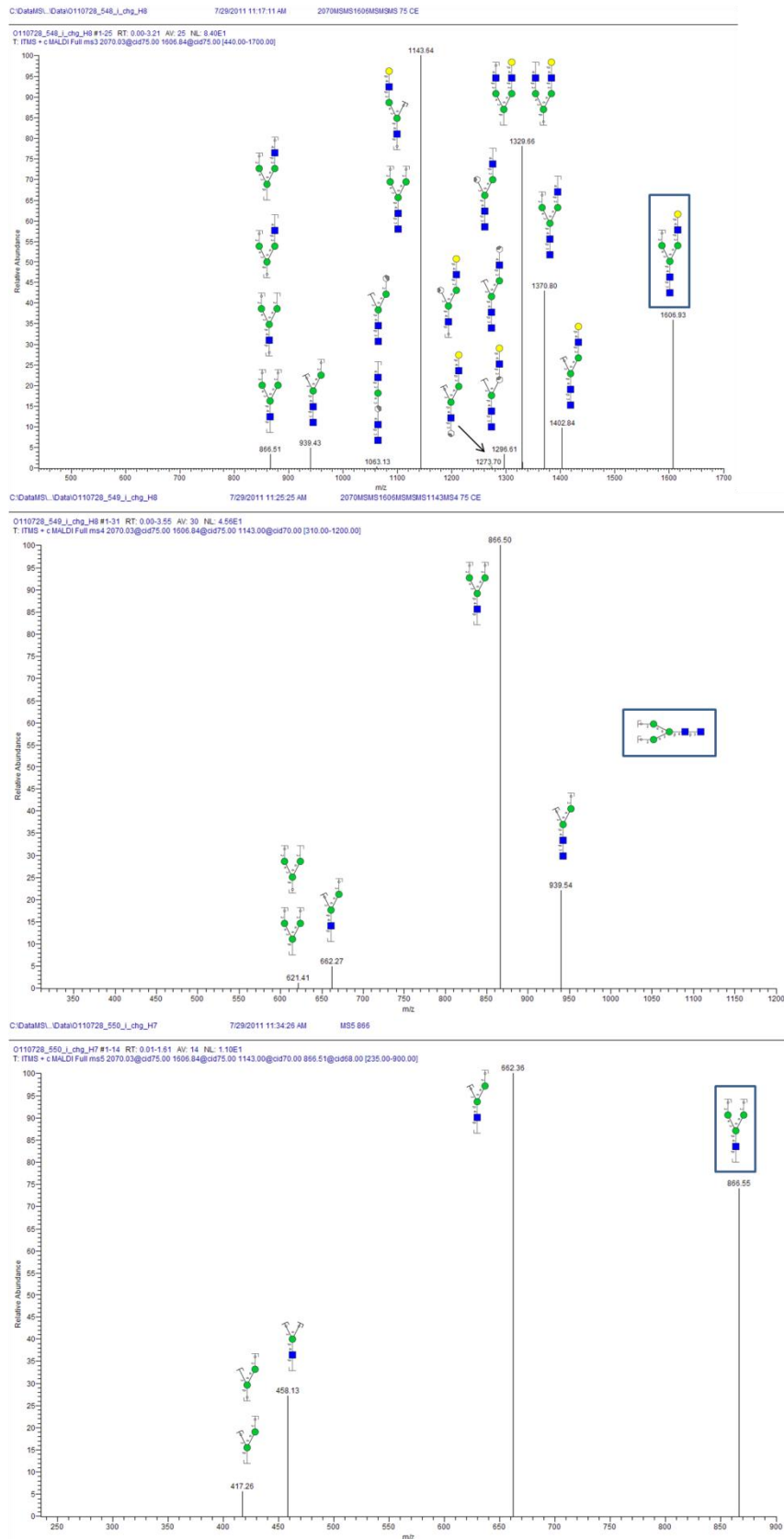


Fig. 13: MSⁿ experiments on fragment ions belonging to the permethylated biantennary structure in Fig.12. In the figure MS³, MS⁴ and MS⁵ are shown in panel A, B and C, respectively. Precursor ions are signed with a rectangle.

Fig.13A represents the MS³ spectrum of fragment ion at m/z 1606.93, corresponding to the Y₅ fragment. In panel B the MS⁴ spectrum of fragment ion at m/z 1143.64 is shown. This fragment was interpreted with two possible attributions from the MS³ spectrum, but the MS⁴ and MS⁵ experiments gave rise to fragments belonging solely to the Y_{4α} Y_{4β} fragment. These considerations led to the conclusion that multi-stage mass spectrometry is a valuable tool for oligosaccharides analysis.

Unfortunately, although permethylation allowed an increment in the number of cleavages, fragmentation spectra were very poor, probably due to sensitivity problems associated with the use of MALDI source. In particular a high sample consumption was observed using this MALDI source. This consumption was probably caused by the need for a high laser energy together with the use of the CPS system (Survey Chrystal Positioning system), which allowed the random choice of shot position by automatic crystal recognition. During the recognition process of DHB crystals the systems was not able to detect most of the sample and the analysis took a very long time leading to a high sample consumption. To conclude, the use of MALDI-LTQ Orbitrap did not satisfied our need for good fragmentation spectra, useful for structural characterization of oligosaccharides, demonstrating that MALDI-TOF/TOF analyses of derivatized sugars is still the gold standard in glycomics. However the extremely high accuracy and a very high resolution of the Orbitrap analyzer can be used for unambiguous assignment of structures during glycan profiling analyses.

II.4. Conclusions

As a whole, in this work an optimized dansylation protocol for glycoproteomic studies by MALDI-MS/MS analysis is reported. The presence of the aromatic dimethylaminonaphthalene group at the reducing end increased the glycan fragmentation leading to a better coverage of the oligosaccharide sequence. This tendency could be rationalized considering that the UV-adsorbing naphthalene group in the reducing end could increase the energy transfer from the laser in PSD-like fragmentation, favoring the glycan bond destabilization/protonation as observed for peptides. Glycan dansylation was demonstrated to be a rapid and cost-effectively valuable method useful to optimize the ionization efficiency of oligosaccharides. Well established techniques were conjugated with novel derivatisation procedure thus providing a new method for the glycoproteome investigation. The set up of a multi-stage mass spectrometry procedure opened the way for structural elucidation of glycans sequences with a novel mass spectrometer, MALDI LTQ Orbitrap.

II.5. References

- [1] Haslam SM, North SJ, Dell A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struct Biol.* 2006 Oct;16(5):584-91.
- [2] Lattová E, Perreault H. Method for investigation of oligosaccharides using phenylhydrazine derivatization. *Methods Mol Biol.* 2009;534:65-77.
- [3] Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wuhler M. Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem.* 2010 Aug;397(8):3457-81.
- [4] Palmese A, De Rosa C, Marino G, Amoresano A. Rapid Commun Mass Spectrom. 2011, 15, 25, 223-31.

- [5] Chiappetta G, Corbo C, Palmese A, Galli F, Piroddi M, Marino G, Amoresano A. *Proteomics*. 2009 Mar;9(6):1524-37. Erratum in: *Proteomics*. 2009 , 9, 3220.
- [6] Amoresano A, Cirulli C, Monti G, Quemeneur E, Marino G. *Methods Mol Biol*. 2009, 527, 173-90, ix.
- [7] Amoresano A, Chiappetta G, Pucci P, Marino G. *Methods Mol Biol*. 2008, 477, 15-29.
- [8] Cirulli C, Chiappetta G, Marino G, Mauri P, Amoresano A. *Anal Bioanal Chem*. 2008, 392, 147-59.
- [9] Amoresano A, Chiappetta G, Pucci P, D'Ischia M, Marino G. *Anal Chem*. 2007, 79, 2109-17.
- [10] Amoresano A, Monti G, Cirulli C, Marino G. *Rapid Commun Mass Spectrom*. 2006, 20, 1400-4.
- [11] Chiappetta G, Ndiaye S, Demey E, Haddad I, Marino G, Amoresano A, Vinh J. *Rapid Commun Mass Spectrom*. 2010;24, 3021-32.
- [12] Carpentieri A, Giangrande C, Pucci P, Amoresano A. Glycoproteome study in myocardial lesions serum by integrated mass spectrometry approach: preliminary insights. *Eur J Mass Spectrom (Chichester, Eng)*. 2010;16(1):123-49.
- [13] North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of N-linked and O-linked glycans. *Curr Opin Struct Biol*. 2009 Oct;19(5):498-506.
- [14] Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008 Apr;7(4):1650-9.
- [15] Mollé D, Jardin J, Piot M, Pasco M, Léonil J, Gagnaire V. Comparison of electrospray and matrix-assisted laser desorption ionization on the same hybrid quadrupole time-of-flight tandem mass spectrometer: application to bidimensional liquid chromatography of proteins from bovine milk fraction. *J Chromatogr A*. 2009 Mar 20;1216(12):2424-32.
- [16] Walker JM. The Dansyl-Edman method for peptide sequencing. *Methods Mol Biol*. 1994;32:329-34.
- [17] Cirulli C, Marino G, Amoresano A. Membrane proteome in *Escherichia coli* probed by MS3 mass spectrometry: a preliminary report. *Rapid Commun Mass Spectrom*. 2007;21(14):2389-97.
- [18] Napoli A, Aiello D, Di Donna L, Moschidis P, Sindona G. Vegetable proteomics: the detection of Ole e 1 isoallergens by peptide matching of MALDI MS/MS spectra of underivatized and dansylated glycopeptides. *J Proteome Res*. 2008 Jul;7(7):2723-32.
- [19] Domon B, Costello CE. A Systematic Nomenclature for Carbohydrate Fragmentations in FABMS/MS of Glycoconjugates., *Glycoconjugate J*. 1988, 5, 397-409.
- [20] Rohmer M, Baeumlisberger D, Stahl B, Bahr U, Karas M. Fragmentation of neutral oligosaccharides using the MALDI LTQ Orbitrap. *Int J Mass Spectrom* 305, (2-3), 15 August 2011:199-208
- [21] Guo T, Gan CS, Zhang H, Zhu Y, Kon OL, Sze SK. Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation. *J Proteome Res*. 2008 Nov;7(11):4831-40.

- [22] Scherl A, Shaffer SA, Taylor GK, Hernandez P, Appel RD, Binz PA, Goodlett DR. On the benefits of acquiring peptide fragment ions at high measured mass accuracy. *J Am Soc Mass Spectrom.* 2008 Jun;19(6):891-901.
- [23] Ojima N, Masuda K, Tanaka K, Nishimura O. Analysis of neutral oligosaccharides for structural characterization by matrix-assisted laser desorption/ionization quadrupole ion trap time-of-flight mass spectrometry. *J Mass Spectrom.* 2005 Mar;40(3):380-8

III. N-glycoproteome in the study of myocarditis

III.1. Introduction

Myocarditis is a group of diseases characterized by the inflammation of the cardiac muscle which might be related to viral (mainly parvovirus B19 and many others), protozoan (*Borrelia burgdoferi*, *Trypanosoma cruzi*, *Toxoplasma gondi*), bacterial (*Brucella*, *Corynebacterium diphtheriae*, *Gonococcus*, *Haemophilus influenzae*, *Actinomyces*, *Tropheryma whipplei*, *Vibrio cholerae*, *Borrelia burgdorferi*, *Leptospira*, *Rickettsia*), fungal (*Aspergillus*) and other non viral pathogens infections [1, 2, 3]; it has been reported that some types of myocarditis are caused by autoimmune mechanisms and this association is evidenced by concomitance with other autoimmune diseases and by the presence of autoantibodies [4]. However, this kind of inflammation might be caused by an hypersensitivity response to drugs [5]. Anyway, in each case the final effect is represented by myocardial infiltration of immunocompetent cells following any kind of cardiac injury and myocytes necrosis.

Myocarditis symptoms are various, ranging from chest pain that spontaneously resolves without treatment, to cardiogenic shock and sudden death [6, 7]. The major long-term consequence is dilated cardiomyopathy with chronic heart failure [8, 9].

Diagnostic tools currently available for this disease are essentially based to general investigations (such as electrocardiography) and analysis of the most abundant serum proteins, whose alteration is related to cardiac pathology, however not specifically related to myocarditis itself.

Inflammatory events generally involves different kinds of biological macromolecules, included proteins. Proteomics investigations offer all useful tools to deeply investigate this kind of alterations. Two dimensional electrophoresis coupled with software mediated image analysis, affinity chromatographies and especially Mass Spectrometry (MS) present high levels of sensitivity, accuracy and reproducibility to use in comparative studies of pathological status.

Thanks to the high sensitivity of the most modern analytical techniques, several research groups focus the research at level of peptides rather than at protein level [10, 11, 12]. In fact, these endogenous peptides are referred to as the peptidome. Initially, peptidomic analyses were conducted as a method to study neuropeptides and peptide hormones; these are signaling molecules that function in a variety of physiological processes [13, 14]. Recent studies have found large numbers of cellular peptides with half-lives of several seconds, raising the possibility that they may be involved in biological functions [15].

Here some preliminary data on the molecular basis of myocarditis have been collected using a proteomic approach. The analyses were focused on the serum proteins profiling, namely the study of the expression level and the characterization of glycoproteins involved in the pathology. A classical two-dimensional gel electrophoresis procedure to obtain protein maps and a "gel-free" comparison of the glycoproteomes in healthy and myocarditis human sera by using advanced mass spectrometry was reported. Free peptides identification in myocarditis affected patients (compared to healthy individual), was achieved in order to provide possible specific markers for this pathology. Moreover a comparison on the glycoproteomes in healthy and pathological human sera was carried out, by using single affinity chromatography (ConA) step and boronate affinity chromatography coupled with mass spectrometry techniques.

III.2. Materials and methods

Materials

Human serum samples from 8 healthy donors (all Caucasian 4 males and 4 females aged between 60 and 80 years, all other clinical informations were covered by laws on privacy) and from 5 myocarditis affected patients (all Caucasian 3 males and 2 females aged between 60 and 85 years, all other clinical informations were covered by laws on privacy), respectively, have been obtained from the "Servizio Analisi" Policlinico, Napoli. Aliquots of serum samples from different donors were pooled in order to obtain an average overview of glycoforms distribution.

Guanidine, dithiothreitol (DTT), trypsin, α -cyano-4-hydroxycinnamic acid were purchased from Sigma. Iodoacetamide (IAM), tris(hydroxymethyl)aminomethane, calcium chloride and ammonium bicarbonate (AMBIC) were purchased from Fluka as well as the MALDI matrix 2,5-, α -cyano-4-hydroxycinnamic. Methanol, trifluoroacetic acid (TFA) and acetonitrile (ACN) are HPLC grade type from Carlo Erba, whereas the other solvents are from Baker. Gel filtration columns PD-10 are from Pharmacia, the HPLC ones from Phenomenex, whereas the pre-packed columns Sep-pak C-18 are from Waters. Comassie Brilliant Blue was from Bio-Rad. PNGase F were purchased from Boehringer. Ion exchange resins Dowex H+ (50W-X8 50-100 mesh) was provided by BDH. Concanavalin A sepharose resin was purchased from Amersham Biosciences.

Protein concentration determination

Sera protein concentration was determined by Bradford assay method, using bovine serum albumin (BSA) as standard. Known amounts of BSA were diluted in 800 μ L of H₂O and then mixed to 200 μ L of Comassie Brilliant Blue. 5 different BSA concentration were determined by measuring absorbance at 595 nm and used to obtain a linear calibration curve. Three different sera dilutions were measured at 595 nm. Absorbance data were interpolated on the calibration curve, allowing the determination of protein concentration in the different samples.

Serum depletion

The depletion of the most abundant proteins from each serum sample was performed using Multiple Affinity Removal Spin Cartridges from Agilent. The procedure was performed at room temperature according to manufacturer's instructions and then immediately frozen to -20°C.

Free peptides analysis

300 μ L acetonitrile was added to 100 μ L of serum, incubated at room temperature for 30 minutes and then centrifuged at 13000 rpm. Supernatants were collected and concentrated by vacuum centrifugation (SAVANT) and resuspended in 20 μ L formic acid 0.1%. Samples were desalted by C18 Zip Tip (Millipore) and diluted 100 times prior mass spectrometry analyses.

Peptides were analyzed by a HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The capillary column works at a flow of 4 μ L/min, concentrating and washing the sample in a 40 nL enrichment column. The sample was then fractionated on a C18 reverse-phase capillary column (75 μ m~43 mm in the Agilent Technologies chip) at flow rate of 400 nL/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min.

Data were acquired through MassHunter software (Agilent Technologies). Proteins identification was achieved by using Mascot software (Matrix science), with a

tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and choosing methionine oxidation and glutamine conversion in pyro-glutamic acid as variable modifications. Peptides were also analyzed by MALDI-TOF/TOF using a 4800 Plus MALDI-TOF/TOF (Applied Biosystems). Samples were mixed on MALDI plate to the matrix consisting of a solution of 10 mg/mL α -cyano-4-hydroxycinnamic, whose preparation consisted in the resuspension of α -cyano-4-hydroxycinnamic in water and acetonitrile 10:1 (v/v). The instrument was calibrated using a mixture of standard peptides (Applied Biosystems). Spectra were register even in reflector positive. MS/MS spectra were performed with CID using air as collision gas. Spectra were manually interpreted.

2-D gel electrophoresis

IEF (first dimension) was carried out on non-linear wide-range immobilized pH gradients (pH 4-7; 7 cm long IPG strips; GE Healthcare, Uppsala, Sweden) and achieved using the EttanTM IPGphorTM system (GE Healthcare, Uppsala, Sweden). Analytical-run IPG-strips were rehydrated with 125 μ g of total proteins in 125 μ l of rehydratation buffer (urea 8 M, 11 CHAPS 2%, 0,5% (v/v) IPG Buffer, bromophenol blue 0,002%) for 12 h at 20°C. The strips were then focused according to the following electrical conditions at 20°C: 500 V for 30 min, from 1000 V for 30 min, 5000 V until a total of 15000 Vt was reached. After focusing IPG strips were equilibrated for 15 min in 6 M urea, 30% (vol/vol) glycerol, 2% (wt/vol) SDS, 0.05 M Tris-HCl, pH 6.8, 2% (wt/vol) DTT, and subsequently for 15 min in the same urea/SDS/Tris buffer solution but substituting the 2% (wt/vol) DTT with 2.5% (wt/vol) iodoacetamide. The second dimension was carried out on 12% polyacrylamide gels at 25 mA/gel constant current until the dye front reached the bottom of the gel. MS gel was stained with colloidal comassie.

Image analysis

Gels images were acquired with an Epson expression 1680 PRO scanner. Computer-aided 2-D image analysis was carried out using the ImageMasterTM 2D Platinum software (GE Healthcare, Uppsala, Sweden). Differentially expressed spots were selected for MS analysis. In order to find differentially expressed proteins, Comassie stained gel image of serum proteins from healthy individual was matched with the one of myocarditis affected patient. The apparent isoelectric points and molecular masses of the proteins were calculated with ImageMaster 2D Platinum 6.0 using identified proteins with known parameters as references. Relative spot volumes (%V) ($V = \text{integration of OD over the spot area}$; $\%V = V \text{ single spot} / V \text{ total spot}$) were used for quantitative analysis in order to decrease experimental errors. The normalized intensity of spots on three replicate 2-D gels was averaged and standard deviation was calculated for each condition.

In situ digestion

Protein spots were excised from the gel and destained by repetitive washes with 0.1 M NH_4HCO_3 pH 7.5 and acetonitrile. Enzymatic digestion was carried out with trypsin (12.5 ng/TI) in 10 mM ammonium bicarbonate buffer pH 7.8. Gel pieces were incubated at 4 °C for 2 h. Trypsin solution was then removed and a new aliquot of the same solution was added; samples were incubated for 18 h at 37 °C. A minimum reaction volume was used as to obtain the complete rehydration of the gel. Peptides were then extracted by washing the gel particles with 10mM ammonium bicarbonate and 1% formic acid 1 in 50% acetonitrile at room temperature.

Mass spectrometry and protein identification

LC-MS/MS analyses were performed as previously described for free peptides. Mass spectrometric obtained data were used for protein identification using the software MASCOT that compare peptide masses obtained by MS and MS/MS data of each tryptic digestion with the theoretical peptide masses from all the proteins accessible in the databases (Peptide Mass Fingerprinting, PMF). Database searches were performed in NCBI databank (National Center for Biotechnology Information), restricting the analysis to the pertinent taxonomies. The parameters used for the identification were: tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and cysteine carbamidomethylation as fixed modification. Variable modifications were methionine oxidation, glutamine conversion in pyro-glutamic acid, and asparagine deamidation.

Reduction and carbamidomethylation

An aliquot (1 mg) of each serum was lyophilized and dissolved in 200 μ L of denaturation buffer (guanidine 6 M, tris 0.3 M, EDTA 10 mM, pH 8). Reduction was carried out by using a 10:1 DTT:cysteins molar ratio. After incubation at 37°C for 2 h, carboxyamidomethylation was performed by adding iodoacetamide in an excess of alkylating agent of 5:1 with respect to the moles of thiol groups. The mixture was then incubated in the dark at room temperature for 30 min. The alkylation reaction was stopped by addition of formic acid, in order to achieve an acidic pH.

The excess of salts and reagents was then removed by gel filtration on PD-10 columns. Elution was performed using AMBIC 10 mM buffer. The collected fractions were analyzed by a spectrophotometer Beckman DU 7500, measuring the absorbance at 220 and 280 nm. Protein containing fractions were collected and concentrated in a centrifuge Speed-Vac.

Enzymatic digestions

Trypsin digestion was carried out in AMBIC 10 mM buffer using a 50:1 trypsin:protein mass ratio (w/w). The sample was incubated at 37°C for 12-16 h.

Concanavalin A Affinity chromatography

The peptide mixtures were incubated in batch, with 50 μ L of ConA sepharose resin, which had been activated and washed in a 20mM Tris-HCl buffer; 0,5M NaCl; 1mM CaCl₂ at pH 7.4, 4°C, for 2 h,. Then the ConA containing eppendorf was centrifuged at 3000 rpm at 4°C. The unbound peptides were removed in the supernatant. The resin was then washed six times with the same buffer described before.

The elution phase was based on methyl- α -D-mannopyranoside 0,25 M competition with glycopeptides and was performed for 30 min at 4°C. This procedure was repeated twice.

Boronate affinity chromatography

Some glycopeptides were purified within the peptide mixture using PBA-bound agarose (Sigma-Aldrich, Munich, Germany).

500 μ L of previously diluted sample (1:1) in equilibration buffer (50 mM taurine/NaOH, pH 8.7, containing 3–10 mM MgCl₂) was incubated with 200 μ L of pre-washed immobilized ligand resin for 1 h on ice and with gentle shaking. After transfer of the resin into 1,5 ml eppendorf tubes, the non-binding fraction was collected by low speed centrifugation (10 s, 500 \times g). The resin was then thoroughly rinsed with equilibration buffer (six washes of 150 μ L each) and 1 N NaCl (three

washes of 150 μ L each). For final elution of the bound fraction, a total of six washes (150 μ L each) with taurine buffer containing 50 mM sorbitol were used. Three successive fractions (150 μ L each) were pooled before further analysis.

RP-HPLC

Both the peptides and the glycopeptides needed a desalting step. Both the unbound and eluted peptides were desalted by HPLC on a C-18 resin. HPLC was performed on an Agilent chromatograph, at a flow rate of 0,2 ml/min and a linear gradient from 5% B (95% acetonitrile, 5% H₂O e 0,07% trifluoroacetic acid) to 95% B in 1 min, allowing the fast elution of the unfractionated peptides.

Deglycosylation

Glycopeptides were lyophilized, resuspended in 10 mM AMBIC and incubated with PNGase F (5 U), for 12-16 h at 37°C. Deglycosylation was carried out also on unbound peptides, in order to release the glycans not retained during the affinity purifications.

Oligosaccharide purification

For glycan purification, the sample was loaded on a C-18 silica based bonded phase (Sep-pak). The column was activated with 10 mL methanol and equilibrated with 10 mL isopropanol, 10 mL H₂O and 10 mL acetic acid 5%. The oligosaccharide component was eluted in 3 mL of 5% acetic acid as eluent whereas peptides were collected in 3 mL of 40% isopropanol and 5% acetic acid. Alternatively, a RP-HPLC step was performed by using a C-18 column and a gradient profile as follows: 7 min at 5% B, 5 to 95% B in 2 min and 10 min at 95% B.

Glycans MALDI-profiling

Glycans were collected and lyophilized. Then they were resuspended in 50 μ L of acetic acid 5% and treated with the ion exchange resin Dowex, in order to eliminate salts and other contaminants.

Both positive and negative MALDI-MS analyses and MALDI PSD experiments were carried out on a Voyager DE-STR-Pro instrument operating in reflectron mode (Applied Biosystems, Framingham, MA). The MALDI matrices were prepared by dissolving 25 mg of DHB (2,5 dihydroxy benzoic acid) or THAP (2,4,6 – Trihydroxyacetophenone) in 1 mL water/acetonitrile 10:1 or 10 mg *-cyano-4-hydroxycinnamic acid in 1 mL of acetonitrile/0.2% trifluoroacetic acid (70:30 v/v). Typically, 1 μ L of matrix was applied to the metallic sample plate and 1 μ L of analyte was then added. Acceleration and reflector voltages were set up as follows: target voltage at 20 kV, first grid at 66% of target voltage, delayed extraction at 200 ns to obtain the best signal-to-noise ratios and the best possible isotopic resolution with two-point external calibration using maltoheptose (exact isotopic mass, 1153.02 Da) and tetrantennary sialylated oligosaccharide (exact isotopic mass, 2686,03 Da) as calibrants, resulting in a mass accuracy of 300 ppm. Each spectrum represents the sum of 1500 laser pulses from randomly chosen spots per sample position. All analyses were conducted in triplicate.

The PSD fragment spectra were acquired after pseudomolecular cation selection using the time ion selector. Fragment ions were refocused by stepwise reducing the reflector voltage by 20%. The individual segments were then stitched together using the Applied Biosystems software. Raw data were analysed using the computer software provided by the manufacturers and are reported as monoisotopic masses.

Peptides LC-MS/MS

Samples were diluted in formic acid 0.1% in order to obtain a final concentration of 20 fmol/ μ L. 8 μ L were loaded and analyzed.

NanoLC-MS² experiments were performed on a 4000 Q-Trap mass spectrometer (Applied Biosystems) coupled to an 1100 nanoHPLC system (Agilent Technologies). Peptides were separated on a Agilent reversed-phase column (Zorbax 300 SBC18, 150 mm), at a flow rate of 0.2 μ L/min using 0.1% formic acid, 2% ACN in water as solvent A 0.1% and formic acid, 2% water in ACN as solvent B. The elution was accomplished by a 5-65% linear gradient of solvent B in 60 min. A micro-ionspray source was used at 2.5 kV with liquid coupling, with a declustering potential of 50 V, using an uncoated silica tip from New Objectives (Ringoos, NJ). Alternatively peptides were analyzed by a HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The capillary column works at a flow of 4 μ L/min, concentrating and washing the sample in a 40 nL enrichment column. The sample was then fractionated on a C18 reverse-phase capillary column (75 μ m~43 mm in the Agilent Technologies chip) at flow rate of 400 nl/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min. Data were acquired through MassHunter software (Agilent Technologies).

Proteins were identified by searching the MS/MS spectra against the SwissProt or NCBI databases using the Mascot in house search engine (Matrix Science, London, UK).

III.3. Results and discussion

Free peptides analysis

The analyses to investigate free peptides both in healthy and in pathological samples were accomplished by using a gel-free approach. The peptide component of the serum was separated precipitating proteins by ACN precipitation and recovering supernatants. Collected fraction were desalted and analyzed by tandem mass spectrometry. The stringency of scoring parameters of the MASCOT algorithm minimized the number of false positive identifications. Most MS/MS spectra giving positive hits were derived from doubly and triply charged precursor ions that resulted predominantly in y-ion series.

Triplicate LC-MS/MS analysis of supernatants after ACN precipitation of serum proteins, showed the occurrence of many free peptides in both analyzed sera. As reported in Table 1, the total number of detected and identified peptides was 41. Among these, 9 peptides were unique in the pathologic sample. It should be noted that some peptides were identified in both samples.

Table 1: List of free peptides identified by LC-MS/MS. Averaged area of chromatographic peaks from healthy(H) and myocarditis (M) ratio, indicates that some of them were differently represented. Peptide sequences were validated by MALDI-TOF/TOF analyses too.

m/z	RT	Sequence	Peptide	Protein	H/M
567.95	17.09	QAGAAGSRMNF RPGVLS	(650-666)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
513.78	17.70	YYLQGAKIPKPEASFSR	(627-644)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
823.17	22.11	MNFRPGVLSRQLGLPGPPDVPDHAAYHPF	(658-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
1005.98	22.78	QLGLPGPPDVPDHAAYHPF	(669-687)	Q14624 Inter-alpha-	0.68 \pm 0.18

				trypsin inhibitor heavy chain H4	
757.71	19.96	SRQLGLPGPPDVPDHAAYHPF	(667-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	0.27±0.09
681.85	22.07	PGVLSSRQLGLPGPPDVPDHAAYHPF	(662-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	0.18±0.03
576.90	20.69	PGVLSSRQLGLPGPPDVPDHAAYHPFR	(662-688)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
379.71	0.55	LAEGGGVR	(28-35)	P02671 Fibrinogen alpha chain	25±3.11
453.24	3.65	FLAEGGGVR	(27-35)	P02671 Fibrinogen alpha chain	2.48±0.57
510.76	11.29	DFLAEGGGVR	(26-35)	P02671 Fibrinogen alpha chain	1.33±0.76
539.27	11.70	GDFLAEGGGVR	(25-35)	P02671 Fibrinogen alpha chain	3.15±1.02
597.77	15.59	SSEGDFLAEGGGV	(22-34)	P02671 Fibrinogen alpha chain	0.85±0.23
603.79	12.57	EGDFLAEGGGVR	(24-35)	P02671 Fibrinogen alpha chain	2.42±0.63
632.30	13.50	GEGDFLAEGGGVR	(23-35)	P02671 Fibrinogen alpha chain	2.72±0.71
655.28	16.03	DSGEGDFLAEGGGV	(21-34)	P02671 Fibrinogen alpha chain	1.23±0.36
675.81	13.67	SSEGDFLAEGGGVR	(22-35)	P02671 Fibrinogen alpha chain	2.32±0.58
690.80	16.13	ADSGEGDFLAEGGGV	(20-34)	P02671 Fibrinogen alpha chain	3.25±0.82
733.33	14.04	DSGEGDFLAEGGGVR	(21-35)	P02671 Fibrinogen alpha chain	1.59±0.12
768.85	14.08	ADSGEGDFLAEGGGVR	(20-35)	P02671 Fibrinogen alpha chain	6.16±1.01
851.71	13.58	SSSYSKQFTSSTSYPNRDSTFES	(576-598)	P02671 Fibrinogen alpha chain	0.65±0.17
693.06	13.06	SSSYSKQFTSSTSYPNRDSTFESKS	(576-600)	P02671 Fibrinogen alpha chain	M: not detected
733.83	14.48	SSSYSKQFTSSTSYPNRDSTFESKSY	(576-601)	P02671 Fibrinogen alpha chain	M: not detected
619.75	17.89	QGVNDNEEGFF	(31-41)	P02675 Fibrinogen beta chain	3.25±0.89
663.26	16.54	QGVNDNEEGFFS	(31-42)	P02675 Fibrinogen beta chain	1.97±0.41
696.28	17.64	QGVNDNEEGFFSA	(31-43)	P02675 Fibrinogen beta chain	2.31±0.53
404.55	18.90	RIHWESASLL	(1310-1319)	P01024 Complement C3	H: not detected
402.22	14.29	THRIHWESASLLR	(1308-1320)	P01024 Complement C3	H: not detected
445.25	16.98	SKITHRIHWESASLL	(1305-1319)	P01024 Complement C3	H: not detected
415.20	7.60	HWESASL	(1312-1318)	P01024 Complement C3	H: not detected
471.74	16.21	HWESASLL	(1312-1319)	P01024 Complement C3	H: not detected
851.07	22.22	TLEIPGNSDPNMIPDGFNSYVR	(957-979)	P0C0L4 Complement C4-A	H: not detected
1054.53	25.86	DDPDAPLQPVTPLQLFEGR	(1429-1447)	P0C0L4 Complement C4-A	H: not detected
489.96	20.26	RHPDYSVLLLLR	(169-180)	P02768 Serum Albumin	0.43±0.17
417.91	16.67	KFQNALLVRY	(426-435)	P02768 Serum Albumin	H: not detected
547.31	16.25	KVPQVSTPTLVEVSR	(438-452)	P02768 Serum Albumin	H: not detected
868.11	21.57	AVPPNNSNAEEDDLPTVELQGVVPR	(14-38)	P00488 Coagulation factor XIII A	2.40±0.84
920.14	21.09	RAVPPNNSNAEEDDLPTVELQGVVPR	(13-38)	P00488 Coagulation factor XIII A	23.60±3.61
837.93	21.39	TAFGRRRAVPPNNSNAEEDDLPTVELQGVVPR	(7-38)	P00488 Coagulation factor XIII A	91.32±7.39
781.37	17.84	TATSEYQTFNPR	(315-327)	P00734 Prothrombin	H: not

					detected
868.46	25.47	TGIFTDQVLSVLKGEE	(86-101)	P02655 Apolipoprotein C-II	H: not detected
572.95	12.91	DALSSVQESQVAQQAR	(45-60)	P02656 Apolipoprotein C-III	H: not detected
803.76	21.14	AATVGLAGQPLQERAQAWGERL	(210-232)	P02649 Apolipoprotein E	0.02

The identification of poorly present peptides demonstrated that LC-MS/MS proved to be a very sensitive and reproducible analysis. A confirmation of these data was provided by another fragmentation technique, MALDI-TOF/TOF. As Fig. 1 shows, the fragmentation pattern of the peptide –ADSGEGDFLAEGGGV R- (from alpha fibrinogen protein) confirmed the data collected by using LC-MS/MS analysis. Most of these peptides are related to different proteolytic activities on proteins involved in acute phase or inflammatory events, such as myocarditis itself. Among identified peptides, the molecular species, namely, peptide 662-668 from inter-alpha-trypsin inhibitor heavy chain H4 and peptide 1305-1319 from complement C3 correspond to free bioactive peptides, whose activity might be related again to inflammatory events [16, 17]. Qualitative and semi-quantitative differences were detected in the analysis of the peptidomas from healthy and pathological samples. In fact, as reported in the Table 1, averaged area of each peak of all detected peptides resulted to be different, showing that some of them were differently represented in each sample. The above mentioned peptides from complement C3 and inter-alpha trypsin inhibitor were poorly represented in the healthy serum samples, whereas fibrinogen alpha chain peptides were strongly represented in these sera, as well as coagulation factor XIII A peptides.

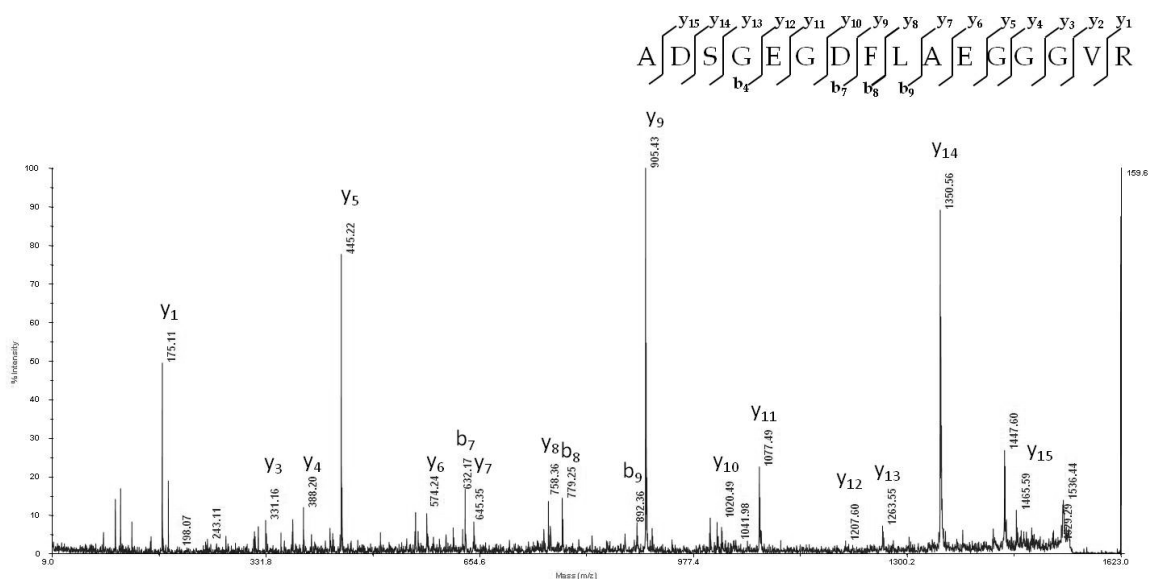


Fig. 1: Positive ion mode MALDI-TOF/TOF fragmentation spectrum of m/z 1536.44 corresponding to peptide 20-35 from fibrinogen alpha chain protein from healthy serum.

2-D gel electrophoresis

After depletion of the most abundant proteins, the two sera samples were separated by two dimensional gel electrophoresis, thus leading to the construction of 2D protein maps of healthy and myocarditis samples. Each gel was blue coomassie stained. Image analysis was performed on the two sets of 2D maps (from healthy control and

pathologic sera) clearly showing that the protein profile is quite different, as reported in figure 2.

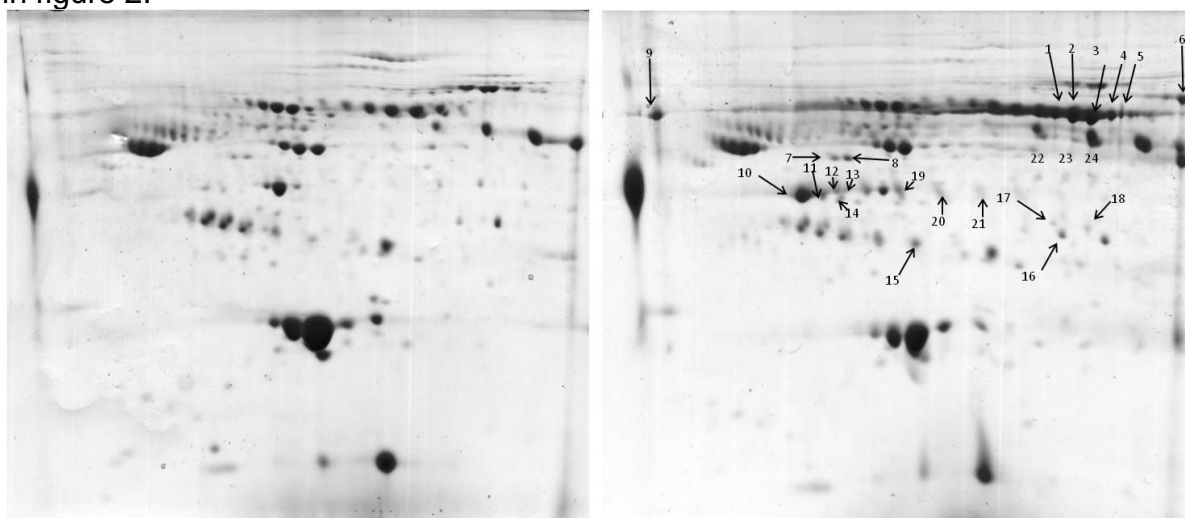


Fig. 2: Two-dimensional electrophoresis gels of healthy (A) and myocarditis affected (B) sera proteins. Arrows and numbers in (B) correspond to numbers of identified spots in Table 2.

All the selected spots were taken by the myocarditis sample and they are summarized in table 2.

Table 2: List of proteins identified by LC-MS/MS of spots excised from two-dimensional gels of sera taken from myocarditis affected patients.

Spot	Accession number	Protein
1	P02790	Hemopexin
2	P02790	Hemopexin
3	P02790	Hemopexin
4	P02790	Hemopexin
5	P02790	Hemopexin
6	P02787	Serotransferrin
7	P01008	Antithrombin III
8	P01008	Antithrombin III
9	P01011	Alpha-1 antichymotrypsin
10	P01024	Complement C3
11	P04196	Histidine rich glycoprotein
12	P01024	Complement C3
13	P00738	Haptoglobin
14	P04196	Histidine rich glycoprotein
15	Q14624	Inter-alpha-trypsin inhibitor heavy chain H4
16	P08603	Complement factor H
17	P08603	Complement factor H
18	P03952	Plasma kallikrein
19	P06727	Apolipoprotein A-IV
20	P00738	Haptoglobin
21	P00738	Haptoglobin
22	P36955	Pigment epithelium-derived factor
23	P36955	Pigment epithelium-derived factor
24	P36955	Pigment epithelium-derived factor

Protein spots indicated as differentially expressed were excised from the gel, *in situ* digested and then submitted to LC-MS/MS according to the peptide mass fingerprinting procedure. MS and MS-MS obtained data were used to search for a non redundant sequence using the in-house MASCOT software, taking advantage of the specificity of trypsin and of the taxonomic category of the samples. The number of measured masses that matched within the given mass accuracy of 20 ppm was recorded and the proteins that had the highest number of peptide matches were examined.

This approach allowed the identification of many proteins differently expressed in the two sera. Among identified proteins, hemopexin [18] complement C3 [19,20], plasma kallikrein [21] are undoubtedly related to inflammatory events and most interestingly they resulted clearly over expressed in the pathologic sample.

A preliminary speculation on identified proteins might be focused on hemopexin. As shown by literature data [18,22,23], hemopexin is a serum protein committed to the heme scavenging; heme is released or lost by the turnover of heme proteins such as haemoglobin, or by haemolysis caused by parasitic infection. Thus hemopexin might have a protection role from the oxidative damage that free heme can cause[23]. Myocarditis itself is not related to haemolysis phenomena, but some viral infections may cause it, therefore, finding an association between very high level of hemopexin and myocarditis affection might open the way to novel putative biomarkers. Moreover we found a certain association between these data and the data collected in the peptidoma analysis; in particular we could detect among free peptides, some from proteins Complement C3, Inter-alpha-trypsin inhibitor heavy chain H4, Antithrombin III which results over expressed in pathologic serum

Assessment of N-glycosylation sites

Aliquots of serum sample from different donors were pooled together in order to obtain an average distribution of glycoforms. Intact serum samples without any previous depletion were used for the determination of N-glycosylation sites occupancy. The two pooled samples were reduced, alkylated, digested with trypsin and then subjected to Concanavalin affinity selection, directly performed in batch after trypsin digestion. The recovered glycopeptides were then subjected to N-glycosidase F treatment in order to enhance the identification of N-glycosylation sites by using LC-MS/MS and obtaining a glycan profiling by MALDI-MS. The strategy is illustrated in Fig. 3.

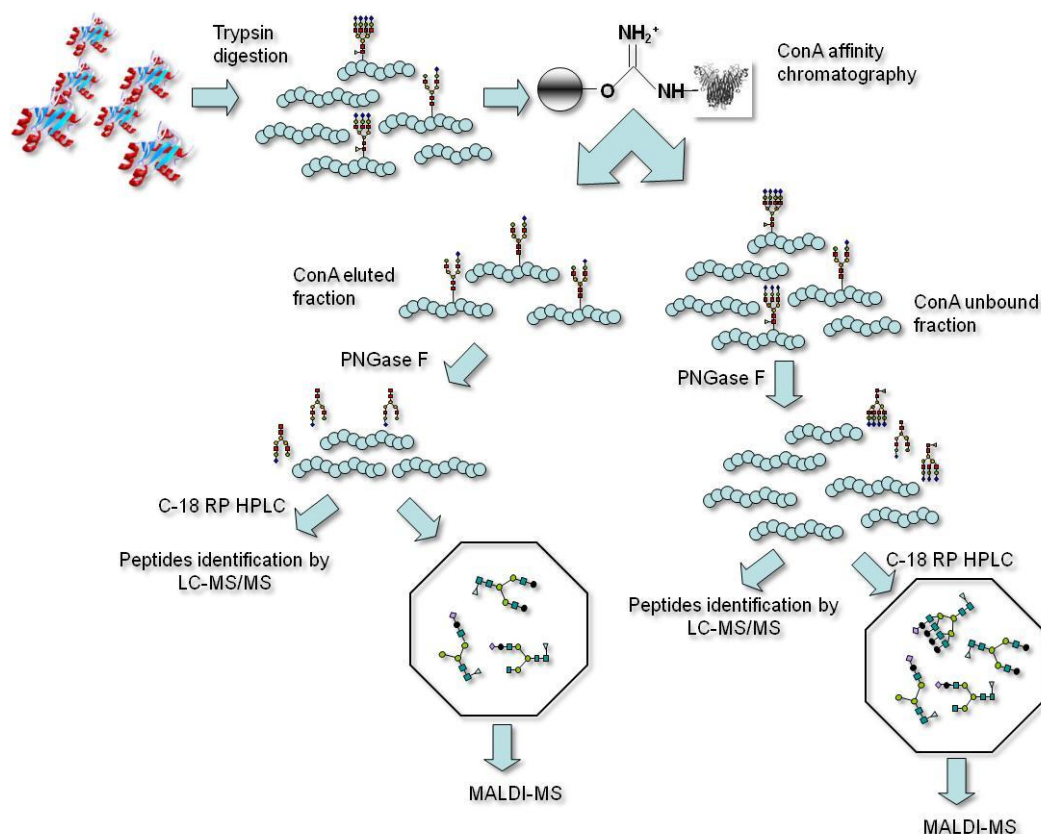


Fig. 3: scheme of the procedure used to obtain an overview of the glycosylation profiling by enriching for N-linked glycopeptides to be analyzed by LC-MS/MS, obtaining a glycan distribution to be analyzed by MALDI-MS.

The presence of putative N-glycosylation sites was inferred by the increase of 0.98 Da in the peptide mass value, due to PNGase F treatment leading to the conversion of asparagine residues to aspartic acid. The fine localization of the modified asparagines was achieved by LC-MS/MS on a 4000 Q-Trap mass spectrometer coupled to a nanoLC system. Fig. 4 shows the daughter ion mass spectrum of a peptide retained by ConA affinity resin after deglycosylation. The MS/MS spectrum of the precursor ion at m/z 878.96 revealed a clear series of y -ions, which make the sequence of the peptide easily interpretable as YLGN*ATAIFFLPDEGK. BLAST analysis of this sequence attributed the peptide as the fragment 268-283 of the human alpha-1-antitrypsin precursor in which Asn271 shows the consensus sequence for glycosylation (Asn-X-Ser/Thr). The effective occurrence of a N-glycosylation site at Asn-271 was demonstrated by the key y^{12} fragment ion at m/z 1308.68 in the MS/MS spectrum. This fragment occurred 0.98 Da lower than expected showing the loss of Asp271 instead of Asn thus confirming the conversion of the asparagine to aspartic acid following PNGase F stripping.

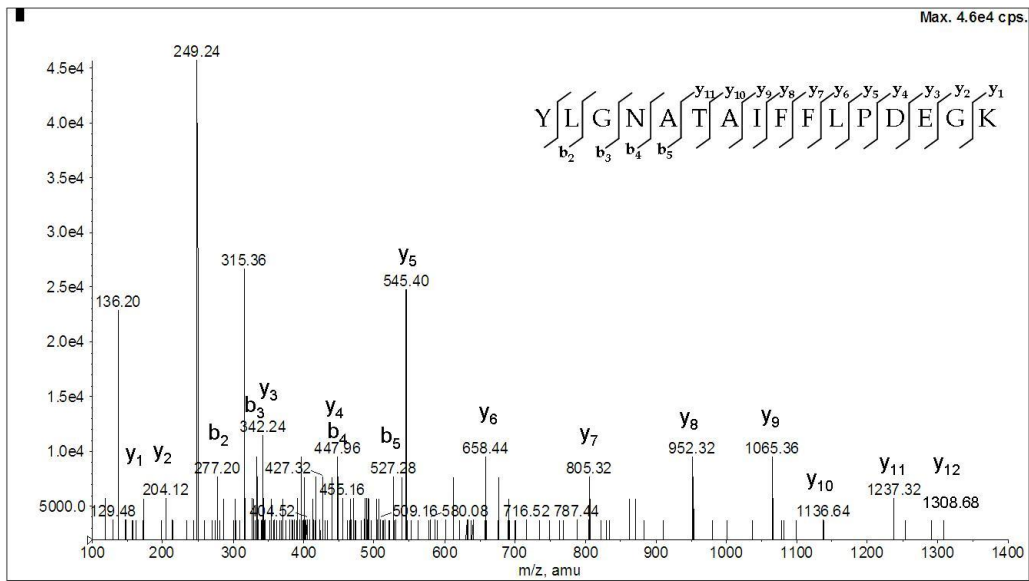


Fig. 4: MS/MS spectrum of the doubly charged precursor ion at m/z 878.96. The fragmentation spectrum shows the presence of a N-glycosylation site at Asn-271

Fig.5 shows the MS/MS spectrum of the triply charged precursor ion at m/z 536.52 corresponding to the fragment 3405-3419 of Apolipoprotein B. The series of y ions from which peptides identification was inferred is clearly visible in the daughter ion spectrum. Again, the N-glycosylation site could easily be assigned by the presence of the key y^8 fragment ion that occurred 1 Da lower than expected (836.40 instead of 837.47) demonstrating that Asn3411 was glycosylated and then had been converted to Asp by PNGase F.

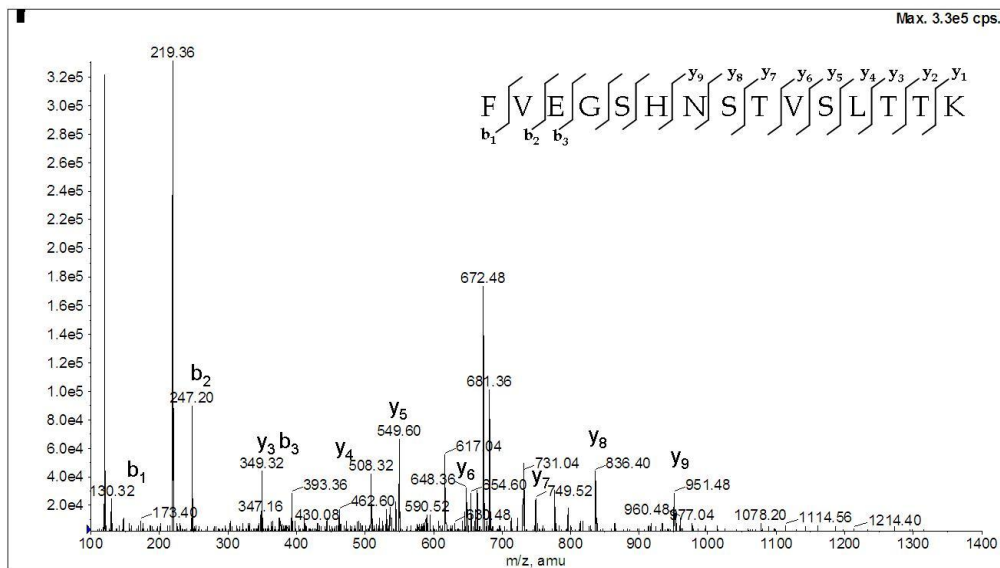


Fig. 5: MS/MS spectrum of the triply charged precursor ion at m/z 536.14. The fragmentation spectrum shows the increment of 0.98 Da for Asn3411 conversion into Asp

Similar analyses were carried out on the unbound Con A fractions, mainly containing non-glycosylated peptides and highly branched glycopeptides. The unbound fraction was also deglycosylated by PNGase F treatment, yielding a mixture of deglycosylated and unmodified peptides that was directly analysed by LC-MS/MS. Figure 6 shows the daughter ion spectrum of the precursor ion at m/z 628.38 that was assigned to the peptide 1004-1014 from the alpha-2-macroglobulin precursor. The key y^5 ion at m/z 624.48 clearly demonstrated the occurrence of an asparagine residue at position 1009 that resulted to be unmodified following the deglycosylation step.

Spontaneous deamidation of Asn residues has been reported, but it seems rather frequent in peptides containing the sequence Asn-Gly [25], therefore for the unambiguous assignment of the N-glycosylation site the presence of the N-X-S/T motif is absolutely necessary.

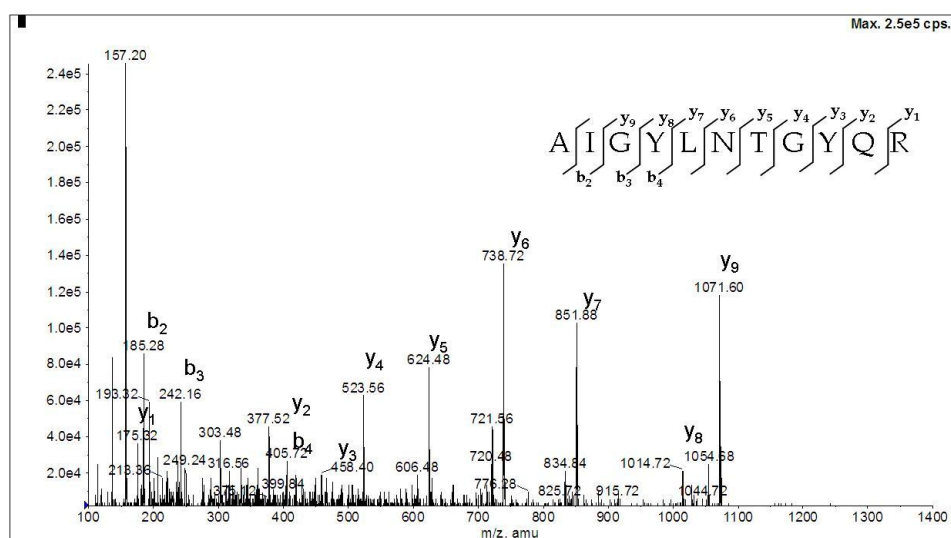


Fig. 6: fragmentation spectrum of a doubly charged precursor ion at m/z 628.38. In this case, asparagines residue is unmodified.

The overall data from the LC-MS/MS analyses of both the ConA bound and unbound fractions were summarised in Table 3. The table exhibits a total of 49 glycoproteins containing 68 N-glycosylation sites. Among these sites, 50 had already been described and were confirmed while 18 were only annotated as potential glycosylation sites due to the presence of the NX(S/T) consensus sequence and were then newly identified.

Table 3: LC/MSMS analysis for the identification of glycosylation sites by ConA (UH = unbound fraction of healthy serum; UL = unbound fraction of myocardial lesions serum; UK = unbound fraction of hepatic carcinoma serum; EH = eluted fraction of healthy serum serum; EL = eluted fraction of myocardial lesions serum; EK = eluted fraction of hepatic carcinoma serum)

	Protein	Sequence	Peptide	
P01011	Alpha-1-antichymotrypsin	YTGN*ASALFILPDQDK ^{UL}	268-283	Confirmed
P02763	Alpha-1-acid glycoprotein	SVQEIQATFFYFTPN*KTEDTIFLR ^{EH} QDQCIYN*TTYLNVQR ^{UH,UL}	58-81 87-101	Confirmed Confirmed
P01009	Alpha-1-antitrypsin	YLGN*ATAIFFLPDEGK ^{EH,UH,EL,UL}	268-283	Confirmed
P01023	Alpha-2-macroglobulin	SLGNVN*FTVSAAEALSEQELCGTEVPSVPEHGRK ^{EH}	864-886	Confirmed
P43652	Afamin	DIENFN*STQK ^{EH,EL} YAEDKFN*ETTEK ^{EH}	28-37 396-407	Confirmed Confirmed
P01008	Antithrombin III	LGACN*DTLQQLMEVFK ^{EH} SLTFN*ETYQDISELVYGAKE ^{EH}	124-139 183-201	Confirmed Confirmed
P04114	Apolipoprotein B-100	FN*SSYLQGTNQITGR ^{EH,EL} FVEGSHN*STVSLTTK ^{EH} AEEEEMLN*VSLVCPK ^{EL} YDFN*SSMLYSTAK ^{EL}	1522-1536 3405-3419 27-41 3462-3474	Confirmed Potential Potential Confirmed
P05090	Apolipoprotein D	ADGTVNQIEGEATPVN*LTEPAK ^{EL}	83-104	Confirmed
P02749	Apolipoprotein H	VYKPSAGN*NSLYR ^{EH} LGN*WSAMPCK ^{EL}	155-167 251-261	Confirmed Confirmed
O75882	Attractin	IDSTGN*VTNELR ^{EH,EL} GPVKMPSQAPTGNFYQPLLN*SSMCLEDSR ^{EH}	411-422 1023-1052	Confirmed Confirmed
P00450	Ceruloplasmin	EHEGAIYPDN*TTDFQR ^{EH,EL} EN*LTAPGSDSAVFFEQGTR ^{EH,EL}	129-144 396-414	Confirmed Confirmed
P08603	Complement factor H	ISEEN*ETTCYMGK ^{EH} MDGASN*VTCINSR ^{EH,EL} IPCSQPPQIEHGTIN*SSR ^{EL} SPDVIN*GSPISQK ^{UH}	907-919 1024-1036 868-885 212-224	Confirmed Potential Confirmed Potential
P10909	Clusterin	LAN*LTQGEDQYYLR ^{EH,EL}	372-385	Confirmed
Q8I WV2	Contactin-4	LN*GTDVDTGMDFR ^{EL}	64-76	Confirmed
P05156	Complement factor 1	FLNN*GTCTAEGK ^{EH,EL}	100-111	Confirmed
P01024	Complement C3	TVLTPATNHMGN*VTFTIPANR ^{EH}	74-94	Confirmed
P0C0L4	Complement C4-A	GLN*VTLSSSTGR ^{EH,EL}	1326-1336	Confirmed
P02748	Complement component C9	AVN*ITSENLI DDVVSLIR ^{EH}	413-430	Confirmed
Q14517	Protocadherin-fat 1	QVYN*LTVRAKDK ^{EL} FSMDYKTGALT VQN*TTQLRSR ^{EL}	994-1005 1930-1950	Potential Potential
P02765	Alpha-2-HS-glycoprotein	VCQDCPLLAPLN*DTR ^{EH} AALAAFNAQNN*GSNFQLEEISR ^{EH}	145-159 166-187	Confirmed Confirmed

Q03591	Complement factor H related protein 1	LQNNENN*ISCVER ^{EH}	120-132	Confirmed
Q13439	Golgin subfamily A member 4	HN*STLKQLMREFNTQLAQK ^{EH}	1990-2008	Potential
P02790	Hemopexin	SWPAVGN*CSSALR ^{EH,EL} ALPQPQN*VTSLLGCTH ^{UH,EH,EL}	181-193 447-462	Confirmed
P00738	Haptoglobin	VVLHPN*YSQVDIGLIK ^{UH,UL,EH,EL} MVSHHN*LTTGATLINEQWLLTTAK ^{EH,EL}	236-251 179-202	Confirmed Confirmed
P04196	Histidine rich glycoprotein	VEN*TTVYYLVLDVQESDCSVLSR ^{EH} VIDFN*CTTSSVSSALANTK ^{EH,EL}	61-83 121-139	Confirmed Confirmed
P05155	Plasma protease C1 inhibitor	VGQLQLSHN*LSLVILVPQNLK ^{EH}	344-364	Confirmed
P01857	Ig alpha-1 chain C region	LSLHRPALEDLLGSEAN*LTCTLTGLR ^{EH}	127-153	Potential
P01859	Ig gamma-2 chain C region	TKPREEQFN*STFR ^{EH} TPLTAN*ITK ^{EH}	168-180 200-208	Confirmed Potential
P01860	Ig gamma-1 chain C region	EEQYN*STYR ^{UH,EL}	136-144	Potential
P01861	Ig gamma-4 chain C region	EEQFN*STYR ^{UH,EH,EL}	173-181	Potential
P01591	Immunoglobulin J chain	EN*ISDPTSPLR ^{EH,EL}	48-58	Confirmed
P40189	Interleukin-6 receptor beta	QQYFKQN*CSQHESSPDISHFER ^{EL}	812-833	Potential
P56199	Integrin alpha-1	SYFSSLN*LTIR ^{EL}	1096-1106	Potential
P29622	Kallistatin	DFYVDEN*TTVR ^{EH}	232-242	Confirmed
P01042	Kininogen-1	LNAENN*ATFYFK ^{UH,EL}	389-400	Potential
P11279	Lysosome-associated membrane glycoprotein 1	DPAFKAAN*GSLR ^{EL}	314-325	Confirmed
P01871	Ig mu chain C region	YKN*NSDISSTR ^{EH} GLTFQQN*ASSMCVDPQDPAIR ^{EL}	44-54 204-223	Confirmed Confirmed
Q9HC10	Otoferlin	NEMLEIQVFN*YSKVFNSK ^{UH,UL,EL}	59-76	Potential
Q5VU65	Nuclear pore membrane glycoprotein 210-like	EVVVN*ASSR ^{EH}	1551-1559	Potential
Q9Y5E7	Protocadherin beta 2	ETRSEYN*ITITVDFGTGR ^{EL}	414-432	Potential
P36955	Pigment epithelium derived factor	VTQN*LTLIEESLTSEFIHDIDR ^{EH}	282-303	Confirmed
P27169	Serum paraoxanase/aryl esterase1	HAN*WTLTPLK ^{EH}	250-259	Confirmed
P49908	Selenoprotein P	EGYSN*ISYIVVNHQGISSR ^{EH}	79-97	Confirmed

Q9Y275	Tumor necrosis factor ligand superfamily member 13B	CIQNMPETLPN*NSCYSAGIAK ^{EL}	232-252	Confirmed
P40225	Thrombopoietin	IHELLN*GTRGLFPGPSRR ^{EL}	250-267	Confirmed
P02787	Serotransferrin	QQQHFLFGSN*VTDCSGNFCLFR ^{EH,EL}	622-642	Confirmed
P07996	Thrombospondin 1	VVN*STTGPGEHLR ^{EH,EL}	1065-1077	Confirmed
P04004	Vitronectin	NN*ATVHEQVGGPSLTSDLQAQSK ^{EL}	85-107	Confirmed
P25311	Zinc-alpha-2-glycoprotein	DIVEYYN*DSN*GSHVLQGR ^{EH,EL}	100-117	Confirmed

These results demonstrated that ConA affinity chromatography on serum tryptic digests was effective in capturing most of the glycopeptides bearing biantennary and high mannose glycan structures, which are the epitope specifically recognized by this lectin. A further advantage of the strategy relies on performing the lectin affinity chromatography on glycopeptides instead of glycoproteins, avoiding fractionation of the extract by chromatographic and/or electrophoretic methods. Moreover, this selection allowed both the identification of the individual glycoprotein and the direct assessment of the N-glycosylation site during the same analysis. Finally, the affinity procedure was instrumental for the detection of the less abundant glycopeptides together with the most abundant ones, such as those deriving from albumins or immunoglobulins.

Results obtained by LC-MS/MS analyses revealed a quite similar profile for myocarditis and control samples, in the same that most of the proteins were identified in both the samples. However some individual glycopeptides were only found in one sample; in a few cases the same glycopeptide was recovered in the ConA bound fraction in one sample and in the unbound fraction in the other. This suggests that small differences can be appreciated in the glycosylation profile of glycoproteins from healthy or pathological sera with these differences very likely confined to the branching structures of the glycan moieties.

To reduce the complexity of the whole sample, in alternative to ConA affinity chromatography Boronate affinity purification can be rapidly performed in batch after tryptic digestion. The principle of purification is based on the vicinal diols binding capacity, so that no discrimination on the basis of the glycan type was performed thus, theoretically, all glycopeptides could be selected. The recovered glycopeptides were then deglycosylated by PNGase F treatment and the peptide mixtures directly analysed by LC-MS/MS.

Even in this case as for ConA the analyses were performed on intact serum samples without any pre-purification step or removal of most abundant proteins. The peptide component of the eluted fraction was analysed by tandem mass spectrometry. Similar analyses were carried out on the unbound Boronate fractions, mainly containing non-glycosylated peptides.

As a whole, using Boronate affinity approach we could confirm most of the previously identified glycosylation sites based on ConcanavalinA enrichment and 5 more glycosylation sites were identified thus refining previous data on myocarditis glycoproteome.

Table 4: LC/MSMS analysis for the identification of glycosylation sites by boronate affinity chromatography (UH = unbound fraction of healthy serum; UL = unbound fraction of myocardial lesions serum; UK = unbound fraction of hepatic carcinoma serum; EH = eluted fraction of healthy serum serum; EL = eluted fraction of myocardial lesions serum; EK = eluted fraction of hepatic carcinoma serum)

	Protein	Sequence	Peptide
P01011	Alpha-1-antichymotrypsin	YTGN*ASALFILPDQDK ^{H,M}	268-283
P02763	Alpha-1-acid glycoprotein	SVQEIQATFFYFTPN*KTEDTIFLR ^{H,M} QDQCIYN*TTYLNVQR ^{H,M}	58-81 87-101
P01009	Alpha-1-antitrypsin	YLGN*ATAIFFLPDEGK ^{H,M} QLAHQSN*STNIFFSPVSIATAFAMLSLGTK ^H ADTHDEILEGLNFN*LTEIPEAQIHEGFQELLR ^{H,M}	268-283 64-93 94-125
P01023	Alpha-2-macroglobulin	SLGNVN*FTVSAEALSEQELCGTEVPSVPEHGRK ^H	864-886
P43652	Afamin	DIENFN*STQK ^{H,M} YAEDKFN*ETTEK ^H	28-37 396-407
P01008	Antithrombin III	LGACN*DTLQQLMEVFK ^{H,M} SLTFN*ETYQDISELVYGAK ^{H,M}	124-139 183-201
P04114	Apolipoprotein B-100	FN*SSYLQGTNQITGR ^{H,M} FVEGSHN*STVSLTTK ^H AEEEMLEN*VSLVCPK ^M YDFN*SSMLYSTAK ^M	1522-1536 3405-3419 27-41 3462-3474
P05090	Apolipoprotein D	ADGTVNQIEGEATPVN*LTEPAK ^M	83-104
P02749	Apolipoprotein H	VYKPSAGN*NSLYR ^H LGN*WSAMPSCK ^M	155-167 251-261
O75882	Attractin	IDSTGN*VTNELR ^{H,M} GPVKMPSQAPTGNFYQPQLLN*SSMCLEDSR ^H	411-422 1023-1052
P00450	Ceruloplasmin	EHEGAIYPDN*TTDFQR ^{H,M} EN*LTAPGSDSAVFFEQGTTR ^{H,M}	129-144 396-414
P08603	Complement factor H	ISEEN*ETTCYMGK ^H MDGASN*VTCINSR ^{H,M} IPCSQPPQIEHGTTIN*SSR ^M SPDVIN*GSPISQK ^H	907-919 1024-1036 868-885 212-224
P10909	Clusterin	LAN*LTQGEDQYYLR ^{H,M}	372-385
Q81WV2	Contactin-4	LN*GTDVDTGMDFR ^M	64-76
P05156	Complement factor 1	FLNN*GTCTAEGK ^{H,M}	100-111
P01024	Complement C3	TVLTPATNHMGN*VTFIPANR ^H	74-94
P0C0L4	Complement C4-A	GLN*VTLSSSTGR ^{H,M}	1326-1336
P02748	Complement component C9	AVN*ITSENLIDDVVSLIR ^{H,M}	413-430
Q14517	Protocadherin-fat 1	QVYN*LTVRAKDK ^M FSMDYKGTGALTVQN*TTQLRSR ^M	994-1005 1930-1950
P02765	Alpha-2-HS-glycoprotein	VCQDCPLLAPLN*DTR ^{H,M} AALAAFNAQNN*GSNFQLEEISR ^H	145-159 166-187
Q03591	Complement factor H related protein 1	LQNNENN*ISCVET ^H	120-132
Q13439	Golgin subfamily A member 4	HN*STLKQLMREFNTQLAQK ^H	1990-2008
P02790	Hemopexin	SWPAVGN*CSSALR ^{H,M} ALPQPQN*VTSLLGCTH ^{H,M}	181-193 447-462
P00738	Haptoglobin	VVLHPN*YSQVDIGLIK ^{H,M} MVSHHN*LTGATLINEQWLLTTAK ^{H,M} NLFLN*HSEN*ATAK ^{H,M}	236-251 179-202 203-215
P04196	Histidine rich glycoprotein	VEN*TTVYYLVLDVQESDCSVLSR ^H VIDFN*CTTSSVSSALANTK ^{H,M}	61-83 121-139
P05155	Plasma protease C1 inhibitor	VGQLQLSHN*LSLVILVPQNLK ^{H,M}	344-364
P01857	Ig alpha-1 chain C region	LSLHRPALEDLLGSEAN*LTCTLTGLR ^H	127-153
P01859	Ig gamma-2 chain C region	TKPREEQFN*STFR ^H TPLTAN*ITK ^H	168-180 200-208
P01860	Ig gamma-1 chain C region	EEQYN*STYR ^{H,M}	136-144
P01591	Immunoglobulin J chain	EN*ISDPTSPLR ^{H,M}	48-58
P40189	Interleukin-6 receptor beta	QQYFKQN*CSQHESPDISHFER ^M	812-833
P56199	Integrin alpha-1	SYFSSLN*LTIR ^M	1096-1106
P29622	Kallistatin	DFYVDEN*TTVR ^H	232-242
P01042	Kininogen-1	LNAENN*ATFYFK ^{H,M}	389-400
P11279	Lysosome-associated membrane glycoprotein 1	DPAFKAAN*GSLR ^M	314-325
P01871	Ig mu chain C region	YKN*NSDISSTR ^H GLTFQQN*ASSMCVDPQDTAIR ^M	44-54 204-223
Q9HC10	Otoferlin	NEMLEIQVFN*YSKVFSNK ^{H,M}	59-76
Q5VU65	Nuclear pore membrane glycoprotein 210-like	EVVVN*ASSR ^H	1551-1559

Q9Y5E7	Protocadherin beta 2	ETRSEYN*ITITVDFGTPR ^M	414-432
P36955	Pigment epithelium derived factor	VTQN*LTIEESLTSEFIHDIR ^H	282-303
P27169	Serum paraoxanase/arylesterase1	HAN*WTLTPLK ^H	250-259
P49908	Selenoprotein P	EGYSN*ISYIVVNHQGISSR ^H	79-97
Q9Y275	Tumor necrosis factor ligand superfamily member 13B	CIQNPETLPN*NSCYSAGIAK ^M	232-252
P40225	Thrombopoietin	IHELLN*GTRGLFPGPSRR ^M	250-267
P02787	Serotransferrin	QQQHLFGSN*VTDCSGNFCLFR ^{H,M} CGLVPVLAENYN*KSDNCEDTPEAGYFAVAVVK ^M	622-642 421-452
P07996	Thrombospondin 1	VVN*STTGPGELHR ^{H,M}	1065-1077
P04004	Vitronectin	NN*ATVHEQVGGPSLTSDLQAQSK ^M	85-107
P25311	Zinc-alpha-2-glycoprotein	DIVEYYN*DSN*GSHVLQGR ^{H,M}	100-117
P02766	Transthyretin	ALGISPFHEHAEVVFTAN*DSGPR ^H	101-123

Glycan profiling

The glycan profiling of the two sera samples was performed in a further attempt to investigate differences in N-glycosylation possibly occurring between myocarditis and healthy donors sera. Several papers have demonstrated an association between N-glycosylation changes in the serum glycome and cancer, with changes concerning the levels of highly branched glycan structures, sialyl Lewis X epitopes and agalactosylated bi-antennary glycans. Moreover, several of these glycans have been linked to chronic inflammatory diseases, thus leading to the hypothesis of possible connections between inflammation and cancer [26].

Samples were subjected to the ConA affinity selection step, desalted by HPLC to remove salts and methyl- α -D-mannopyranoside, then the entire set of oligosaccharides was released from glycopeptides occurring in both the Con A bound and unbound fractions by incubation with PNGase F to isolate high mannose, hybrid and biantennary structures as well as higher branched glycans.

Oligosaccharides released from enzymatic hydrolysis were desalted by in batch ion exchange chromatography on a Dowex mixed bed resin and directly analyzed by MALDI-MS on a reflectron instrument, both in positive and negative mode. Since significant desialylation might occur during analysis of the samples, MALDI mass spectra were also performed using THAP as matrix that was reported to avoid most of the problems associated with loss of sialic acids by fragmentation [27].

Each peak detected in the MALDI spectra was tentatively attributed to a glycan structure based on the knowledge of N-glycan biosynthetic pathway and literature data. Manual interpretation was confirmed by using ExPASy Glycomod software, available on line for the prediction of possible oligosaccharidic structures from mass values. Selected mass signals were also fragmented by MALDI PSD experiments to confirm glycan structures and to discriminate between isobaric components. As a whole, these MALDI analyses constituted a real oligosaccharide profiling of the samples allowing us to describe the microheterogeneity of glycosylation within the different serum samples in details and to identify specific changes in glycan structures occurring in the pathological sera.

Fig.7 shows the positive MALDI spectra of glycans deriving from the healthy (A) and myocarditis (B) glycopeptides retained on the ConA affinity step and incubated with PNGase F. The oligosaccharides composition and structures corresponding to the recorded mass signals are summarised in Table 5. More than fifty structures were detected in the MALDI profiling, essentially belonging to high mannose and biantennary type, thus indicating a high degree of glycan heterogeneity. These data indicated that the two ConA bound fractions essentially consisted of peptides carrying biantennary and hybrid as well as high mannose structures. This finding is in perfect agreement with the reported affinity of ConA that generally binds complex

biantennary, hybrid type and high-mannose-type N-linked glycans. The MALDI profiling revealed the occurrence of only three triantennary structures, at m/z 2028.7, 2174.8 and 2320.8 respectively that were attributed to HexNAc₅Hex₆, FucHexNAc₅Hex₆ and HexNAc₅Hex₆Sia glycans. PSD analysis performed on the precursor ion at m/z 2174.8 localised the fucose residue on the 3-antenna. In the fucosylated triantennary structure the position of the fucose residue plays a very important biological role. As an example, alpha-1-acid glycoprotein, one of the identified glycoproteins in this study, is a non-specific marker, whose temporary increase in the serum level is generally associated with many acute phase syndromes (infections, inflammation, cancer)[28]. Previous data reported that the fucose residue on the triantennary glycans in alpha-1-acid glycoprotein was generally found on the 3-antenna and not on the terminal core GlcNAc [29,30].

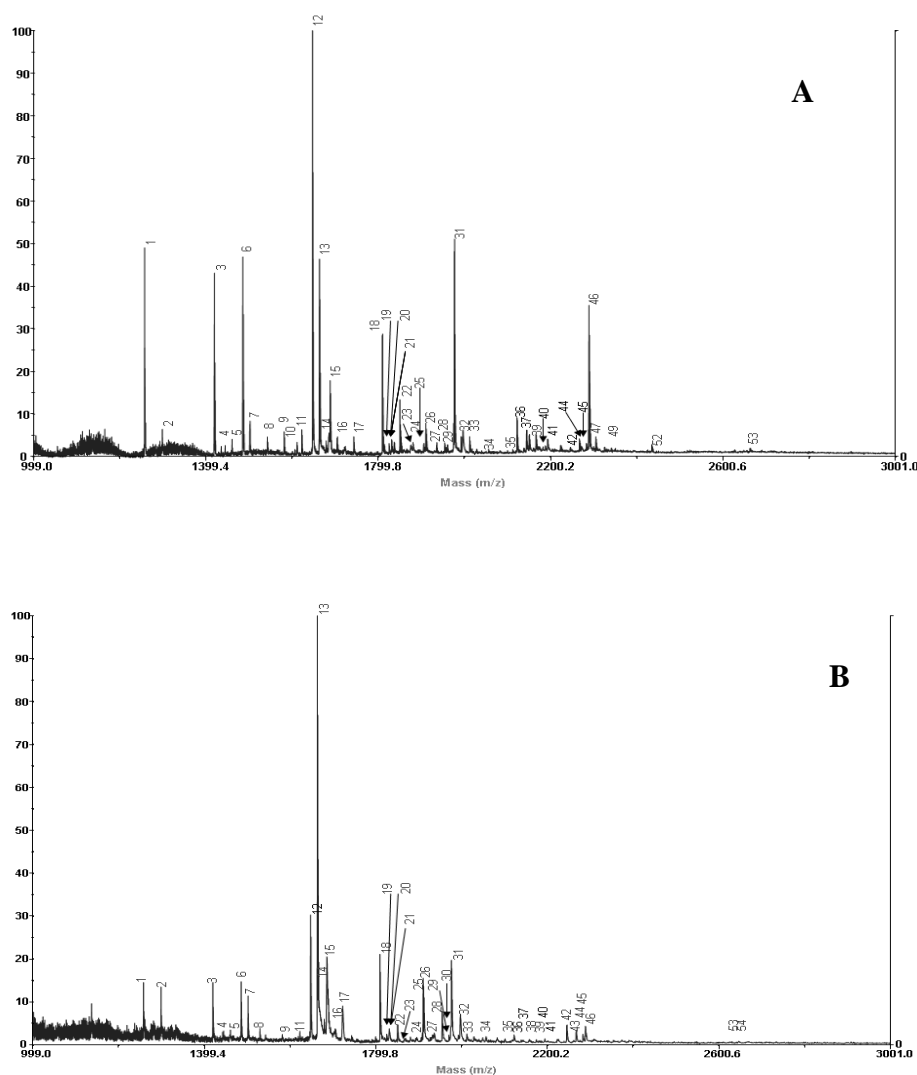


Fig. 7: Positive ion mode spectra of glycans deriving from ConA bound fractions of healthy (panel A) and myocarditis sera samples (panel B)

Table 5: Glycans detected by MALDI MS analysis of the glycans from ConA eluates of the different sera samples (H = healthy serum; L = myocardial lesions; K = hepatic carcinoma)

MALDI (m/z [M+Na] ⁺)			Monosaccharides composition	Structure		PSD
	Calculated	Found				
1	1257.4	1257.4 ^H 1257.3 ^L	HexNAc ₂ Hex ₅	High mannose		X
2	1298.5	1298.4 ^H 1298.3 ^L	HexNAc ₃ Hex ₄	Complex, monoantennary		X
3	1419.5	1419.5 ^H 1419.3 ^L	HexNAc ₂ Hex ₆	High mannose		X
4	1444.5	1444.5 ^H 1444.4 ^L	FucHexNAc ₃ Hex ₄	Complex, fucosylated		X
5	1460.5	1460.4 ^H 1460.4 ^L	HexNAc ₃ Hex ₅	Hybrid		
6	1485.5	1485.4 ^H 1485.4 ^L	FucHexNAc ₄ Hex ₃	Complex, biantennary, fucosylated		X
7	1501.5	1501.5 ^H 1501.3 ^L	HexNAc ₄ Hex ₄	Complex, biantennary		X
8	1542.6	1542.4 ^H 1542.4 ^L	HexNAc ₅ Hex ₃	Complex, bisecting		X
9	1581.5	1581.5 ^H 1581.4 ^L	HexNAc ₂ Hex ₇	High mannose		
10	1611.5	1611.4 ^H	HexNAc ₃ Hex ₄ SiaNa	Complex, monoantennary, sialylated		
11	1622.5	1622.5 ^H 1622.4 ^L	HexNAc ₃ Hex ₆	Hybrid		
12	1647.6	1647.5 ^H 1647.3 ^L	FucHexNAc ₄ Hex ₄	Complex, biantennary, fucosylated		X
13	1663.6	1663.5 ^H 1663.3 ^L	HexNAc ₄ Hex ₅	Complex, biantennary		X
14	1685.6	1685.8 ^H 1685.4 ^L	HexNAc ₄ Hex ₅ Na	Complex, biantennary	Calculated	
15	1688.6	1688.5 ^H 1688.3 ^L	FucHexNAc ₅ Hex ₃	Complex, fucosylated, bisecting		X
16	1704.6	1705.5 ^H	HexNAc ₅ Hex ₄	Complex, bisecting		
17	1743.6	1743.4 ^H	HexNAc ₂ Hex ₈	High mannose		
18	1809.6	1809.5 ^H 1809.3 ^L	FucHexNAc ₄ Hex ₅	Complex, biantennary, fucosylated		X
19	1814.6	1815.4 ^H 1816.3 ^L	HexNAc ₄ Hex ₄ SiaNa	Complex, biantennary, sialylated	Calculated	
20	1825.6	1825.5 ^H 1825.3 ^L	HexNAc ₄ Hex ₆	Hybrid	Calculated	
21	1832.6	1831.3 ^L	FucHexNAc ₄ Hex ₅ Na	Complex, fucosylated, biantennary	Calculated	
22	1850.7	1850.5 ^H 1850.3 ^L	FucHexNAc ₅ Hex ₄	Complex, fucosylated, bisecting		
23	1866.7	1866.5 ^H 1866.4 ^L	HexNAc ₅ Hex ₅	Complex, bisecting		
24	1881.6	1880.8 ^H 1880.5 ^L	Fuc ₂ HexNAc ₃ Hex ₄ Sia	Complex, difucosylated, sialylated, monoantennary	Calculated	
25	1905.6	1906.5 ^H 1905.4 ^L	HexNAc ₂ Hex ₉	High mannose		
26	1910.7	1910.5 ^H 1910.3 ^L	HexNAc ₅ Hex ₅ Na ₂	Complex, triantennary/bisecting Complex, biantennary, fucosylated, sulfated	Calculated	

27	1935.6	1936.5 ^H 1936.3 ^L	HexNAc ₃ Hex ₆ SiaNa	Hybrid, sialylated		
28	1955.7	1954.5 ^H 1954.3 ^L	HexNAc ₄ Hex ₅ Sia Fuc ₂ HexNAc ₄ Hex ₅	Complex, biantennary, sialylated Complex, biantennary, difucosylated		X
29	1960.7	1958.5 ^H 1957.3 ^L	FucHexNAc ₄ Hex ₄ SiaNa	Complex, fucosylated, sialylated, biantennary	Calculated	
30	1971.7	1970.3 ^L	FucHexNAc ₄ Hex ₆	Hybrid, fucosylated	Calculated	
31	1976.7	1975.4 ^H 1976.3 ^L	HexNAc ₄ Hex ₅ SiaNa	Complex, biantennary, sialylated		
32	1995.7 1996.7	1996.5 ^H 1997.2 ^L	HexNAc ₅ Hex ₄ Sia Fuc ₂ HexNAc ₅ Hex ₄	Sialylated, triantennary/bisecting Difucosylated, triantennary/bisecting	Calculated	
33	2012.7	2012.5 ^H 2012.3 ^L	FucHexNAc ₅ Hex ₅	Complex, fucosylated, bisecting		
34	2028.7	2029.5 ^H 2029.3 ^L	HexNAc ₅ Hex ₆	Complex, triantennary		
35	2100.7	2100.5 ^H 2101.4 ^L	FucHexNAc ₄ Hex ₅ Sia	Complex, fucosylated, sialylated, biantennary		X
36	2123.7	2122.5 ^H 2122.3 ^L	FucHexNAc ₄ Hex ₅ SiaNa	Complex, fucosylated, sialylated, biantennary	Calculated	
37	2144.7	2145.4 ^H 2144.2 ^L	FucHexNAc ₄ Hex ₅ SiaNa ₂	Complex, fucosylated, sialylated, biantennary		X
38	2157.1	2158.3 ^L	HexNAc ₅ Hex ₅ Sia	Complex, sialylated, triantennary/bisecting	Calculated	
39	2166.7	2166.3 ^H	FucHexNAc ₄ Hex ₅ SiaNa ₃	Complex, fucosylated, sialylated, biantennary		X
40	2174.8	2175.5 ^H 2175.2 ^L	FucHexNAc ₅ Hex ₆	Complex, fucosylated, triantennary		
41	2197.7	2193.6 ^H 2194.5 ^L	FucHexNAc ₅ Hex ₆ Na	Complex, fucosylated, triantennary	Calculated	
42	2245.8	2246.5 ^H 2245.3 ^L	HexNAc ₄ Hex ₅ Sia ₂	Complex, disialylated, biantennary		X
43	2262.8	2262.2 ^L	FucHexNAc ₄ Hex ₆ Sia	Hybrid, sialylated, fucosylated	Calculated	
44	2268.8	2267.5 ^H 2267.4 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, disialylated, biantennary		X
45	2284.8	2283.4 ^H 2283.2 ^L	FucHexNAc ₄ Hex ₆ SiaNa	Hybrid, fucosylated, sialylated		X
46	2290.8	2289.5 ^H 2289.3 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na ₂	Complex, disialylated, biantennary		
47	2306.8	2305.4 ^H	FucHexNAc ₄ Hex ₆ SiaNa ₂	Hybrid, fucosylated, sialylated	Calculated	
48	2325.8	2326.5 ^H	FucHexNAc ₅ Hex ₅ SiaNa	Complex, fucosylated, sialylated, triantennary/bisecting	Calculated	
49	2434.8	2435.5 ^H	FucHexNAc ₄ Hex ₅ Sia ₂ Na ₂	Complex, fucosylated, disialylated, biantennary	Calculated	
50	2627.9	2625.4 ^H 2625.5 ^L	FucHexNAc ₅ Hex ₇ Sia	Hybrid, sialylated	Calculated	
54	2649.8	2647.0 ^L	FucHexNAc ₅ Hex ₇ SiaNa	Hybrid, fucosylated, sialylated	Calculated	

Negative MALDI spectra performed on the same ConA bound fractions essentially confirmed the above reported data showing the presence of few sialylated variants, corresponding to biantennary glycans.

Higher branched glycopeptides were not retained onto the lectin originating the unbound fraction of the affinity purification. Thus, ConA affinity chromatography led to the relative concentration of tri- and tetraantennary oligosaccharides in this fraction. Without this step, these usually less abundant glycans could have been escaped detection. PNGase F was directly added to the unbound fraction, glycans were

separated from peptides by HPLC and directly analysed by MALDI-MS following the procedures described above.

Figure 8 shows the positive MALDI-MS analyses of glycans deriving from the unbound fractions of the healthy (A) and pathological (B) serum samples. Mass signals were tentatively assigned as described before and, when necessary, confirmed by PSD fragmentation experiments. Negative MALDI analyses were also performed on the same ConA unbound fractions to highlight the presence of sialylated structures (data not shown). All the data are summarised in Tables 6.

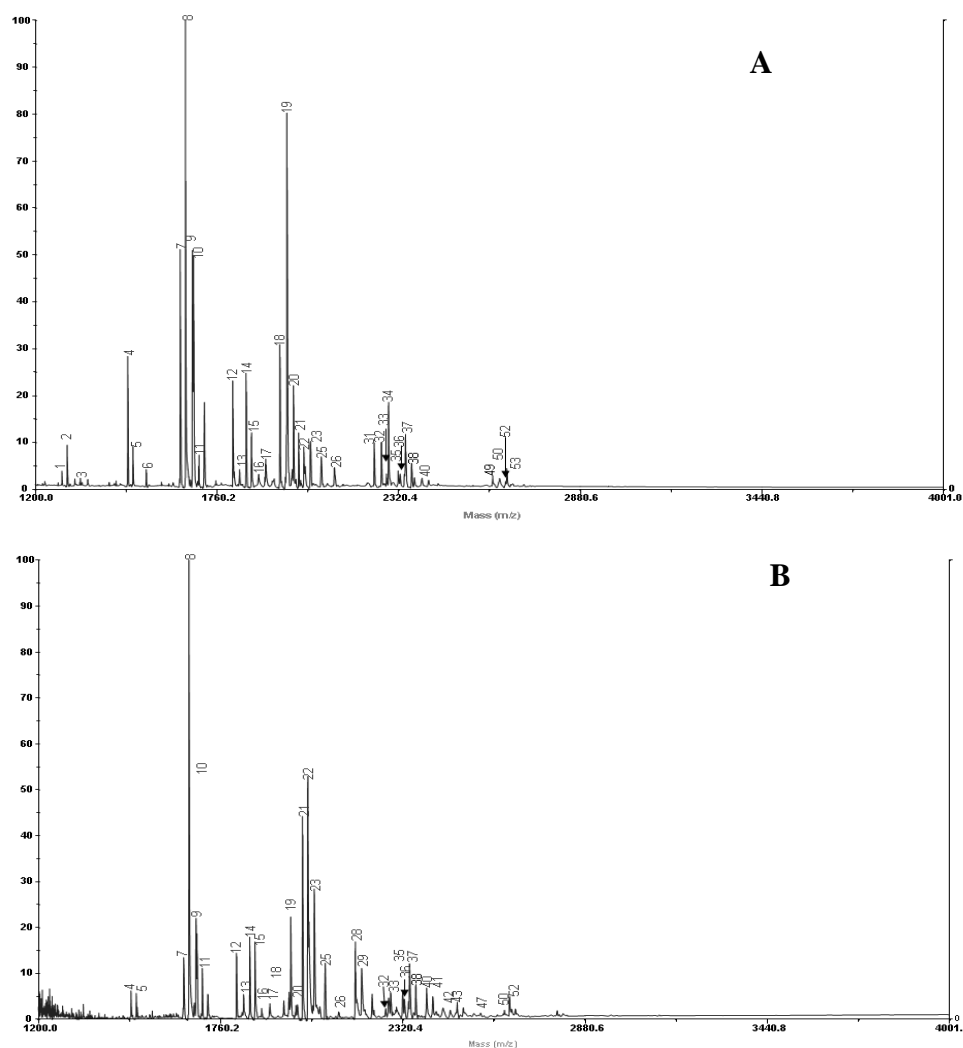




Fig. 8: Positive ion mode spectra of glycans deriving from ConA unbound fractions of healthy (panel A) and myocarditis sera samples (panel B)

Table 6: Glycans detected by MALDI MS analysis of the glycans from ConA unbound fractions of the different sera samples (H = healthy serum; L = myocardial lesions; K = hepatic carcinoma)

MALDI (m/z [M+Na] ⁺)			Monosaccharides composition	Structure	PSD	
	Calculated	Found				
1	1282.4	1282.1 ^H	FucHexNAC ₃ Hex ₃	Complex, fucosylated, monoantennary		
2	1298.5	1298.1 ^H	HexNAC ₃ Hex ₄	Complex, monoantennary		X

3	1339.5	1339.1 ^H	HexNAc ₄ Hex ₃	Complex, biantennary		
4	1485.5	1485.1 ^H 1485.0 ^L	FucHexNAc ₄ Hex ₃	Complex, fucosylated, biantennary		X
5	1501.5	1501.0 ^H 1500.9 ^L	HexNAc ₄ Hex ₄	Complex, biantennary		X
6	1542.6	1542.0 ^H	HexNAc ₅ Hex ₃	Complex, bisecting		X
7	1647.6	1647.0 ^H 1646.9 ^L	FucHexNAc ₄ Hex ₄	Complex, fucosylated, biantennary		X
8	1663.6	1663.0 ^H 1662.9 ^L	HexNAc ₄ Hex ₅	Complex, biantennary		X
9	1685.6	1684.6 ^H 1685.1 ^L	HexNAc ₄ Hex ₅ Na	Complex, biantennary	Calculated	
10	1688.6	1688.0 ^H	FucHexNAc ₅ Hex ₃	Complex, fucosylated, bisecting		
11	1704.6	1705.0 ^H 1703.9 ^L	HexNAc ₅ Hex ₄	Complex, bisecting		
12	1809.6	1809.0 ^H 1809.9 ^L	FucHexNAc ₄ Hex ₅	Complex, biantennary, fucosylated		X
13	1831.6	1830.1 ^H 1830.8 ^L	FucHexNAc ₄ Hex ₅ Na	Complex, fucosylated, biantennary		X
14	1850.7	1850.0 ^H 1849.9 ^L	FucHexNAc ₅ Hex ₄	Complex, fucosylated, bisecting		
15	1866.7	1867.2 ^H 1865.8 ^L	HexNAc ₅ Hex ₅	Complex, bisecting		
16	1889.7	1889.2 ^H 1886.4 ^L	HexNAc ₅ Hex ₅ Na	Complex, triantennary/bisecting	Calculated	
17	1913.7	1912.0 ^H 1911.8 ^L	HexNAc ₃ Hex ₆ Sia	Hybrid, sialylated		X
18	1955.7	1954.0 ^H 1953.8 ^L	HexNAc ₄ Hex ₅ Sia Fuc ₂ HexNAc ₄ Hex ₅	Complex, sialylated, biantennary Complex, biantennary, difucosylated		X
19	1977.7	1975.9 ^H 1975.8 ^L	HexNAc ₄ Hex ₅ SiaNa	Complex, sialylated, biantennary		X
20	1995.7 1996.7	1996.0 ^H 1996.8 ^L	HexNAc ₅ Hex ₄ Sia Fuc ₂ HexNAc ₅ Hex ₄	Complex, sialylated, triantennary/bisecting Complex, difucosylated, triantennary/bisecting	Calculated	
21	2012.7	2011.9 ^H 2011.8 ^L	FucHexNAc ₅ Hex ₅	Complex, fucosylated, bisecting		
22	2028.7	2028.0 ^H 2028.7 ^L	HexNAc ₅ Hex ₆	Complex, triantennary		X
23	2050.7	2049.9 ^H 2047.8 ^L	HexNAc ₅ Hex ₆ Na	Complex, triantennary	Calculated	
24	2081.6	2081.3 ^H 2081.7 ^L	FucHexNAc ₃ Hex ₆ Sia	Hybrid, fucosylated, sialylated	Calculated	
25	2123.7	2121.9 ^H	FucHexNAc ₄ Hex ₅ SiaNa	Complex, fucosylated, sialylated, biantennary		X
26	2174.8	2174.7 ^L	FucHexNAc ₅ Hex ₆	Complex, fucosylated, triantennary		
27	2179.8	2179.6 ^L	HexNAc ₅ Hex ₅ SiaNa	Complex, sialylated, triantennary/bisecting	Calculated	
28	2245.8	2244.9 ^H	HexNAc ₄ Hex ₅ Sia ₂	Complex, disialylated, biantennary		X
29	2267.8	2266.9 ^H 2267.5 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, disialylated, biantennary		X
30	2284.8	2282.8 ^H 2283.7 ^L	FucHexNAc ₄ Hex ₆ SiaNa	Hybrid, fucosylated, sialylated	Calculated	
31	2290.8	2288.9 ^H	HexNAc ₄ Hex ₅ Sia ₂ Na ₂	Complex, disialylated, biantennary		
32	2320.8	2318.8 ^H 2319.7 ^L	HexNAc ₅ Hex ₆ Sia Fuc ₂ HexNAc ₅ Hex ₆	Complex, sialylated, triantennary Complex, difucosylated, triantennary		X
33	2325.8	2324.9 ^H	FucHexNAc ₅ Hex ₅ SiaNa	Complex, fucosylated,	Calculated	

		2324.7 ^L		sialylated, triantennary/bisecting		
34	2342.8	2341.8 ^H 2340.7 ^L	HexNAc ₅ Hex ₆ SiaNa	Complex, sialylated, triantennary	Calculated	
35	2360.9	2359.9 ^H 2359.8 ^L	HexNAc ₆ Hex ₅ Sia	Complex, sialylated, tetraantennary	Calculated	
36	2393.7	2392.9 ^H 2392.6 ^L	HexNAc ₆ Hex ₇	Complex, tetraantennary		X
37	2413.8	2412.3 ^L	FucHexNAc ₄ Hex ₅ Sia ₂ Na	Complex, fucosylated, disialylated, biantennary	Calculated	
38	2466.8	2465.6 ^L	FucHexNAc ₅ Hex ₆ Sia	Complex, fucosylated, sialylated, triantennary	Calculated	
39	2488.8	2487.7 ^L	FucHexNAc ₅ Hex ₆ SiaNa	Complex, fucosylated, sialylated, triantennary	Calculated	
40	2539.9	2539.4 ^L	FucHexNAc ₆ Hex ₇	Complex, fucosylated, tetraantennary		
41	2611.9	2610.7 ^H	HexNAc ₅ Hex ₆ Sia ₂	Complex, disialylated, triantennary	Calculated	
42	2633.9	2632.8 ^H 2632.6 ^L	HexNAc ₅ Hex ₆ Sia ₂ Na	Complex, disialylated, triantennary	Calculated	
43	2651.9	2649.7 ^H 2648.6 ^L	HexNAc ₆ Hex ₅ Sia ₂	Complex, disialylated, tetraantennary	Calculated	
44	2654.9	2653.8 ^H	HexNAc ₅ Hex ₆ Sia ₂ Na ₂	Complex, disialylated, triantennary	Calculated	

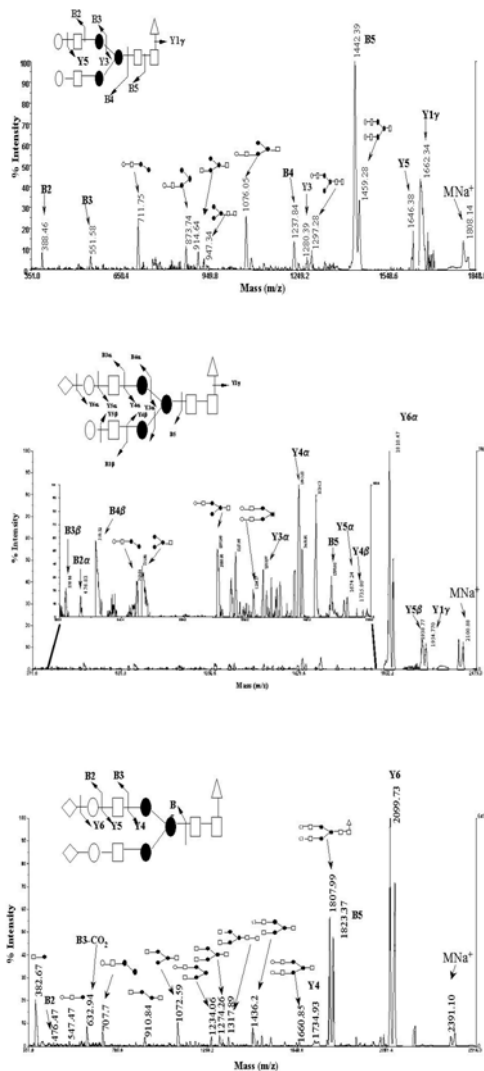


Fig. 9: PSD analysis of the non sialylated (panel A), monosialylated (panel B) and disialylated biantennary structure (panel C).

Compared to the glycans of the bound fraction, these oligosaccharides showed the presence of higher branched, sialylated and fucosylated structures. Table 6 shows the presence of essentially sialylated triantennary glycans with the occurrence of some fucosylated and sialylated tetraantennary structures and few biantennary oligosaccharides. When peaks could not be unambiguously interpreted, the ultimate definition of the correct glycan structure was achieved by tandem mass spectrometry experiments using the MALDI PSD technique. As an example, the signal at m/z 2392.8 in the MALDI spectrum was tentatively assigned to two different isobaric structures. Following PSD fragmentation, the correct biantennary sialylated and fucosylated glycoform could be inferred.

As shown in Tables 5 and 6, the microheterogeneity of glycans detected in the healthy sera resulted to be in general very similar to the glycan profiling observed for the myocardial lesions serum. However, a number of differences could be detected in the high mass region of the spectra and particularly in the unbound fraction of the myocarditis sample with respect to the healthy glycans. Myocardial glycan profiling showed the occurrence of peculiar structures, the persistence of incomplete and truncated structures, the accumulation of precursors and the appearance of novel structures, thus leading to a higher degree of heterogeneity, with respect to the

healthy one. In particular, the myocardial lesions serum revealed the presence of hyper-fucosylated oligosaccharides. It has been demonstrated that the N-linked glycans attached to proteins generally associated with inflammation, like haptoglobin, alpha1-acid glycoprotein and a1-antichymotrypsin, have increased SLex epitopes, during chronic inflammation.

It is clear that this work is by far a non-exhaustive analysis of a system as complex as serum. A complete profiling of the glycosylation in the different sera was accomplished, but a detailed quantitative evaluation of all the oligosaccharide structures detected was not performed. However, preliminary quantitative speculations can be made considering the most abundant glycan structures observed in the myocarditis sera in comparison with the healthy ones. The analysis performed on the ConA eluted glycan fractions showed a quite similar amount of the different glycan structures in both the pathological samples. However, the relative abundance of the healthy oligosaccharides resulted to be quite 5 times higher. On the contrary, the same analysis carried out on the unbound ConA fractions revealed a higher level of heterogeneity within the relative abundance of the glycan structures. In fact, some species resulted to be expressed at lower level with respect to the healthy sample; other structures resulted to be over expressed. More interestingly, the myocarditis samples showed the occurrence of structures completely absent in the reference sample. MALDI-MS analysis of glycans deriving from the unbound fraction allowed the detection of plethora of different glycans, together with those identified in the eluted fraction. The glycan profiling of myocarditis samples was characterized by abnormal glycosylation patterns, taking into account branching, sialylation and fucosylation, presenting an excessive expression of peculiar structures, the persistence of incomplete and truncated structures, the accumulation of precursors and the appearance of novel structures, thus leading to a higher degree of heterogeneity, with respect to the healthy one. In details, the onset of this disease seems to be connected with the presence of hyper-fucosylated signals. These data are in agreement with previous data. It has been demonstrated, in fact, that the N-linked glycans attached to proteins generally linked to inflammation like haptoglobin, a1-acid glycoprotein and a1-antichymotrypsin have increased SLex epitopes in cancer and again, more generally, during chronic inflammation [31].

III.4. Conclusions

In biomedical applications, a comparative approach is usually employed to identify proteins that are up and down regulated in a disease specific manner for use as diagnostic markers or therapeutic targets.

This work is not only concentrated on the study of glycosylation but it can be considered an overview of the investigation at molecular level of myocarditis by using a proteomic approach. Serum proteins and glycoproteins and free peptides occurring in human sera from healthy donors were compared to the ones from myocarditis patients. The high resolution, the sensitivity and the reproducibility of the used techniques led to the identification of some up regulated proteins in the serum from a myocarditis affected patient, all these proteins are connected to inflammatory events and one in particular (hemopexin) opens the way to new speculations in serum proteins as a specific marker for pathologic state. Mass spectrometry techniques allowed the detection of some very weakly represented free peptides, whose presence might be connected to the pathology itself. This procedure, allowed the identification of several N-glycosylation sites, contemporarily allowing the profiling and characterization of several oligosaccharides once released from peptides.

Finally, this proteomic and glycometabolic approach represents a new opportunity for therapeutics and early diagnostics, for the screening of proteic biomarkers in pathological status. Finding a biomarker molecule that precisely indicates certain kind of pathology, is something quite difficult to achieve since it requires a huge background in many different fields of clinical investigation.

III.5. References

- [1] Rezkalla SH, Kloner RA. Influenza-related viral myocarditis. *WMJ*. 2010 Aug;109(4):209-13
- [2] Blauwet LA, Cooper LT. Myocarditis. *Prog Cardiovasc Dis*. 2010 Jan-Feb;52(4):274-88.
- [3] Cihakova D, Rose NR. Pathogenesis of myocarditis and dilated cardiomyopathy. *Adv Immunol*. 2010 8;99:95-114.
- [4] Leuschner F, Katus HA, Kaya Z. Autoimmune myocarditis: past, present and future. *J Autoimmun*. 2009 Nov-Dec;33(3-4):282-9
- [5] Kühl U, Schultheiss HP. Viral myocarditis: diagnosis, aetiology and management. *Drugs*. 2009 Jul 9;69(10):1287-302
- [6] Kühl U, Schultheiss HP. Myocarditis in children. *Heart Fail Clin*. 2010 Oct;6(4):483-96.
- [7] Taylor CL, Eckart RE. Chest pain, ST elevation, and positive cardiac enzymes in an austere environment: Differentiating smallpox vaccination-mediated myocarditis and acute coronary syndrome in operation Iraqi freedom. *J Emerg Med*. 2009 Jan 30.
- [8] Lv H, Havari E, Pinto S, Gottumukkala RV, Cornivelli L, Raddassi K, Matsui T, Rosenzweig A, Bronson RT, Smith R, Fletcher AL, Turley SJ, Wucherpfennig K, Kyewski B, Lipes MA. Impaired thymic tolerance to α -myosin directs autoimmunity to the heart in mice and humans. *J Clin Invest*. 2011 Mar 23.
- [9] Stensaeth KH, Hoffmann P, Fossum E, Mangschau A, Sandvik L, Klow NE. Cardiac magnetic resonance visualizes acute and chronic myocardial injuries in myocarditis. *Int J Cardiovasc Imaging*. 2011 Feb 24.
- [10] Taylor-Papadimitriou J, Epenetos AA. Exploiting altered glycosylation patterns in cancer: progress and challenges in diagnosis and therapy. *Trends Biotechnol*. 12, 227-33 (1994).
- [11] Amado F, Lobo MJ, Domingues P, Duarte JA, Vitorino R. Salivary peptidomics. *Expert Rev Proteomics*. 2010 Oct;7(5):709-21
- [12] Menschaert G, Vandekerckhove TT, Baggerman G, Schoofs L, Luyten W, VanCriekinge W. Peptidomics coming of age: a review of contributions from Bioinformatics angle. *J Proteome Res*. 2010 May 7;9(5):2051-61.
- [13] Ludwig M. Are neuropeptides brain hormones? *J Neuroendocrinol*. 2011 Apr;23(4):381-2.
- [14] Colgrave ML, Xi L, Lehnert SA, Flatscher-Bader T, Wadensten H, Nilsson A, Andren PE, Wijffels G. Neuropeptide profiling of the bovine hypothalamus: Thermal stabilization is an effective tool in inhibiting post-mortem degradation. *Proteomics*. 2011 Apr;11(7):1264-76.
- [15] Gorman PM, Yip CM, Fraser PE, Chakrabarty A. Alternate aggregation pathways of the Alzheimer beta-amyloid peptide: A beta association kinetics at endosomal pH. *J Mol Biol*. 2003 Jan 24;325(4):743-57.

- [16] van den Broek I, Sparidans RW, Schellens JH, Beijnen JH. Sensitive liquid chromatography/tandem mass spectrometry assay for absolute quantification of ITIH4-derived putative biomarker peptides in clinical serum samples. *Rapid Commun Mass Spectrom*. 2010 Jul 15;24(13):1842-50.
- [17] ter Weeme M, Vonk AB, Kupreishvili K, van Ham M, Zeerleder S, Wouters D, Stooker W, Eijnsman L, Van Hinsbergh VW, Krijnen PA, Niessen HW. Activated complement is more extensively present in diseased aortic valves than naturally occurring complement inhibitors: a sign of ongoing inflammation. *Eur J Clin Invest*. 2010 Jan;40(1):4-10.
- [18] Dooley H, Buckingham EB, Criscitiello MF, Flajnik MF. Emergence of the acute-phase protein hemopexin in jawed vertebrates. *Mol Immunol*. 2010 Nov-Dec;48(1-3):147-23 52.
- [19] Adamsson Eryd S, Smith JG, Melander O, Hedblad B, Engström G. Inflammation-sensitive proteins and risk of atrial fibrillation: a population-based cohort study. *Eur J Epidemiol*. 2011 Mar 19 18;144(4):539-50.
- [20] Onat A, Can G, Rezvani R, Cianflone K. Complement C3 and cleavage products in cardiometabolic risk. *Clin Chim Acta*. 2011 Mar 23.
- [21] Kolte D, Bryant J, Holsworth D, Wang J, Akbari P, Gibson G, Shariat-Madar Z. Biochemical characterization of a novel high-affinity and specific plasma kallikrein inhibitor. *Br J Pharmacol*. 2011 Apr;162(7):1639-49.
- [22] Mauk MR, Smith A, Grant Mauk A. An alternative view of the proposed alternative activities of hemopexin. *Protein Sci*. 2011 Mar 14.
- [23] Larsen R, Gozzelino R, Jeney V, Tokaji L, Bozza FA, Japiassú AM, Bonaparte D, Cavalcante MM, Chora A, Ferreira A, Marguti I, Cardoso S, Sepúlveda N, Smith A, Soares MP. A central role for free heme in the pathogenesis of severe sepsis. *1 Sci Transl Med*. 2010 Sep 29;2(51):51.
- [25] Wakankar AA, Borchardt RT. Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. *J Pharm Sci*. 2006 Nov;95(11):2321-36.
- [26] Arnold JN, Saldova R., Abd Hamid UM, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics*. 8, 3284-3293 (2008)
- [27] Harvey DJ, Mass matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates, *Spectrometry Reviews*. 18, 349- 451 (1999)
- [28] T. Kremmer, É. Szöllösi, M. Boldizsár, B. Vincze, K. Ludányi, T. Imre, G. Schlosser and K. Vékey Liquid chromatographic and mass spectrometric analysis of human serum acid alpha-1-glycoprotein. *Biomed Chromatogr*. 2004 Jun;18(5):323-9
- [29] Van Dijk W, Brinkman-Vander Linden ECM, Navenaar EC. Glycosylation of α -1-acid glycoprotein (orosomucoid) in healthy and disease: occurrence, regulation and possible functional implications. *TIGG*. 1998;10:235–245
- [30] Hashimoto S, Asao T, Takahashi J, Yagihashi Y, Nishimura T, Saniabadi AR, Poland DC, van Dijk W, Kuwano H, Kochibe N, Yazawa S. α 1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis. *Cancer*. 2004 Dec 15;101(12):2825-36
- [31] Arnold JN, Saldova R., Abd Hamid UM, Rudd PM. Evaluation of the serum n-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics*. 8, 3284-3293 (2008)

IV. *Arabidopsis thaliana* glycoproteome characterization by integrated mass spectrometry techniques

IV.1. Introduction

The increasing demand for recombinant proteins, in particular for the pharmaceutical industry, has encouraged the search for novel expression systems, to overcome the low production capacities of the mammalian ones. Plants can represent a valuable alternative to mammalian and microbial expression systems for their scalability and low production costs, as well as for their ability to produce perfectly folded mammalian proteins [1]. Recently, biotechnological relevant proteins, such as monoclonal antibodies, vaccine antigens, blood proteins, cytokines, interleukins, growth factors and growth hormones, have benefited from the use of plant-based production platforms [2,3,4], essentially based on “molecular farming”. These platforms rely on several techniques including whole-plants, suspension cells, hairy roots, moss, duckweed, microalgae, etc. Plant cell cultures are prepared from the callus, an undifferentiated tissue, forming single cells and small aggregates, able to grow in shaker flasks or fermenters [5]. An alternative is constituted by field-grown crops, producing high quantities of recombinant proteins for human use, at the competitive costs of agricultural production and without the risk of human and animal pathogens contamination, leading to safer products [6]. Seed-specific expression is becoming an attractive alternative to conventional leaf-based platforms, due to the protecting function of seed environment, allowing long time storage at room temperature of the protein products accumulated in the small volume of seeds [7].

Nevertheless, even if plants expression systems can provide very homogeneous protein products as regards N-glycosylation, their glycosylation deviates to some extent from mammalian one and this can have positive and negative effects on recombinant proteins. The positive aspect can lie in the exploitation of plant-specific N-glycans as “cis-adjuvants” in vaccine administration, for their ability to stimulate antigen uptake via lectin receptors, followed by degradation and antigen presentation of dendritic cells to T cells. The negative effects are restricted to a minor immunological response, leading to allergic reactions [8]. For these reasons, the set up of plant-based expression systems for glycoproteins production and the possibility to exploit plant glycans as adjuvant in vaccines formulations, have focused investigators’ attention on the study of glycosylation in plants.

Plant and mammals glycosylation differ in the synthesis of complex type N-glycans, occurring in the Golgi compartment. Plants are characterized by the presence of β -1,2-xylose and α -1,3-fucose residues on the core structure, but they lack β -1,4-galactose and sialic acid.

Several proteomic approaches have been applied in the analysis of plant cell wall proteins, but the deep characterisation of glycoproteome is still poor. Recently, Zhang and co-workers proposed a complex and interesting study addressed to the implementation of information on plant cell wall glycoproteome by combining different biochemical strategies and bioinformatics tools [9]. Here, we report the investigation of glycoproteome from total protein extract in *Arabidopsis thaliana* by using ConA affinity purification and LC-MS/MS for protein identification. Oligosaccharides analysis was carried out by MALDI-TOF/TOF.

IV.2. Materials and methods

Protein extraction from Arabidopsis thaliana

Leaves of *Arabidopsis thaliana* Columbia-0 ecotype were collected from young plants at three weeks after sowing. Plants were grown in a controlled growth chamber with a 16-h-light/8-h-dark photoperiod at 22/18 °C. Tissues were homogenized by grinding them with a pestle and mortar in liquid nitrogen. Then, 1 ml of 2X ice cold extraction buffer (Na₂HPO₄ 10 mM, KH₂PO₄ 2 mM, NaCl 136.9 mM, KCl 2.7 mM, at pH 7.4) and 1 mM PhenylMethyl-Sulfonyl Fluoride were added for each gr of fresh tissue. Two centrifugation steps were performed at 4 °C for 20 minutes at 14000 rpm to remove the cell debris.

Protein concentration determination

Arabidopsis thaliana extract concentration was determined by BCA assay. 5 different concentrations of BSA were incubated with 980 µL of reagent A and 20 µL of reagent B. Such points were used to obtain a linear calibration curve.

The sample was prepared in the same way and everything was incubated at 37°C for 30 min. At the end absorbance at 562 nm was measured.

SDS-PAGE and selective staining of glycoproteins

SDS-PAGE was performed, loading on a 1.5 mm, 12.5 % gel three different amounts of total proteins, 20, 50 and 100 µg respectively, together with a negative and a positive control. The gel was run at constant 25 mA for 1h and stained by a glycoprotein staining kit (Thermo Scientific). Briefly, after electrophoresis, the gel was fixed by completely immersing it in 100 ml of 50% methanol for 30 minutes. Then it was washed three times in 3% acetic acid and transferred in 25 ml of an Oxidizing Solution for 15 minutes. Then it was washed three times in 3% acetic acid for 5 minutes and after transferred in 25 mL of Glycoprotein staining reagent, under gentle agitation for 15 min. 25 mL of Reducing solution were used for 5 min and then the gel was washed extensively with 3% of acetic acid and water.

In situ hydrolysis

Samples were treated as reported in chapter III.

Reduction and alkylation and trypsin digestion.

Samples were treated as reported in chapter II.

Concanavalin A affinity chromatography.

Samples were treated as reported in chapter II.

RP-HPLC.

Samples were treated as reported in chapter II.

Deglycosylation

Glycopeptides from *Arabidopsis thaliana* were lyophilized and resuspended in 100 mM sodium citrate phosphate (Fluka) pH 5 and incubated with 5 U of PNGase A (Roche).

Oligosaccharide purification

Samples were treated as reported in chapter II.

LC-MS/MS

Samples were treated as reported in chapter II.

MALDI-TOF/TOF analysis of glycans

Samples were treated as reported in chapter II. Firstly spectra were manually interpreted and checked against GlycoWorkbench [10] software and Simglycan software [11].

IV.3. Results and discussion

The possibility to obtain plant produced biopharmaceuticals is promising as a more compliant alternative "factory" to classical systems. However, approaches to overcome and solve the associated challenges of this culture system that include a non-mammalian glycosylation are emerging in the field of plant cell-based protein production [12]. In this study we provided a viable strategy to gain new insights to increase the information available on plant glycoproteome. The strategy is basically based on three steps: a) a simple enrichment glycoproteins treatment; b) identification of protein and localisation of the glycosylation sites by mass spectral techniques, c) N-linked glycosylation characterization. Prior to the enrichment, *Arabidopsis thaliana* protein extract was subjected to SDS-PAGE and the resulting gel was stained by a specific glycoprotein staining kit. As a valid step for enrichment, selecting specific glycoforms and simplifying the initial peptide mixture, concanavalin A affinity chromatography was applied to complex matrix from *Arabidopsis thaliana* total cell extract.

SDS-PAGE and glycoproteins staining.

In order to specifically realize the content in glycosylated proteins of the *A. thaliana* extracts, a gel containing separated proteins was treated with an oxidation reagent which exploits the oxidizing properties of periodate directed to the cis-diol sugar groups in glycoproteins. The treatment determines the formation of aldehyde groups, which react with the staining reagent to form Schiff-base bonds, producing a magenta staining corresponding to glycosylated proteins. Different amounts of total extract were loaded on the gel, as well as a positive and a negative control. Fig.1 shows an intense band in the positive control lane, and no staining was detected for the negative control.

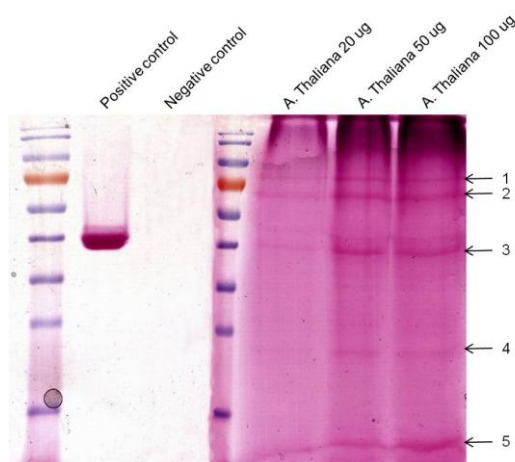


Fig. 1: SDS-PAGE gel stained with glycoprotein staining kit. The first two lanes contain a positive and negative control, A.thaliana extract was loaded in three different concentration

Some more intense bands were detected in the *A.thaliana* extracts, and some bands were excised from the gel as reported in Fig.1 in order to detect potentially glycosylated proteins by *in situ* hydrolysis and LC-MS/MS. Table 1 summarizes all the results obtained. Among the entire pool of identified proteins, most of them were glycosylated, as reported in Swissprot database.

Table 1: List of glycoproteins identified after in situ digestion and LC-MS/MS analyses

Potentially glycosylated proteins
Band 1
Myrosinase 1 P37702
Fasciclin-like arabinogalactan protein 8 O22126
Fasciclin-like arabinogalactan protein 10 Q9LZX4
Endoplasmic homolog Q9STX5
Beta-galactosidase 6 Q9FFN4
Probable glycerophosphoryl diester phosphodiesterase 3 Q7Y208
Fasciclin-like arabinogalactan protein 1 Q9FM65
Subtilisin-like protease O65351
Fasciclin-like arabinogalactan protein 7 Q9SJ81
Endoglucanase 2 Q9FXI9
Band 2
Myrosinase 1 P37702
Protein disulfide isomerase-like 1-1 Q9XI01
Beta-glucosidase 18 Q9SE50
Monocopper oxidase-like protein SKU5 Q9SU40
Reticuline oxidase-like protein Q9SVG4
Fasciclin-like arabinogalactan protein 8 O22126
ABC transporter B family member 17 Q9LSJ6
Protein disulfide isomerase-like 1-2 Q9SRG3
Fasciclin-like arabinogalactan protein 10 Q9LZX4
ABC transporter B family member 22 Q9LSJ2
Probable cellulose synthase A catalytic subunit 9 [UDP-forming] Q9SJ22
Band 3
GDSL esterase/lipase ESM1 Q9LJG3
Peroxidase 34 Q9SMU8
Peroxidase 33 P24101
Peroxidase 12 Q96520
Myrosinase 1 P37702
Fasciclin-like arabinogalactan protein 9 Q9ZWA8
Peroxidase 32 Q9LHB9
Fasciclin-like arabinogalactan protein 13 Q9FFH6
Peroxidase 22 P24102
Peroxidase 38 Q9LDA4
Cysteine proteinase RD21a P43297
Peroxidase 68 Q9LVL1
GDSL esterase/lipase At1g54020 Q9C5N8
Fasciclin-like arabinogalactan protein 2 Q9SU13
Beta-glucosidase 18 Q9SE50
Band 4
Myrosinase 1 P37702
Vegetative storage protein 1 O49195
Germin-like protein subfamily 3 member 3 P94072
Pectinesterase/pectinesterase inhibitor 3 O49006
Germin-like protein subfamily 3 member 1 P94040
Auxin-induced in root cultures protein 12 Q94BT2
Germin-like protein subfamily 2 member 2 Q9FZ27

GDSL esterase/lipase ESM1 Q9LJG3
Peroxidase 34 Q9SMU8
Vegetative storage protein 2 O82122
Band 5
Myrosinase 1 P37702
Vegetative storage protein 1 O49195

Unfortunately this procedure was not able to pinpoint if a protein, indicated as a potentially glycoprotein in databases, is actually glycosylated, and where the glycosylation site is located.

Concanavalin A affinity chromatography on Arabidopsis thaliana.

To answer the above mentioned questions investigations about *Arabidopsis thaliana* glycoproteome were carried out. The philosophy behind the choice to study glycoproteome in this specie relies in the lack of information about global plant glycoproteins and glycosylation pattern. Moreover, the availability of the entire genome sequence, a huge set of natural variants and an ever increasing number of molecular tools, made *A. thaliana* a useful biological system for proteomic approach [13]. It should be noted that generally, procedures for PTM characterization are developed to investigate yeast and mammalian biology and later adopted to elucidate plant systems. In mammalian systems, the study of N-linked glycoproteins or glycopeptides is generally accomplished by removing the glycans and the associated microheterogeneity. Following this step, the relative abundance of the modified peptides increases greatly resulting in an improved peptide identification [14].

Although the general pathway of N-glycan biosynthesis and maturation is conserved in plants, their structure has been reported to be different in different species. However, the plant machinery generally produces glycoproteins with high-Mannose N-glycans, as well as complex N-linked oligosaccharides containing β 1,2 xylose, α 1,3 fucose and terminal GlcNAc or Le^a antennae [15]. The main difference with respect mammalian glycoproteins relies in the presence of α 1,3-fucose instead of α 1,6-fucose at GlcNAc proximal to Asn [16]. This substitution makes α 1,3-fucose not recognizable by PNGase-F. Thus, we used the enzyme PNGase A for the release of the glycan from the protein moiety.

After cold acetone precipitation, proteins were subjected to: reduction, carbamidomethylation and trypsin hydrolysis. Then the peptides mixture was subjected to lectin affinity chromatography (Krusius et al. 1976), by using a specific Concanavalin A, a lectin binding glucopyranosidic and mannopyranosidic groups (Ogata, 1975). Such step not only facilitate glycopeptides selection in the bound fraction, but even in the unbound fraction thus removing the suppression phenomena and leading to less abundant species to be observed. Affinity chromatography was carried out in order to isolate glycopeptides in a complex mixture and to identify N-glycosylation sites by LC-MS/MS. ConA bound peptides were then incubated with PNGase A, in order to cleave N-glycan chains linked to asparagine from glycopeptides. The enzyme catalyzes the cleavage of structures with or without alpha-1,3-linked core fucose residues, characteristic of insects and plants. Peptides were separated from oligosaccharides by RP-HPLC, using a C-18 as stationary phase. Oligosaccharides were characterized by MALDI-MS (see later), whereas peptides were analysed by LC-MS/MS. We performed the enrichment on tryptic digests so that we identified glycoproteins together with N-glycosylation sites. Peptides component of ConA bound and unbound fractions were analysed via LC-MS/MS demonstrating the efficiency of such chromatography in the identification of

glycoproteins. In fact, all the selected peptides contained the conserved motif N-X-S/T for the N-glycosylation.

LC-MS/MS on deglycosylated peptides enabled the identification of several N-glycosylation sites as shown in Table 1. As expected, ConA bound fraction sub-glycoproteome was characterized uniquely by peptides containing one or more putative N-glycosylation sites. Such presence was supported by the increment of 0.984 Da on the mass of the peptide, due to the conversion of asparagine into aspartic acid as an effect of N-glycosydase treatment. In contrast, LC-MS/MS analyses performed on the unbound fraction identified more abundant peptides with not glycosylated sites. These results underlined the high sensitivity of concanavalin A to bind specific oligosaccharidic structures as high mannose. The superior mass accuracy of the Q-TOF 6520 enabled N-glycosylation sites identification with high reliability. The main advantage of the strategy relies on the lectin affinity chromatography that was performed on glycopeptides instead of glycoproteins, without the SDS-PAGE step and the isolation and identification of the individual glycoproteins by in situ digestion.

Table 2: List of N-glycosylation sites identified after ConA affinity chromatography, deglycosylation and LC-MS/MS analysis. N-glycosylation sites are in bold.

Protein	Peptide	Sequence
Peroxidase 34 (Q9SMU8)	40-49 314-328	SCPNVTNIVR MGNITPTTGTQQQIR
Myrosinase (P37702)	100-113	DIDVMDELNSTGYR
Early nodulin-like protein 3 (Q8LC95)	77-97	DAYINCNTTNPAAANYSGDTK
Peroxidase 12 (Q96520)	246-261	TCPTANSSNTQVNDIR
Lamin-like protein (Q39131)	96-104	DIVTLNQTK
Fasciclin-like arabinogalactan protein 8 (O22126)	272-284	NNISTLATNGAGK
LysM domain-containing GPI-anchored protein2 (O23006)	66-79	SILGANNLPLNTR
GDSL esterase/lipase ESM1 (Q9LJG3)	143-152	SNWNDSYIEK
Monocopper oxidase-like protein SKU5 (Q9SU40)	291-298 424-434	VVNETIWR VATSIINGTYR
Uncharacterized GPI anchored protein At5g19240 (Q84VZ5)	42-50	ESVNLTLNK
Fasciclin-like arabinogalactan protein 7 (Q9SJ81)	99-107	NPPLSNLTK
Uncharacterized GPI-anchored protein At3g06035 (Q84MC0)	41-50	TTQNLTLISK
Wall-associated receptor kinase 3 (Q9LMN8)	293-308	DIDECISDTHNCSDPK
Probable pectate lyase 22 (Q93Z25)	63-70	KINESISR

However, some non-specific peptides, namely non-glycosylated peptides, were detected in the eluted ConA fraction, and identified as belonging to most abundant proteins like RuBisCO. As a whole, the LC-MS/MS analysis led to the selective identification of 13 glycosylated proteins present in the *Arabidopsis thaliana* sample. 2 of them being identified as arising by more than one unique peptide sequence and

11 arising from single-peptide matching. No N-glycosylation sites were detected in the unbound fraction.

As for *Arabidopsis thaliana*, proteins databases only contain the indication of the putative N-glycosylation sites. Most of the data on glycosylation in plants come from studies using transgenic plants or from the biochemical characterization of few proteins [17].

Methods for enrichment in glycoproteins and mass spectrometry (MS) techniques were applied to the characterization of cell wall glycoproteome of etiolated hypocotyls in *A. thaliana* by Zhang [9]. Most glycoproteins were found to be putative N-glycoproteins and N-glycopeptides were predicted from MS data using the ProTerNyc bioinformatics software. The results obtained in this paper, complement the data available on plant glycoproteome so far. For the first time, in this work, several plant glycosylation sites were effectively demonstrated. This methodology, in fact, confirmed the genuine glycosylation of those putative sites, giving new insights to the structural elucidation of glycoproteome in such plant specie.

Characterization of oligosaccharides

Considering the growing interest in the exploitation of novel protein expression systems and the development of glycoengineering as a new tool for the production of recombinant proteins with specific N-glycan profiles, the study of glycosylation in plants is arousing a particular interest.

In this field, MALDI-TOF/TOF mass spectrometry can be considered the best technique in terms of speed and sensitivity in the analyses to analyse oligosaccharide mixtures and to detect any difference between the analytes.

In order to obtain a more complete and detailed glycan profiling without any discrimination, PNGase A was directly added to the bound and unbound fraction after Concanavalin A affinity chromatography. Glycans were separated from peptides by a HPLC step and analysed by MALDI-MS and MALDI-MS/MS. Fig.2A-B shows the positive ion mode spectrum of the oligosaccharide-bound and unbound fractions of the *A. thaliana* sample. Figure 2A reported the MALDI-TOF spectrum of the bound fraction and the attribution of each peak as reported in table 3. High mannose type glycans were the main identified structures.

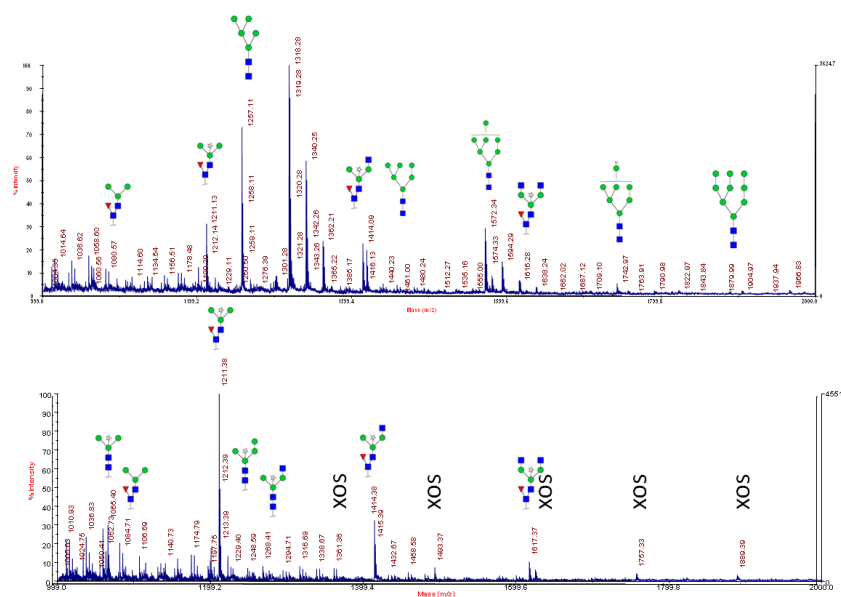


Fig. 2: positive ion mode MALDI-MS of ConA bound(A) and unbound (B) fractions of oligosaccharides extracted from A.thaliana glycoproteins

Table 3: list of oligosaccharides structures identified by MALDI-MS in ConA bound and unbound fractions.

m/z	Composition		ConA fraction
1065.22	XylHexNAc ₂ Hex ₃		Bound, unbound
1078.86	FucHexNAc ₂ Hex ₃		Bound, unbound
1095.20	HexNAc ₂ Hex ₄		Bound
1211.21	XylFucHexNAc ₂ Hex ₃		Unbound
1257.56	HexNAc ₂ Hex ₅		Bound
1268.32	XylHexNAc ₃ Hex ₃		Unbound
1414.23	FucXylHexNAc ₃ Hex ₃		Bound, unbound
1419.11	HexNAc ₂ Hex ₆		Bound
1581.11	HexNAc ₂ Hex ₇		Bound
1617.11	FucXylHexNAc ₄ Hex ₃		Unbound
1742.05	HexNAc ₂ Hex ₈		Bound
1905.01	HexNAc ₂ Hex ₉		Bound

The higher representative structures in the unbound fractions (fig 2B and table3) belong to paucimannosidic compounds. These data are in agreement with the literature data on the characterisation of N-glycans in plant consisting only of the Man₃GlcNAc₂ core or of a restricted Man₂GlcNAc₂ core or even shorter modifications carrying α -1,3-linked Fuc and β -1,2 Xyl [18]. Moreover, MALDI spectra showed the occurrence of peaks that have been interpreted as polymers of xilose and mannose. The main differences between the fractions consisted in the presence of high mannose structures solely in the ConA eluted fraction.

Peaks were manually attributed taking in account the biosynthetic pathway on plant N-linked glycans and then confirmed by MALDI-MS/MS. Fragmentation spectra were interpreted manually with the help of Glycoworkbench software [10] which simulates

glycans fragmentation and compares the experimental values for m/z of fragments with the theoretical ones.

In addition to this, one more software was used and tested to support glycans structural characterization. This software, Simglycan was developed by Premierbiosoft [11] and can be used as a bioinformatic tool in different phases of glycomics studies: the most intuitive application of the software is the support in the prediction of glycans structures on the base of the MS/MS spectra and the precursor m/z, matching them with specific databases. The software generates a list of probable glycan structures, scoring them on the base of how they are close to the experimental data. It is able to provide other biological information about the identified structures with links to numerous databases. Fig.3 depicts the output of the prediction of the structure of ion at m/z 1065.27. It is worth noting that the software predicted 4 different structures and scores were assigned considering the number of glycosidic fragments and cross ring fragments.

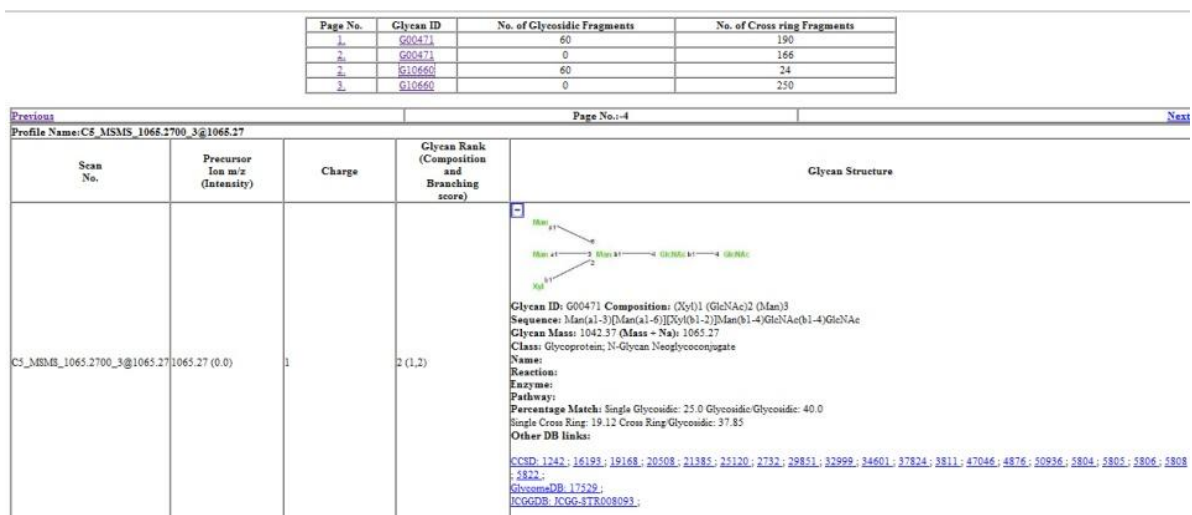


Fig. 3: Simglycan identification for the ion at 1065.27 on the basis of MS/MS spectrum

Another form of output of the software is the creation of an annotated spectrum with matched fragment ions. Fig.4A represents the experimental MS/MS spectrum of ion at m/z 1065.27 and Fig.4B depicts the spectrum generated by Simglycan, where the x-axis contains the m/z values and the y-axis the experimental intensities of each peak extrapolated from the original spectrum. Each peak was automatically annotated with the correct fragment ion. Fig.5A-B shows the fragmentation spectrum of ion at m/z 1257.32. Even in this case an annotated spectrum was generated (Fig.5B).

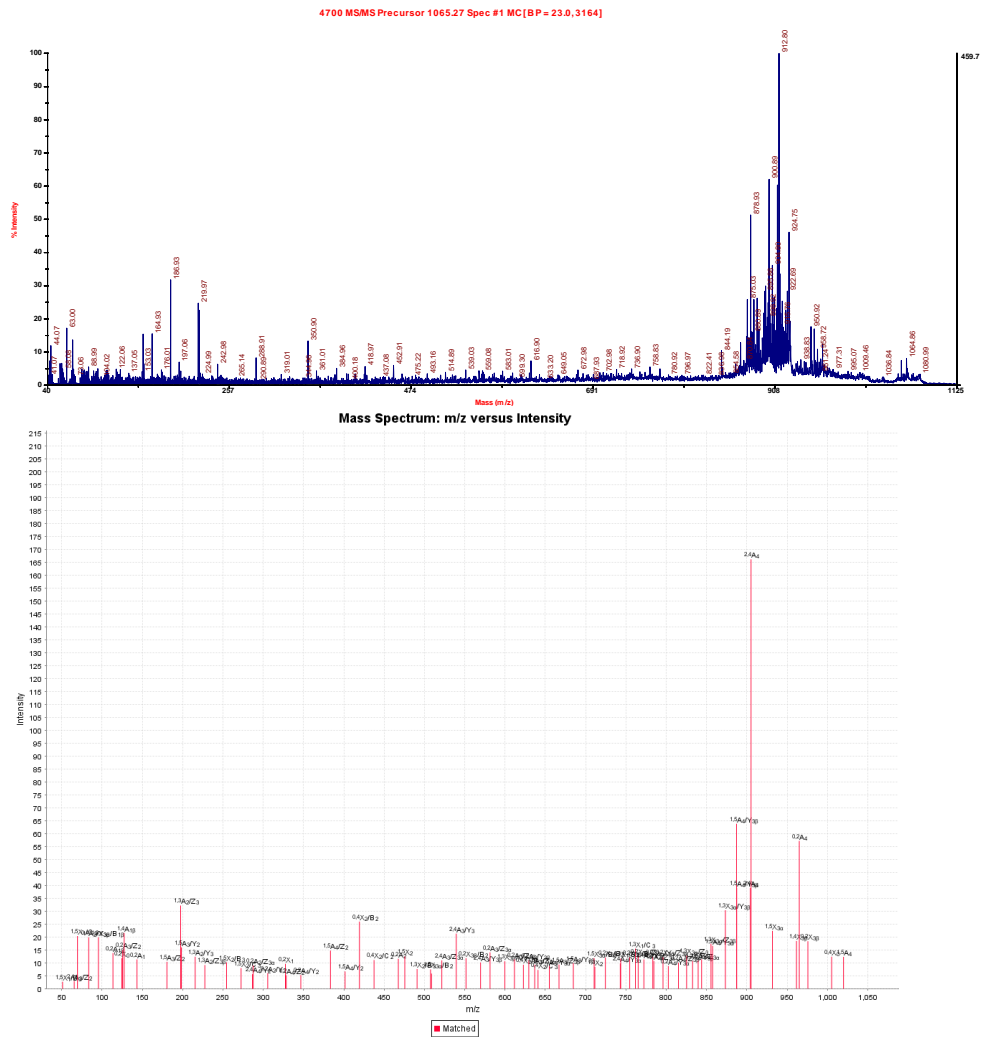


Fig. 4: (A) fragmentation spectrum obtained on a MALDI-TOF/TOF spectrometer, corresponding to ion at m/z 1065.27. (B) spectrum generated by Simglycan, where the x-axis contains the m/z values and the y-axis the experimental intensities of each peak extrapolated from the original spectrum. This spectrum was automatically annotated

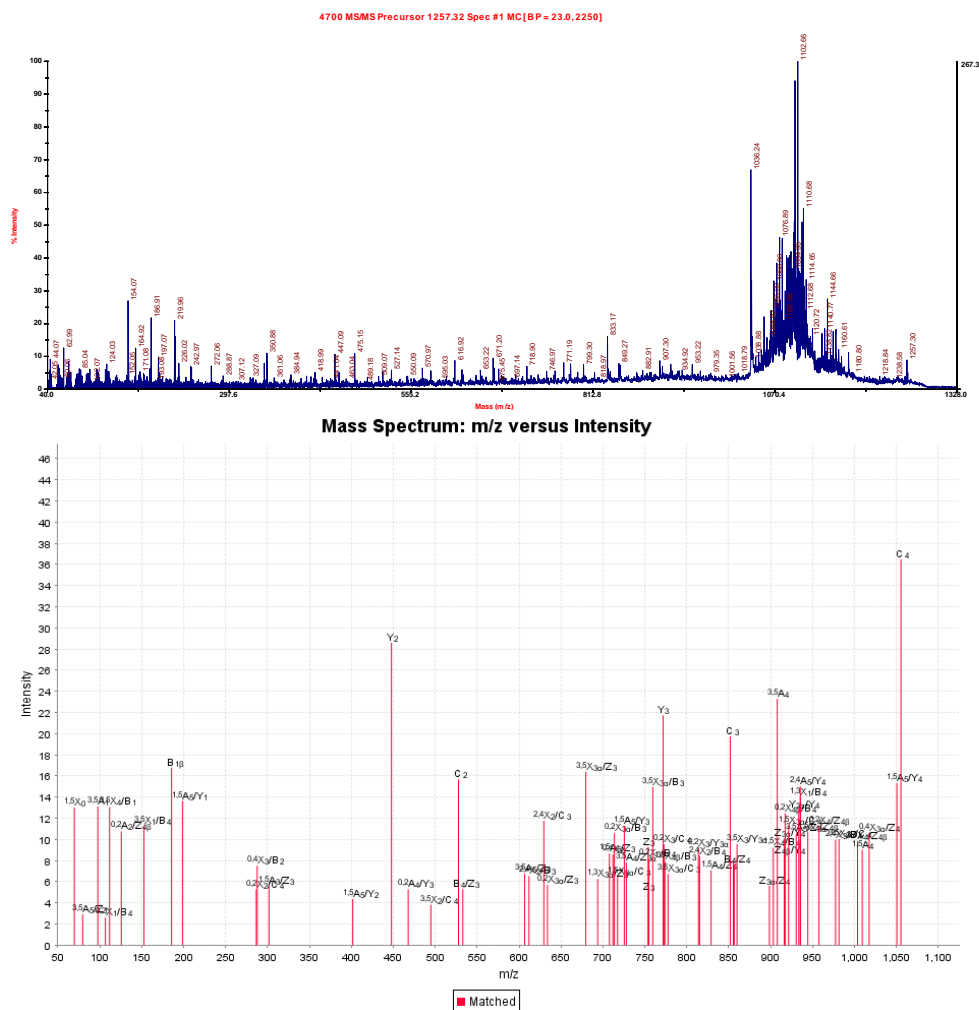


Fig. 5: (A) fragmentation spectrum obtained on a MALDI-TOF/TOF spectrometer, corresponding to ion at m/z 1257.32. (B) spectrum generated by Simglycan, where the x-axis contains the m/z values and the y-axis the experimental intensities of each peak extrapolated from the original spectrum. This spectrum was automatically annotated

Moreover the software presents the possibility to extract many kinds of report, in particular it is possible to visualize graphical representations of ion fragments.

IV.4 Conclusions

As a whole, this paper report on the isolation and analysis of glycoproteins in *Arabidopsis thaliana* by coupling lectin affinity chromatography with MS/MS techniques. Thus revealing that this integration could be considered as a powerful strategy to study the glycoproteome in plant systems. However, to the best of our knowledge only few works are still present [19] on the analysis of plant glycoproteins. The value of this approach is illustrated by the characterization of glycoproteins that are expressed in ripe tomato (*Solanum lycopersicum*) fruit. This strategy allowed the isolation of a fraction of *A. thaliana* glycoproteome with an extremely high proportion of proteins that are predicted to be glycosylated, and the identification of new putative N-glycosylation sites. The structural characterization of glycan moieties was performed by using MALDI-TOF/TOF mass spectrometry. The elucidation of glycan structures resulted of interest in the analysis of the plant system for protein production. Moreover, the direct localization of N-glycosylation sites and the structural characterization of N-linked glycans were reported thus providing better insights in glycoproteome investigation useful for plant cell-based protein factories.

IV.5 References

- [1] Ma JK, Hiatt A, Hein M, Vine ND, Wang F, Stabila P, van Dolleweerd C, Mostov K, Lehner T. Generation and assembly of secretory antibodies in plants. *Science*. 1995 May 5;268(5211):716-9.
- [2] Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. Rapid high-yield expression of full-size IgG antibodies in plants coinfecting with noncompeting viral vectors. *Proc Natl Acad Sci U S A*. 2006 Oct 3;103(40):14701-6.
- [3] Sainsbury F, Thuenemann EC, Lomonossoff GP. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J*. 2009 Sep;7(7):682-93.
- [4] Kwon TH, Seo JE, Kim J, Lee JH, Jang YS, Yang MS. Expression and secretion of the heterodimeric protein interleukin-12 in plant cell suspension culture. *Biotechnol Bioeng*. 2003 Mar 30;81(7):870-5.
- [5] Hellwig S, Drossard J, Twyman RM, Fischer R. Plant cell cultures for the production of recombinant proteins. *Nat Biotechnol*. 2004 Nov;22(11):1415-22.
- [6] Xu J, Dolan MC, Medrano G, Cramer CL, Weathers PJ. Green factory: Plants as bioproduction platforms for recombinant proteins. *Biotechnol Adv*. 2011 Sep 8.
- [7] Loos A, Van Droogenbroeck B, Hillmer S, Grass J, Kunert R, Cao J, Robinson DG, Depicker A, Steinkellner H. Production of monoclonal antibodies with a controlled N-glycosylation pattern in seeds of *Arabidopsis thaliana*. *Plant Biotechnol J*. 2011 Feb;9(2):179-92.
- [8] Bosch D, Schots A. Plant glycans: friend or foe in vaccine development? *Expert Rev Vaccines*. 2010 Aug;9(8):835-42.
- [9] Zhang Y, Giboulot A, Zivy M, Valot B, Jamet E, Albenne C. Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry*. 2011 Jul;72(10):1109-23.
- [10] Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008 Apr;7(4):1650-9.
- [11] Apte A, Meitei NS. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol Biol*. 2010;600:269-81.
- [12] Xu J, Ge X, Dolan MC. Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures. *Biotechnol Adv*. 2011 May-Jun;29(3):278-99.
- [13] Wienkoop S, Baginsky S, Weckwerth W. *Arabidopsis thaliana* as a model organism for plant proteome research. *J Proteomics*. 2010 Oct 10;73(11):2239-48.
- [14] Ytterberg AJ, Jensen ON. Modification-specific proteomics in plant biology. *J Proteomics*. 2010 Oct 10;73(11):2249-66.
- [15] Gomord V, Fitchette AC, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D, Faye L. Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J*. 2010 Jun;8(5):564-87.
- [16] Gomord V, Chamberlain P, Jefferis R, Faye L. Biopharmaceutical production in plants: problems, solutions and opportunities. *Trends Biotechnol*. 2005 Nov;23(11):559-65. Epub 2005 Sep 15.

- [17] Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R. Recent advances in plant cell wall proteomics. *Proteomics*. 2008 Feb;8(4):893-908.
- [18] Pattison RJ, Amtmann A. N-glycan production in the endoplasmic reticulum of plants. *Trends Plant Sci*. 2009 Feb;14(2):92-9.
- [19] Catalá C, Howe KJ, Hucko S, Rose JK, Thannhauser TW. Towards characterization of the glycoproteome of tomato (*Solanum lycopersicum*) fruit using Concanavalin A lectin affinity chromatography and LC-MALDI-MS/MS analysis. *Proteomics*. 2011 Feb 25.

V. Cell surface glycoconjugates: innate immunity probed by lipopolysaccharides affinity strategy and proteomics

V.1.Introduction

As it is illustrated in chapter I, glycans and glycan-binding proteins (lectins) play a crucial role in the host-pathogen interaction as concerning the onset of innate and adaptive immune processes. Bacteria synthesize several glycoconjugates, including lipopolysaccharides (LPS), peptidoglycan, glycoproteins, capsules, teichoic acids and exopolysaccharides by using nucleotide-activated sugars as substrates for different glycosyltransferases. LPS and glycoproteins biosynthetic pathways involve homologous proteins, often characterized by a wider substrate specificity, that make them the ideal candidates to be exploited in engineered living bacteria for industrial glycoconjugate production, generating potential vaccines and therapeutics [1]. The full exploitation of glycoconjugates potential for industrial biotechnological and pharmaceutical applications requires a deep knowledge of the immunological mechanisms at the base of host-pathogen recognition.

In a biological context, the term immunity refers to the set of events and biological cascades that protect organisms against infectious diseases. This protection system is based on complex interconnections between cells and molecules whose synergic action determines the immune response. The main physiological function of immune system is the protection against infectious agents or simply against exogenous elements. The ability to distinguish “self” from “non-self” is critical to prepare an effective immune response [2].

Immune system response can be classified into two categories: adaptive immunity and innate immunity. The adaptive immune system is based on specialized processes whose function is to eliminate external threats during the late phase of infection with the generation of immunological memory. Specificity is developed by clonal gene rearrangements form a range of antigen-specific receptors on lymphocytes [3]. Only vertebrates are provided with this type of immunity, allowing the organism to recognize and remember the “non-self”. Innate immune system is necessary to activate adaptive immune system. It acts in a non-specific manner and does not confer any memory of the challenge. It is widespread amongst all animals and even in plants, and its main function is to provide immediate protection against pathogens. Innate immunity relies on specific cells as primary mediators, namely phagocytic cells and antigen presenting cells (APCs), such as granulocytes, macrophages, and dendritic cells (DCs) [4]. Innate immunity mechanisms do not have specific pathogen targets, the principal targets are evolutionarily conserved molecular structures, termed “pathogen associated molecular patterns” (PAMPs) whose expression is common to different pathogens. These structures are generally recognized by receptors on the membrane of the host cells that are responsible for the immune response. Such structures are called “pattern recognition receptors” (PRR). PRRs are constitutively expressed in the host on all cells of a given type [5]. Lipopolysaccharide (LPS) is a glycolipid, generally referred to as endotoxin in Gram negative bacteria, belonging to PAMP family. It is a potent inducer of the innate immune system and the main cause of septic shock [6]. Lipopolysaccharides are amphiphilic macromolecules composed of a hydrophilic heteropolysaccharide covalently linked to a lipophilic moiety termed lipid A, which anchors these macromolecules to the outer membrane. The polysaccharidic region is divided into a core oligosaccharide and a region called O-chain. Lipopolysaccharides not possessing the O-chain are termed rough lipopolysaccharides or

lipooligosaccharides (LOSs). Lipid A is the lipophilic portion of LPS responsible for anchoring the entire macromolecule to the outer membrane of bacteria through electrostatic and hydrophobic interactions and it represents the endotoxic principle of Gram negative. The basic structure of lipid A consists of two units of glucosamine called GlcN I and GlcN II, linked by a glycosidic linkage β -(1 \rightarrow 6), and repeated throughout the lipopolysaccharide. Linked to the disaccharidic backbone there are fatty acid chains responsible of the hydrophobic nature of lipid A. Positions 1 and 4' of the lipid backbone are occupied by two phosphate groups known as "polar heads", not present in stoichiometric quantities. They can be linked to different substituents such as ethanolamine or glucosamine, to neutralize the negative charge of phosphates. In position 6 of GlcN II a characteristic monosaccharide of LPSs is linked, namely 2-Keto-3-deoxy-D-manno-octonic acid which represents the conjunction factor between the lipid (lipid A) and the sugar (the core) portion. The Core consists of about fifteen monosaccharides, and is divided into two distinct domains called inner-core and outercore. The inner-core consists of characteristic monosaccharides, such as KDO, L-glycero-D-manno-heptose, D-glycero-D-manno-heptopyranose; the outer-core instead, consists of a few glycosidic residues: neutral sugars, uronic acids and amino sugars. The core is always negatively charged due to the presence of acid or phosphorylated sugar substituents. The O-specific chain is the structure protruding outside from the cell surface and contains up to fifty repeating oligosaccharide units. Most of lipopolysaccharides heterogeneity is given by the extremely variable structure of O-chain [7].

The recognition of this endotoxin is achieved by the interplay of a variety of proteins, either receptors or serum secreted proteins. The cascades of events following LPS infection has been investigated in numerous studies. However, most of them focused on the interaction of single specific proteins with LPS, such as the LPS-binding protein (LBP), whose mechanism of action was clarified in many works [8,9]. Further studies showed that other proteins possess LPS-binding domains [10,11,12]. Intracellular pathways elicited by LPS infection was investigated by in vitro and in vivo analyses on different cell types, ranging from immune cells to epithelial cells [13;14].

Toll-like receptor 4 (TLR-4) is the signal transducing receptor of lipopolysaccharides; lipid A is the real effector required to activate TLR4 signalling pathway in conjunction with a soluble co-receptor protein myeloid differentiation protein 2 (MD2), which directly and physically binds to LPS [15;16]. Entry of LPS or lipid A into the blood stream is responsible for the onset of septic shock [17]. Aggregates of LPS bind to certain plasma proteins, such as albumin, LBP and soluble CD14 to facilitate the interaction with host cells, via TLR4. The complex of protein bound LPS, TLR4, and MD2, activates two intracellular pathways [18]: the MyD88-dependent pathway, responsible for early-phase NF- κ B and MAPK activation, which controls the induction of proinflammatory cytokines TNF- α , IL-1, IL-6, and the MyD88-independent, TRIF-dependent pathway that activates IRF3, which is necessary for the induction of IFN- β - and IFN-inducible genes [19]. The activation of the TLR4-MD2 complex triggers the induction of inflammatory cytokines acting as endogenous mediator of infection, as well as the superoxide anion (O $_2^-$), hydroxyl radicals (OH), nitric oxide and antimicrobial peptides [20].

The low and balanced concentrations of these mediators and soluble immune response modulators leads to a resulting inflammation, one of the most important and ubiquitous aspect of the immune host defence against invading microorganisms. Beside these positive effects, an uncontrolled and massive immune response, due to

the circulation of large amount of endotoxins, leads to the symptoms of the sepsis and of the septic shock. Therefore identification of all the molecular actors involved in the recognition cascade is of pivotal importance with particular emphasis on the differences in elicitation and reaction by different LPSs.

Numerous studies have been performed following pro-inflammatory expression patterns via microarray analyses [21] and the phage display method was used to capture new LPS-binding peptides [22].

Recently, proteomics approaches were addressed to the study of innate immunity. Comparative analysis of the human plasma proteome prior and after LPS injection was performed [23] and the human innate immunity interactome was investigated using 58 genes involved in transcriptional regulation of type I IFN tagged with the FLAG epitope following cells stimulation with LPS [24].

The aim of this study was the setting up of a new methodology for the development of lipopolysaccharides analyses to be used to capture LPS-interacting proteins in order to better elucidate the molecular mechanism of the innate immunity. Here, a strategy based on chemical manipulations, biochemical procedure and mass spectrometry techniques, integrated in a functional proteomic workflow is proposed. The procedure was developed using an LPS derived from *Salmonella typhimurium* and the proteomic study was carried out by using human serum from a healthy donor as a model.

V.2. Material and methods

LPS extraction

Dried cells of *Salmonella typhimurium* were extracted by the phenol/water method [25]. The extracted phases were dialysed three times against distilled water and subjected to enzymatic digestions in order to remove nucleic acids and protein contaminants. Both water and phenol fractions were analysed through SDS-PAGE 13.5%.

LPS biotinylation

LPS biotinylation was achieved by a transesterification reaction with biotin-p-nitrophenylester (Sigma Aldrich). Lyophilized LPS was solubilised in 200 μ L pyridine (Romil), to which a certain amount of biotin-p-nitrophenylester was added. LPS: biotin-p-nitrophenylester ratio was calculated as 1:3 (weight:weight). The reaction was carried out in the dark for 2 hours at 80°C. The reaction mixture was dried under nitrogen and then the sample was dissolved in MilliQ water. The excess of reagents was removed by molecular exclusion chromatography on PD-10 columns (GE-Healthcare). LPS containing fractions were pooled and lyophilized. Unlabelled and biotinylated LPS were dissolved in Laemmli buffer, supplemented with 100 mM DTT and separated by SDS-PAGE in a 12.5% polyacrylamide gel. LPS was detected by silver nitrate staining for carbohydrates (Tsai Frash, 1982) or electrotransferred onto a PVDF membrane (Millipore). The membrane was blocked with 3% non fat dry milk, 1% BSA (Sigma Aldrich), in PBS containing 0.05% Tween 20 (Sigma Aldrich). Then it was incubated with 1:1000 streptavidin/HRP conjugate (Sigma Aldrich). Biotinylated sample was detected using SuperSignal West Femto Chemiluminescent Kit from Pierce.

LPS immobilization onto avidin beads

Biotin-LPS immobilization was carried out using 3 mg biotin-LPS/mL of settled avidin agarose resin (settled gel, Pierce).

An aliquot of resin was extensively washed with binding buffer (0.1 M Na₃PO₄; 0.3 M NaCl; pH 7.2) and incubated with biotin-LPS at 4°C overnight under shaking. Then the resin was washed three times with binding buffer.

Biotin-LPS and human serum interaction

The affinity experiment was preceded by a “precleaning step”. A protease inhibitors cocktail (Sigma Aldrich) was added to 5 mg human serum (Servizio Analisi, Policlinico), to a final concentration of 1 mM. This sample was diluted to a final volume of 600 µL with binding buffer and incubated with 150 µL of settled resin overnight at 4°C. The unbound fraction of precleaning step was therefore incubated with 150 µL of biotin-LPS immobilized resin overnight at 4°C. Later on, the two aliquots of resin, that we term “control” and “test” were subjected to repeated washes with binding buffer. Elution was performed boiling the resin in Laemmli buffer and DTT.

SDS-PAGE was performed, loading on a 1.5 mm, 12.5 % gel all the samples deriving from the different steps of the fishing experiment. The gel was run at constant 25 mA for 1h. The gel was stained with colloidal Coomassie (Pierce).

In gel trypsin digestion

The analysis was performed on the Coomassie blue-stained protein bands excised from the gels. Gel particles were washed first with acetonitrile and then with 0.1 M ammonium bicarbonate. Protein samples were reduced incubating the bands with 10 mM dithiothreitol (DTT) for 45 min at 56 C. Cysteines were alkylated by incubation in 5 mM iodoacetamide for 15 min at room temperature in the dark. The bands were then washed with ammonium bicarbonate and acetonitrile. Enzymatic digestion was carried out with trypsin (12.5 ng/µl) in 50 mM ammonium bicarbonate buffer, pH 8.5. Gel particles were incubated at 4°C for 2h, in order to allow the enzyme to enter the gel. The buffer solution was then removed and a new aliquot of buffer solution was added for 18 h at 37 C. A minimum reaction volume, enough for complete gel rehydration was used. At the end of the incubation the peptides were extracted by washing the gel particles 0.1% formic acid in 50% acetonitrile at room temperature and then lyophilised.

LC-MS/MS analyses

Peptides mixtures were analyzed by LC-MS/MS, using a HPLC-Chip LC system (Agilent 1200) connected to a Q-TOF 6520 (Agilent Technologies). Samples were diluted in 10 µL 0.1% formic acid. After loading, the peptide mixtures were concentrated and washed at 4 µL/min in 40 nL enrichment column with 0.2% formic acid in 2% acetonitrile. Fractionation was carried out on a C-18 reverse phase column (75 µm x 43 mm) at a flow rate of 0,4 µL/min with a linear gradient of eluent B (95% acetonitrile and 0.2% formic acid) in A (2% acetonitrile and 0.1% formic acid) from 7% to 80% in 51 min.

Mass spectrometry analyses were performed using data dependent acquisition MS scans (mass range 300-2400 m/z), followed by MS/MS scans (mass range 100-2000 m/z) of the most intense ions of a chromatographic peak.

Raw data from LC-MS/MS were converted to mzData. The spectra were searched against the NCBI database (2006.10.17 version) using the licensed version of Mascot 2.1 (Matrix Science). The MASCOT search parameters were: taxonomy

Homo sapiens; allowed number of missed cleavages 2; enzyme trypsin; variable post-translational modifications, methionine oxidation, pyro-glu N-term Q; peptide tolerance 10 ppm and MS/MS tolerance 0.6 Da; peptide charge, from +2 to +3.

V.3.Results and discussion

The innate immune system constitutes the first line of defence against microorganisms and plays a primordial role in the activation and regulation of adaptive immunity. In humans, components of the innate immune system include members of the complement cascade and soluble pattern recognition molecules (PRRs), functional ancestors of antibodies [26]. In particular, the complement system plays a key role in the elimination of micro-organisms after entrance in the human host.

The integrated strategy based on functional proteomic approach represents a new starting point to elucidate the molecular mechanisms elicited by bacterial LPS and involved in the different steps of innate immunity response. Identification of the specific LPS-binding proteins in human serum was rarely addressed since most investigations focused on the LPS-mediated effects on cellular pathways and transcriptional responses. Only recently, a single example was reported using a solid-phase chemically immobilised LPS [27].

Immobilisation of the LPS represents the key step in the entire strategy to generate a labeled bait that could be linked to agarose beads. Biotinylation of LPS was chosen as the best suited derivatization process both for the simple modification reaction and to take advantage of the strong and specific interaction of biotin with avidin.

Biotinylated LPS was analysed by SDS-PAGE following both silver staining and immunodetection with streptavidin/HRP conjugate. Figure 1 shows the corresponding gel. Staining procedures displayed the occurrence of two sample bands. The low molecular mass bands corresponds to the LipidA core component of LPS (R-type LPS) whereas the high molecular mass band represents the whole LPS molecule comprising the O-chain moiety (S-type LPS). Both components were responsive to western blot analysis demonstrating that they had been derivatised by biotin.

Biotin-LPS was conjugated to avidin agarose beads by incubation in 0.1 M Na₃PO₄ overnight. Both unbound and bound fractions were analyzed by SDS-PAGE followed by blotting and incubation with streptavidin/HRP conjugate. Fig. 2 shows the results of the immobilization procedure; two immunoresponsive bands corresponding to LPS and the Lipid A moiety were clearly detected in the bound fraction while no immunoresponsive bands occurred in the unbound material.

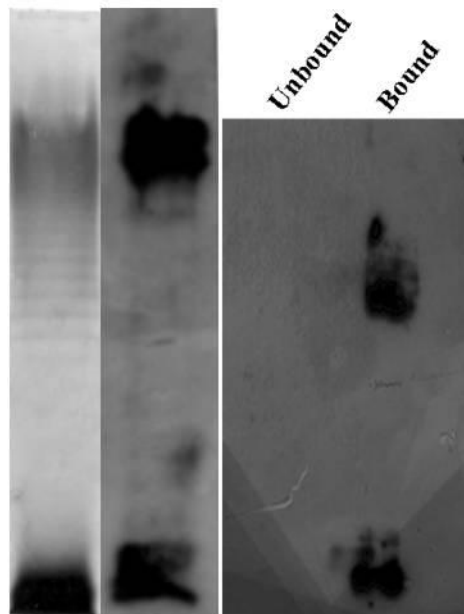


Fig. 1: (A) SDS-PAGE of LPS after biotin-p-nitrophenylester derivatization detected by silver staining (on the left) and blotting with streptavidin/HRP incubation (on the right). (B) immobilization procedure probed by blotting with streptavidin/HRP incubation. Two bands corresponding to LPS and Lipid A moiety were clearly detected in the bound fraction while no bands occurred in the unbound material

The final step of the proposed strategy relies on the use of the immobilised LPS as bait in a functional proteomic experiment [28] aiming at capturing specific LPS interactors to investigate innate immunity molecular mechanisms in human serum. An aliquot of human serum proteins was incubated with underivatized avidin agarose beads overnight at 4°C as a pre-cleaning step to remove all the proteins that non specifically interact with agarose or avidin. The unbound fraction was then incubated with biotin-LPS avidin agarose beads and after extensively washing in binding buffer both the bound fractions (control and sample) were eluted in Laemmli buffer and fractionated by SDS-PAGE gel. An aliquot of 15 µg of human serum was also loaded onto the gel (Fig.3). Due to the complexity of the gel patterns and the low resolution of 1D electrophoresis, several proteins can occur in the same gel band. Therefore, protein bands specifically present in the sample lane and absent in the control lane cannot be identified by simply comparing the two gel profiles. Thus the entire sample lane of the gel was cut in 25 slices. Each slice was destained and in situ trypsin digested. To check for non specific proteins, the same procedure was applied to the control lane. The resulting peptide mixtures were extracted from the gel and submitted to nanoLC-MS/MS analysis generating sequence information on individual peptides.

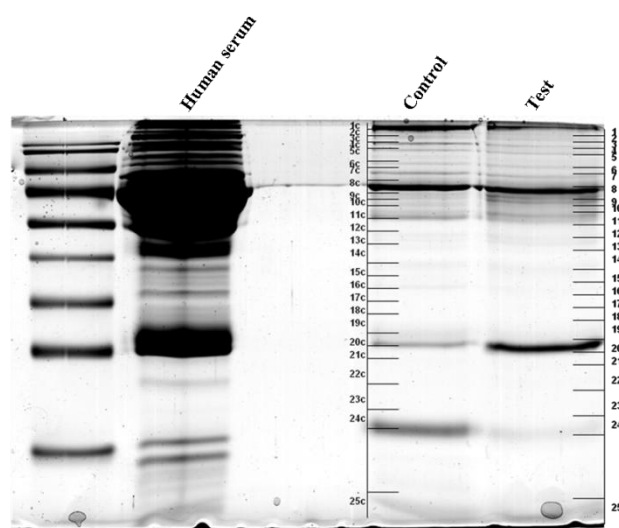


Fig. 2: 12,5% SDS-PAGE gel containing 15 µg of human serum together with control and sample bound fractions after elution in Laemmli buffer. The entire control and sample lanes of the gel were cut in 25 slices. Each slice was destained and trypsin digested, in situ.

This information, together with the peptide mass values, was then used to search protein databases, leading to the identification of the protein components. The experiments were performed in two biological replicates and each protein mixture was run through LC/MS/MS three times. All proteins identified in both the sample and control lanes were discarded, while only those solely occurring in the sample lane and absent in the control lane were selected as putative LPS interactions. As further selection criteria, only proteins found in both biological replicates and identified with at least 2 significant peptides in MASCOT search were considered, providing a full list of putative LPS interactors. The results are summarised in Table 1.

Table 1: Table 1: list of putative LPS-interacting proteins after in situ digestion of the bands from control and sample bound fractions. As selection criteria only proteins solely present in the sample lane and completely absent in the control lane were considered putative interactors. Analyses were carried out in replicates and only proteins present in both replicates and identified with at least 2 peptides in MASCOT search were considered

ID	Protein	Score	Peptides
P01023	Alpha 2 macroglobulin	170	8
P08603	Factor H	44	2
P00450	Ceruloplasmin	136	6
P80108	Phospholipase D	114	4
P00734	Coagulation factor II (Thrombin)	56	4
P07225	S plasma protein	140	8
P04196	Histidine rich glycoprotein	54	2
P04003	C4 binding protein C4bp	266	16
P02749	Beta 2 glycoprotein I	91	5
P02647	Proapolipoprotein A1	86	3
P0C0L5	Complement C4B	92	3
P05090	Apolipoprotein D	56	2
P06727	Apolipoprotein A-IV	74	4
P12273	Prolactin inducible precursor	64	2

The strategy based on affinity capture procedure using biotin LPS as a bait was optimized and human serum from a healthy donor was used as model. A number of proteins were specifically retained by the LPS bait and subsequently identified by mass spectrometric methodologies. The identity of these proteins represented per se a validation of the developed affinity-capture procedure. These proteins, in fact, essentially gathered within two functional groups, regulation of the complement system (Factor H, C4b, C4BP and alpha 2 macroglobulin) and inhibition of LPS induced inflammation (HRG and Apolipoproteins). Both groups are functionally correlated with the biological role of LPS.

The complement system is an essential component of the innate immune system that participates in elimination of pathogens. Three different pathways synergistically contribute to the protection mechanism elicited by the complement proteins, the classical pathway, the alternative pathway and the lectin pathway. Despite the specific pathway they belong to, complement proteins provide host defence towards bacterial infections by binding to cell surface components of exogenous microorganisms. In contrast, pathogenic microorganisms have evolved several strategies to escape these defence mechanisms. The most common procedure consists in the recruitment of complement inhibitory proteins to the bacterial surface thus impairing the immune system.

In this respect, the identification of well characterized protein inhibitors of the complement system among the LPS interactors has a clear biological significance. C4BP, Factor H and alpha 2 macroglobulin are all complement inhibitors able to downregulate activation of the classical, alternative and lectin pathways respectively. *Streptococcus pyogenes*, the etiologic agent of important human infections, was shown to bind Factor H, C4BP and other complement proteins as a crucial step in the pathogenesis of these infections [29]. *Bordetella pertussis* was able to escape the classical pathway of complement by binding the classical pathway inhibitor C4BP. In addition, very recently it was shown that this pathogen can also evade the alternative complement pathway by recruitment of Factor H on the cell surface [30].

On the other hand, alpha 2 macroglobulin exerts a specific inhibitory activity towards the lectin pathway by binding the mannose-binding lectin (MBL) and the MBL associated serine proteases MASP-1 and MASP-2 [31]. Adding of alpha 2 macroglobulin to human serum totally reversed killing of *Neisseria gonorrhoeae*, preventing MBL mediated activation of the complement system [32].

However, the most effective complement inhibitor is Factor H, a soluble protein regulator essential for controlling the alternative immune pathway [33]. Recognition and binding Factor H constitutes the infection mechanism adopted by several pathogens as a common immune evasion strategy. In addition several recent reports pointed out that besides its role as an alternative pathway downregulator, Factor H has an additional complement regulatory role in inhibiting activation of the classical pathway [34,35].

A totally different biological role can be ascribed to the second group of proteins identified by the LPS affinity capture strategy. HRG and Apolipoproteins, in fact, are known to exert protective effects against LPS induced systemic inflammation preventing inhibition of the complement immune system. HRG binds strongly to several complement protein inhibitors including Factor H and C4BP thus assisting in maintenance of normal immune function and enhancing complement activation [36].

S. pyogenes was shown to grow more efficiently in HRG-deficient plasma while the presence of overexpressed HRG greatly increased clots formation, bacterial entrapment and killing [37].

Apolipoproteins AI, A-IV and D are all components of the high density lipoproteins (HDL) that have long been reported to bind bacterial LPS neutralizing its toxicity and preventing initiation of innate immunity. More recently evidences for an Apo AI-LPS specific interaction have been obtained [38].

Moreover, the adenovirus mediated overexpression of this protein led to protection of mice against LPS-mediated systemic inflammation [39].

A further consideration on this study is the fact that most of the identified proteins are glycosylated, confirming previous data on glycoproteins involvement as PRRs (Patterns Recognition Receptors) in innate immunity response in plants and animals [5].

V.4.Conclusions

In conclusion, the novel LPS-mediated affinity capture strategy developed in this work proved to be very effective in specifically binding proteins involved either in inhibiting the activation of the complement system or in preventing bacterial infection by sequestering the LPS moiety. Both processes are of relevant biological significance and the results obtained makes sense with the bait used. Moreover, although a proper sensitivity test was not performed, the LPS affinity procedure led to the identification of low abundant serum proteins even without any depletion step. The sensitivity of the LC-MS/MS analyses allowed us to identify specific proteins even when the protein band showed only a very faint staining. A further advantage of this procedure concerns the possibility of determining both the proteins directly bound to LPS bait and other components involved in the functional complex but not linking the LPS moiety. It will be now possible to foresee applications of this strategy to investigate biological mechanisms exerted by pathogenic bacteria at the molecular level and to possibly define new mechanisms of infection.

V.5.References

- [1] Hug I, Feldman MF. Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology*. 2011 Feb;21(2):138-51.
- [2] Beutler BA. TLRs and innate immunity. *Blood*. 2009 Feb 12;113(7):1399-407.
- [3] Medzhitov R, Janeway C Jr. Innate immunity. *N Engl J Med*. 2000 Aug 3;343(5):338-44.
- [4] Iwasaki A, Medzhitov R. Toll-like receptor control of the adaptive immune responses. *Nat Immunol*. 2004 Oct;5(10):987-95.
- [5] Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell*. 2006 Feb 24;124(4):783-801.
- [6] Miller SI, Ernst RK, Bader MW. LPS, TLR4 and infectious disease diversity. *Nat Rev Microbiol*. 2005 Jan;3(1):36-46.
- [7] Caroff M, Karibian D, Cavaillon JM, Haeffner-Cavaillon N. Structural and functional analyses of bacterial lipopolysaccharides. *Microbes Infect*. 2002 Jul;4(9):915-26.
- [8] Su GL, Simmons RL, Wang SC. Lipopolysaccharide binding protein participation in cellular activation by LPS. *Crit Rev Immunol*. 1995;15(3-4):201-14.
- [9] Wurfel MM, Kunitake ST, Lichenstein H, Kane JP, Wright SD. Lipopolysaccharide (LPS)-binding protein is carried on lipoproteins and acts as a cofactor in the neutralization of LPS. *J Exp Med*. 1994 Sep 1;180(3):1025-35.

- [10] Wang ZQ, Xing WM, Fan HH, Wang KS, Zhang HK, Wang QW, Qi J, Yang HM, Yang J, Ren YN, Cui SJ, Zhang X, Liu F, Lin DH, Wang WH, Hoffmann MK, Han ZG. The novel lipopolysaccharide-binding protein CRISPLD2 is a critical serum protein to regulate endotoxin function. *J Immunol.* 2009 Nov 15;183(10):6646-56.
- [11] Nakamura T, Tokunaga F, Morita T, Iwanaga S. Interaction between lipopolysaccharide and intracellular serine protease zymogen, factor C, from horseshoe crab (*Tachypleus tridentatus*) hemocytes. *J Biochem.* 1988 Feb;103(2):370-4.
- [12] Tan NS, Ho B, Ding JL. High-affinity LPS binding domain(s) in recombinant factor C of a horseshoe crab neutralizes LPS-induced lethality. *FASEB J.* 2000 May;14(7):859-70.
- [13] Xu Z, Huang CX, Li Y, Wang PZ, Ren GL, Chen CS, Shang FJ, Zhang Y, Liu QQ, Jia ZS, Nie QH, Sun YT, Bai XF. Toll-like receptor 4 siRNA attenuates LPS-induced secretion of inflammatory cytokines and chemokines by macrophages. *J Infect.* 2007 Jul;55(1):e1-9.
- [14] Guillot L, Medjane S, Le-Barillec K, Balloy V, Danel C, Chignard M, Si-Tahar M. Response of human pulmonary epithelial cells to lipopolysaccharide involves Toll-like receptor 4 (TLR4)-dependent signaling pathways: evidence for an intracellular compartmentalization of TLR4. *J Biol Chem.* 2004 Jan 23;279(4):2712-8.
- [15] Poltorak A, Ricciardi-Castagnoli P, Citterio S, Beutler B. Physical contact between lipopolysaccharide and toll-like receptor 4 revealed by genetic complementation. *Proc Natl Acad Sci U S A.* 2000 Feb 29;97(5):2163-7.
- [16] Ohto U, Fukase K, Miyake K, Satow Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science.* 2007 Jun 15;316(5831):1632-4.
- [17] Andrä J, Gutschmann T, Müller M, Schromm AB. Interactions between lipid A and serum proteins. *Adv Exp Med Biol.* 2010;667:39-51.
- [18] Marshall JC. Lipopolysaccharide: an endotoxin or an exogenous hormone? *Clin Infect Dis.* 2005 Nov 15;41 Suppl 7:S470-80.
- [19] Mogensen TH. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev.* 2009 Apr;22(2):240-73, Table of Contents.
- [20] Liaw WJ, Tzao C, Wu JY, Chen SJ, Wang JH, Wu CC. Inhibition by terbutaline of nitric oxide and superoxide anion levels of endotoxin-induced organs injury in the anesthetized rat. *Shock.* 2003 Mar;19(3):281-8.
- [21] Kollipara RK, Perumal NB. Motif prediction to distinguish LPS-stimulated pro-inflammatory vs. antibacterial macrophage genes. *Immunome Res.* 2010 Sep 21;6:5.
- [22] Matsumoto M, Horiuchi Y, Yamamoto A, Ochiai M, Niwa M, Takagi T, Omi H, Kobayashi T, Suzuki MM. Lipopolysaccharide-binding peptides obtained by phage display method. *J Microbiol Methods.* 2010 Jul;82(1):54-8. Epub 2010 Apr 20. Erratum in: *J Microbiol Methods.* 2010 Oct;83(1):87.
- [23] Qian WJ, Jacobs JM, Camp DG 2nd, Monroe ME, Moore RJ, Gritsenko MA, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Smith RD. Comparative proteome analyses of human plasma following in vivo lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. *Proteomics.* 2005 Feb;5(2):572-84.

- [24] Li S, Wang L, Berman M, Kong YY, Dorf ME. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. *Immunity*. 2011 Sep 23;35(3):426-40.
- [25] Westphal O, Jann K. Bacterial lipopolysaccharides: extraction with phenol-water and further applications of the procedure. *Methods in Carbohydrate Chemistry*, 1965, 5, 83–91.
- [26] Deban L, Jaillon S, Garlanda C, Bottazzi B, Mantovani A. Pentraxins in innate immunity: lessons from PTX3. *Cell Tissue Res*. 2011 Jan;343(1):237-49.
- [27] Kim YG, Yang YH, Kim BG. High-throughput characterization of lipopolysaccharide-binding proteins using mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2010 Dec 1;878(31):3323-6.
- [28] Monti M, Cozzolino M, Cozzolino F, Tedesco R, Pucci P. Functional proteomics: protein-protein interactions in vivo. *Ital J Biochem*. 2007 Dec;56(4):310-4.
- [29] Oliver MA, Rojo JM, Rodríguez de Córdoba S, Alberti S. Binding of complement regulatory proteins to group A *Streptococcus*. *Vaccine*. 2008 Dec 30;26 Suppl 8:175-8.
- [30] Amdahl H, Jarva H, Haanperä M, Mertsola J, He Q, Jokiranta TS, Meri S. Interactions between *Bordetella pertussis* and the complement inhibitor factor H. *Mol Immunol*. 2011 Jan;48(4):697-705.
- [31] Ambrus G, Gál P, Kojima M, Szilágyi K, Balczer J, Antal J, Gráf L, Laich A, Moffatt BE, Schwaeble W, Sim RB, Závodszy P. Natural substrates and inhibitors of mannan-binding lectin-associated serine protease-1 and -2: a study on recombinant catalytic fragments. *J Immunol*. 2003 Feb 1;170(3):1374-82.
- [32] Gulati S, Sastry K, Jensenius JC, Rice PA, Ram S. Regulation of the mannan-binding lectin pathway of complement on *Neisseria gonorrhoeae* by C1-inhibitor and alpha 2-macroglobulin. *J Immunol*. 2002 Apr 15;168(8):4078-86.
- [33] Ferreira VP, Pangburn MK, Cortés C. Complement control protein factor H: the good, the bad, and the inadequate. *Mol Immunol*. 2010 Aug;47(13):2187-97.
- [34] Tan LA, Yu B, Sim FC, Kishore U, Sim RB. Complement activation by phospholipids: the interplay of factor H and C1q. *Protein Cell*. 2010 Nov;1(11):1033-49.
- [35] Tan LA, Yang AC, Kishore U, Sim RB. Interactions of complement proteins C1q and factor H with lipid A and *Escherichia coli*: further evidence that factor H regulates the classical complement pathway. *Protein Cell*. 2011 Apr;2(4):320-32.
- [36] Manderson GA, Martin M, Onnerfjord P, Saxne T, Schmidtchen A, Mollnes TE, Heinegård D, Blom AM. Interactions of histidine-rich glycoprotein with immunoglobulins and proteins of the complement system. *Mol Immunol*. 2009 Oct;46(16):3388-98.
- [37] Shannon O, Rydengård V, Schmidtchen A, Mörgelin M, Alm P, Sørensen OE, Björck L. Histidine-rich glycoprotein promotes bacterial entrapment in clots and decreases mortality in a mouse model of sepsis. *Blood*. 2010 Sep 30;116(13):2365-72.
- [38] Henning MF, Herlax V, Bakás L. Contribution of the C-terminal end of apolipoprotein AI to neutralization of lipopolysaccharide endotoxic effect. *Innate Immun*. 2011;17(3):327-37.

- [39] Li Y, Dong JB, Wu MP. Human ApoA-I overexpression diminishes LPS-induced systemic inflammation and multiple organ damage in mice. *Eur J Pharmacol.* 2008 Aug 20;590(1-3):417-22.

Publications

- 1) Magherini F, Pieri L, Guidi F, Giangrande C, Amoresano A, Bucciantini M, Stefani M, Modesti A. Proteomic analysis of cells exposed to prefibrillar aggregates of HypF-N. *Biochim Biophys Acta*. 2009 Aug;1794(8):1243-50.
- 2) Amoresano A, Carpentieri A, Giangrande C, Palmese A, Chiappetta G, Marino G, Pucci P. Technical advances in proteomics mass spectrometry: identification of post-translational modifications. *Clin Chem Lab Med*. 2009;47(6):647-65.
- 3) Carpentieri A, Giangrande C, Pucci P, Amoresano A. Glycoproteome study in myocardial lesions serum by integrated mass spectrometry approach: preliminary insights. *Eur J Mass Spectrom (Chichester, Eng)*. 2010;16(1):123-49.
- 4) Russo A, Siciliano G, Catillo M, Giangrande C, Amoresano A, Pucci P, Pietropaolo C, Russo G. hnRNP H1 and intronic G runs in the splicing control of the human rpl3 gene. *Biochim Biophys Acta*. 2010 May-Jun;1799(5-6):419-28.
- 5) Capparelli R, Nocerino N, Lanzetta R, Silipo A, Amoresano A, Giangrande C, Becker K, Blaiotta G, Evidente A, Cimmino A, Iannaccone M, Parlato M, Medaglia C, Roperto S, Roperto F, Ramunno L, Iannelli D. Bacteriophage-resistant *Staphylococcus aureus* mutant confers broad immunity against staphylococcal infection in mice. *PLoS One*. 2010 Jul 22;5(7):e11720.
- 6) Andrea Carpentieri, Chiara Giangrande, Piero Pucci and Angela Amoresano (2011). *A Proteomic Approach to Investigate Myocarditis*, Myocarditis, Daniela Cihakova (Ed.), ISBN: 978-953-307-289-0, InTech
- 7) Giangrande C, Melchiorre M, Marino G, Amoresano A. Oligosaccharides characterization by dansyl labelling and integrated mass spectrometry techniques.
Submitted to *Glycoconjugate Journal*.
- 8) Giangrande C, Colarusso L, Lanzetta R, Molinaro A, Pucci P, Amoresano A. Innate immunity probed by lipopolysaccharides affinity strategy and proteomics.
Submitted to *Molecular Biosystems*
- 9) Giangrande C, Melchiorre M, Chiaiese P, Palomba F, Filippone E, Marino G, Amoresano A. *Arabidopsis thaliana* glycoproteome characterization by integrated mass spectrometry techniques.
Manuscript in preparation.
- 10) Lettera V, Danilo Donnarumma, Chiara Giangrande, Angela Amoresano, Gennaro Marino and Giovanni Sannia. Functional analysis of the autoinducer 2-hydroxy-4-nitrobenzoic acid in the white-rot fungus *Pleurotus ostreatus*.
Manuscript in preparation.

Congress communications

- 1) TECHNICAL ADVANCES IN PROTEOMICS AND MASS SPECTROMETRY: IDENTIFICATION OF POST –TRANSLATIONAL MODIFICATIONS. Giangrande C, Amoresano C, Carpentieri A, Palmese A, Chiappetta G, Marino G, Pucci P. 4th ItPA Conference, Milano (22-25 giugno 2009)
- 2) ENDOGENOUS AUTOREGULATORS INFLUENCE ON PLEUROTUS OSTREATUS LIGNOLITIC SYSTEMS. Vincenzo Lettera, Chiara Giangrande, C. Del Vecchio, Angela Amoresano, Gennaro Marino and Giovanni Sanna. Lignobiotech One Symposium 2010, Reims (Francia) (28 marzo - 1 aprile 2010)
- 3) GLYCOPROTEOME CHARACTERIZATION BY INTEGRATED MASS SPECTROMETRY TECHNIQUES. Chiara Giangrande, Maura Melchiorre, Pasquale Chiaiese, Edgardo Filippone, Angela Amoresano. 5th ItPA Conference, Firenze (9- 12 giugno 2010) Winner of the Best poster prize ItPA 2010
- 4) GLYCOPROTEOME CHARACTERIZATION BY INTEGRATED MASS SPECTROMETRY TECHNIQUES. Chiara Giangrande, Maura Melchiorre, Pasquale Chiaiese, Edgardo Filippone, Angela Amoresano. 4th EUPA meeting and 6th ProCura meeting, Estoril (Portogallo) (23-27 ottobre 2010)
- 5) I PARTNERS MOLECOLARI DEL MODULATORE TRASCRIZIONALE DI E. coli AidB. Pamela Di Pasquale, Chiara Giangrande, Valentina Rippa, Angela Amoresano, Angela Duilio. Giornate Scientifiche 2010, Napoli (25-26 novembre 2010)
- 6) NEW INSIGHTS INTO INNATE IMMUNITY IN ARABIDOPSIS: DISCOVERY OF A LOS SIGNALLING NETWORK BY MULTIDISCIPLINARY APPROACHES. Proietti S, Bertini L, Amoresano A, Giangrande C, Caporale C, Caruso C. PR-Proteins and Induced Resistance Against Pathogens and Insects 2011. Neuchâtel (Svizzera) 4-8 september 2011)

Visiting appointment

June-July 2011

Dr. Joelle Vinh's laboratory. École supérieure de physique et de chimie industrielles de la ville de Paris



Proteomic analysis of cells exposed to prefibrillar aggregates of HypF-N

Francesca Magherini ^{a,*}, Laura Pieri ^a, Francesca Guidi ^a, Chiara Giangrande ^c, Angela Amoresano ^c,
Monica Bucciantini ^{a,b}, Massimo Stefani ^{a,b}, Alessandra Modesti ^a

^a Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, Italy

^b Research Centre on the Molecular Basis of Neurodegeneration, Università degli Studi di Firenze, Italy

^c Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Italy

ARTICLE INFO

Article history:

Received 16 February 2009

Received in revised form 20 April 2009

Accepted 21 April 2009

Available online 3 May 2009

Keywords:

2D-GE

MALDI-TOF

Prefibrillar protein aggregate

Protein expression profiling

HypF-N

ABSTRACT

Several human diseases are associated with the deposition of stable ordered protein aggregates known as amyloid fibrils. In addition, a large wealth of data shows that proteins not involved in amyloidoses, are able to form, in vitro, amyloid-like prefibrillar and fibrillar assemblies indistinguishable from those grown from proteins associated with disease. Previous studies showed that early prefibrillar aggregates of the N-terminal domain of the prokaryotic hydrogenase maturation factor HypF (HypF-N) are cytotoxic, inducing early mitochondria membrane depolarization, activation of caspase 9 and eventually cell death. To gain knowledge on the molecular basis of HypF-N aggregate cytotoxicity, we performed a differential proteomic analysis of NIH-3T3 cells exposed to HypF-N prefibrillar aggregates in comparison with control cells. Two-dimensional gel electrophoresis followed by protein identification by MALDI-TOF MS, allowed us to identify 21 proteins differentially expressed. The changes of the expression level of proteins involved in stress response (Hsp60 and 78 kDa glucose-regulated protein) and in signal transduction (Focal adhesion kinase1) appear particularly interesting as possible determinants of the cell fate. The levels of some of the differently expressed proteins were modified also in similar studies carried out on cells exposed to A β or α -synuclein aggregates, supporting the existence of shared features of amyloid cytotoxicity.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Amyloid diseases are a group of protein misfolding pathologies including either systemic forms (i.e. type II diabetes mellitus,) and neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington and prion diseases) (reviewed in [1,2]). The molecular basis of these clinically different pathologies can be traced back to the presence, in the affected tissues and organs, of proteinaceous deposits of fibrillar aggregates of one out of a number of peptides or proteins, each found aggregated specifically in each disease (reviewed in [1,3]). In the last ten years it has become increasingly clear that the ability to oligomerize into amyloid assemblies is not a specific feature of the proteins and peptides found aggregated in tissues affected by amyloid diseases; in fact, since 1998, an increasing number of reports support the idea that protein misfolding following mutations, chemical modifications, presence of destabilizing surfaces or any other alteration of the chemical environment, can result a structural reorganiza-

tion, favouring oligomerization/polymerization of peptides and proteins into amyloids [4,5].

In addition to amyloid aggregation, aggregate toxicity has also recently resulted as a generic property of proteins and peptides. In particular, it is increasingly recognized that amyloid oligomers, preceding the appearance of mature fibrils, known as prefibrillar aggregates, are the most toxic species, among amyloid assemblies associated or not associated with disease [6,7]. This view supports the idea that any protein can potentially become the source of toxic species impairing cell viability and that the cytotoxicity of prefibrillar aggregates results, at least in part, from shared basic structural features of the latter [8]. Moreover, a growing number of studies suggest that, in most cases, the cell membrane, can also favour protein/peptide misfolding favouring the appearance of aggregating nuclei [9]. Conversely, toxic prefibrillar aggregates from disease-associated or disease-unrelated proteins can interact with the cell membranes modifying their structural order resulting in the early impairment of ion and redox homeostasis [10,11]. Intense efforts are presently dedicated at unravelling the molecular basis of the appearance in tissue of protein aggregates and their cytotoxicity. However, much must still be learnt to gain enough knowledge to allow designing therapeutic strategies aimed at counteracting the clinical symptoms of amyloid diseases. On this respect, it can be important to investigate the cytotoxic effects on cells exposed to amyloid aggregates of proteins/peptides not associated with any amyloid

Abbreviations: 2D-GE, two dimensional gel electrophoresis; IEF, isoelectrofocusing; MS, mass spectrometry; MALDI-TOF, matrix assisted laser desorption/ionization time-of-flight; LC-MS/MS, liquid chromatography, tandem mass spectrometry; CHAPS, 3-[(3cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; ECL, Enhanced Chemiluminescence; HypF-N, hydrogenase maturation factor; HypF-N, N-terminal domain of the prokaryotic hydrogenase maturation factor HypF

* Corresponding author.

E-mail address: Francesca.magherini@unifi.it (F. Magherini).

disease. Such study can highlight cell modifications resulting from shared structural features in amyloids excluding those associated with the specific features of any aggregated peptide/protein.

The advent of proteomics has allowed the simultaneous analysis of changes in the expression pattern of multiple proteins in complex biological systems. This appears particularly important in the case of cell dysfunctions resulting from protein aggregation. The effects protein aggregates have on cells appear highly complex and heterogeneous. Such a complex pattern of cell impairment makes proteomics one of the most useful tools to integrate these modifications into a whole systematic picture. The reports that recently appeared on the proteomic analysis of amyloid diseases such as Alzheimer's disease, Parkinson's disease and others have provided valuable data on some cell modifications allowing to explain, at least in part, cell impairment in these diseases [12–16]. Proteomic studies provided useful information on the changes in pattern of protein expression in cells exposed to toxic aggregates of specific peptides/proteins found aggregated in the corresponding diseases.

In the present study we performed a proteomic analysis of NIH-3T3 cells exposed to toxic aggregates of a protein domain not involved in any amyloid disease, the N-terminal domain of the prokaryotic HypF hydrogenase maturation factor (HypF-N) whose aggregation properties and aggregate cytotoxicity were previously characterized [8,10,17]. To our knowledge, this is the first proteomic study on the alterations of the protein expression profiles in cells exposed to toxic amyloid aggregates of a disease-unrelated protein.

Our results highlight some generic changes in protein expression pattern elicited by the shared features of amyloids such as the basic cross-beta structure and the exposure of hydrophobic patches. We found significant differences in the protein expression patterns in the exposed cells, with a number of up- or down-regulated proteins. Some of these proteins were also found in similar studies carried out on cells exposed to A β or α -synuclein aggregates in agreement with the generic nature of the cellular changes underlying amyloid cytotoxicity. Among the differentially expressed proteins, the reduced expression of Fak1 observed in the exposed cells can be related to the apoptotic process, whereas the increased expression of Hsp60 can provide protection against cell stress induced by HypF-N prefibrillar aggregates. Furthermore the treated cells showed a marked increase of the expression of both glyceraldehyde 3-phosphate dehydrogenase (Gapdh) and enolase. These two proteins are involved in energy metabolism and Gapdh has also been shown to bind the β -amyloid precursor protein [18] and to be involved in transcriptional regulation of cell-cycle [19]. Finally, we observed an increase in the expression level of actin as previously shown in cell exposed to the intracellular domain of the β -amyloid precursor protein [20].

2. Materials and Methods

2.1. HypF-N expression and purification

HypF-N was expressed and purified as previously described [17]. HypF-N prefibrillar aggregates were obtained by incubating the protein for 48 h at room temperature at a concentration of 0.3 mg/ml in 30% (v/v) trifluoroethanol, 50 mM sodium acetate, 2 mM dithiothreitol (DTT), pH 5.5, as previously reported [17]. At the end of the incubation the solution was centrifuged, and the resulting pellet was dried under N₂ to remove the residual solvent, dissolved in DMEM at 200 μ M (monomeric protein concentration) and immediately added to the cell culture medium at 2 μ M final concentration.

2.2. Cell culture and treatment

Cell culture media and other reagents, unless otherwise stated, were from Sigma-Aldrich Fine Chemicals Co. NIH-3T3 murine

fibroblasts (ATCC, Manassas, VA) were routinely cultured in Dulbecco's modified Eagle's medium (DMEM) with 4.5 g/l glucose, containing 10% bovine calf serum (HyClone Lab, Perbio Company, Celbio), 3 mM glutamine, 100 U/ml penicillin and 100 μ g/ml streptomycin, in a 5% CO₂ humidified environment at 37 °C. Cells were used for a maximum of 10 passages. Sub-confluent NIH-3T3 cells were treated for differing lengths of time with 2 μ M toxic HypF-N prefibrillar aggregates. Under these conditions it was previously shown that the aggregates are stable in the culture media [8]. Controls were performed by exposing the cells to the same amount of native, soluble HypF-N. At the end of each treatment, the cells were washed twice with phosphate-buffered saline (PBS), dried and stored at –80 °C.

2.3. Sample preparation and 2D-GE

Cells were scraped in RIPA buffer (50 mM Tris-HCl pH 7.0, 1% NP-40, 150 mM NaCl, 2 mM EGTA, 100 mM NaF) containing a cocktail of protease inhibitors (Sigma). The cells were sonicated (10 s) and protein extracts were clarified by centrifugation at 8000 g for 10 min. Proteins were precipitated following a chloroform/methanol protocol [21] and the pellet was resuspended in 8 M urea, 4% CHAPS and 20 mM DTT. Three independent experiments were performed and each sample was run in triplicate in order to assess biological and analytical variation. IEF (first dimension) was carried out on non-linear wide-range immobilized pH gradients (pH 3.0–10; 18 cm long IPG strips; GE Healthcare, Uppsala, Sweden) and achieved using the Ettan™ IPGphor™ system (GE Healthcare, Uppsala, Sweden). Analytical-run IPG strips were rehydrated with 60 μ g of total proteins in 350 μ l of lysis buffer and 0.2% carrier ampholyte for 1 h at 0 V and for 8 h at 30 V, at 20 °C. MS-preparative IPG strips were loaded with 400 μ g of proteins. The strips were focused at 20 °C according to the following electrical conditions: 200 V for 1 h, from 300 V to 3500 V in 30 min, 3500 V for 3 h, from 3500 V to 8000 V in 30 min, and 8000 V until a total of 80,000 V/h was reached. After focusing, analytical and preparative IPG strips were equilibrated for 12 min in 6 M urea, 30% glycerol, 2% SDS, 2% DTT in 0.05 M Tris-HCl buffer, pH 6.8, and subsequently for 5 min in the same urea/SDS/Tris buffer solution where DTT was substituted with 2.5% iodoacetamide. The second dimension was carried out on 9–16% polyacrylamide linear gradient gels (18 cm \times 20 cm \times 1.5 mm) at 10 °C and 40 mA/gel constant current until the dye front reached the bottom of the gel. Analytical gels were stained with ammoniacal silver nitrate as previously described [22]; MS-preparative gels were stained with colloidal Coomassie [23].

2.4. Western blotting analysis of proteomic candidates

For 1-DE 30 μ g of protein extracts was separated by 12% SDS-PAGE and transferred onto a PVDF membrane (Millipore). To confirm the results obtained from 2D-GE analysis, the relative amount of Hsp60 and Fak proteins were assessed by Western blot with appropriate antibodies (Santacruz). For quantification, the blots were stained with Coomassie brilliant blue R-250 and subjected to densitometric analysis performed using Quantity One Software (Bio-Rad). Statistical analysis of the data was performed by Student's *t*-test; *p*-values 0.05 were considered statistically significant. The intensity of the immunostained bands were normalized with the total protein intensities measured by Coomassie brilliant blue R-250 from the same blot.

2.5. Image analysis and statistics

The gel and Western blot images were acquired with an Epson expression 1680 PRO scanner. For each condition, three biological replicates were performed and only the spots present in all the replicates were taken in consideration for subsequent analysis. Computer-aided 2D image analysis was carried out using ImageMaster 2-DE Platinum software version 6.0 (GE Healthcare). The

relative spot volume calculated as %V (V single spot/ V total spots, where V =integration of OD over the spot area) was used for quantitative analysis in order to decrease experimental errors. The normalized intensity of the spots on replicate 2D gels was averaged and standard deviation was calculated for each condition. A two-tiled non paired Student's t -test was performed using ORIGIN 6.0 (Microcal Software, Inc.) to determine whether the relative change was statistically significant.

2.6. In-gel trypsin digestion and MALDI-TOF mass spectrometry

The analysis was performed on the Coomassie blue-stained spots excised from the gels. The spots were washed first with acetonitrile and then with 0.1 M ammonium bicarbonate. Protein samples were reduced by incubation in 10 mM dithiothreitol (DTT) for 45 min at 56 °C. The cysteines were alkylated by incubation in 5 mM iodoacetamide for 15 min at room temperature in the dark. The gel particles were then washed with ammonium bicarbonate and acetonitrile. Enzymatic digestion was carried out with trypsin (12.5 ng/ μ l) in 50 mM ammonium bicarbonate buffer, pH 8.5, at 4 °C for 4 h. The buffer solution was then removed and a new aliquot of the enzyme/buffer solution was added for 18 h at 37 °C. A minimum reaction volume, enough for complete gel rehydration was used. At the end of the incubation the peptides were extracted by washing the gel particles with 20 mM ammonium bicarbonate and 0.1% TFA in 50% acetonitrile at room temperature and then lyophilised. Positive Reflectron MALDI spectra were recorded on a Voyager DE STR instrument (Applied Biosystems, Framingham, MA). The MALDI matrix was prepared by dissolving 10 mg of alpha cyano in 1 ml of acetonitrile/water (90:10 v/v). Typically, 1 μ l of matrix was applied to the metallic sample plate and then 1 μ l of analyte was added. Acceleration and reflector voltages were set up as follows: target voltage at 20 kV, first grid at 70% of target voltage, delayed extraction at 100 ns to obtain the best signal-to-noise ratios and the best possible isotopic resolution with multipoint external calibration using a peptide mixture purchased from Applied Biosystems. Each spectrum represents the sum of 1500 laser pulses from randomly chosen spots per sample position. Raw data were analyzed using the computer software provided by the manufacturers and are reported as monoisotopic masses.

2.7. nanoLC mass spectrometry

A mixture of peptide solution was subjected to LC-MS analysis using a 4000Q-Trap (Applied Biosystems) coupled to an 1100 nano HPLC system (Agilent Technologies). The mixture was loaded on an Agilent reverse-phase pre-column cartridge (Zorbax 300 SB-C18, 5 \times 0.3 mm, 5 μ m) at 10 μ l/min (A solvent 0.1% formic acid, loading time 5 min). The peptides were separated on an Agilent reverse-phase column (Zorbax 300 SB-C18, 150 mm \times 75 μ m, 3.5 μ m), at a flow rate of 0.3 μ l/min with a 0% to 65% linear gradient in 60 min (A solvent 0.1% formic acid, 2% acetonitrile in MQ water; B solvent 0.1% formic acid, 2% MQ water in acetonitrile). Nanospray source was used at 2.5 kV with liquid coupling, with a declustering potential of 20 V, using an uncoated silica tip from NewObjectives (O.D. 150 μ m, I.D. 20 μ m, T.D. 10 μ m). The data were acquired in information-dependent acquisition (IDA) mode, in which a full scan mass spectrum was followed by MS/MS of the 5 most abundant ions (2 s each). In particular, spectra acquisition of MS-MS analysis was based on a survey Enhanced MS Scan (EMS) from 400 m/z to 1400 m/z at 4000 amu/s. This scan mode was followed by an Enhanced Resolution experiment (ER) for the five most intense ions and then MS2 spectra (EPI) were acquired using the best collision energy calculated on the basis of m/z values and charge state (rolling collision energy) from 100 m/z to 1400 m/z at 4000 amu/s. The data were acquired and processed using the Analyst software (Applied Biosystems).

2.8. MASCOT analysis

The spectral data were analyzed using the Analyst software (version 1.4.1) and the MS-MS centroid peak lists were generated using the MASCOT.dll script (version 1.6b9). The MS-MS centroid peaks were threshold at 0.1% of the base peak. MS-MS spectra with less than 10 peaks were rejected. The spectra were searched against the Swiss Prot database (2006.10.17 version) using the licensed version of Mascot 2.1 (Matrix Science), after converting the acquired MS-MS spectra in MASCOT generic file format. The MASCOT search parameters were: taxonomy mus musculus; allowed number of missed cleavages 2; enzyme trypsin; variable post-translational modifications, methionine oxidation, pyro-glu N-term Q; peptide tolerance 200 ppm and MS/MS tolerance 0.6 Da; peptide charge, from +2 to +3 and top 20 protein entries. Spectra with a MASCOT score <25 having low quality were rejected. The score used to evaluate the quality of matches for the MS-MS data was higher than 30. However, the spectral data were manually validated and contained sufficient information to assign peptide sequence.

3. Results

3.1. Comparative proteomic analysis between control cells and cells exposed to HypF-N prefibrillar aggregates

Previous experiments performed by Bucciantini et al. showed that prefibrillar HypF-N aggregates induced early Ca^{2+} increase and oxidative stress followed by mitochondria depolarization and caspase activation in exposed NIH-3T3 cells. After 24 h, the cells died with necrotic features possibly since the ATP levels were too low to sustain the initially triggered apoptotic program [10]. In order to investigate the changes (if any) in protein expression induced in the same cells upon exposure to HypF-N prefibrillar aggregates, we performed a 2D-GE followed by mass spectrometry. In all the experiments carried out in this study, the cells were exposed to 2 μ M prefibrillar aggregates. This protein concentration was chosen to investigate finely regulated biochemical processes, such as the activation of pro-apoptotic factors, which could be hidden by a stronger cell injury [10]. The cells were treated for 5 and 24 h and proteins extracts were prepared as described under [Materials and methods](#). Then the proteins were separated by 2D-GE and the resulting silver-stained gels were analyzed using the ImageMaster 2D Platinum 6.0 software. The differences of protein expression between control and treated cells were taken into consideration if the relative volume of the spots differed reproducibly more than 1.5-fold and this difference was statistically significant. An average of about 1300 spots was detected in each silver-stained gel. Cell exposure to HypF-N prefibrillar aggregates did not affect the overall proteomic profiles both after 5 h and after 24 h (Figs. 1 and 2). However, the computer analysis highlighted 19 variations between cells treated for 5 h with 2 μ M HypF-N prefibrillar aggregates (Fig. 1B) and the control cells treated for the same length of time with an equivalent amount of native HypF-N (Fig. 1A). Among these variations, only two were still present after 24 h of treatment with aggregates (Fig. 2) indicating that the alteration of the expression is a transient event, at least for this group of proteins, except two. On the other hand, the comparison between cells treated for 24 h with 2 μ M HypF-N prefibrillar aggregates and control cells showed a variation of 9 spots, whose expression was not affected after 5 h, indicating that some proteins are up- or down-regulated as a consequence of the prolonged exposure to the aggregates.

3.2. Identification of differentially expressed proteins

In order to identify the proteins of interest, 400 μ g of protein lysates was loaded on preparative gels and stained with colloidal

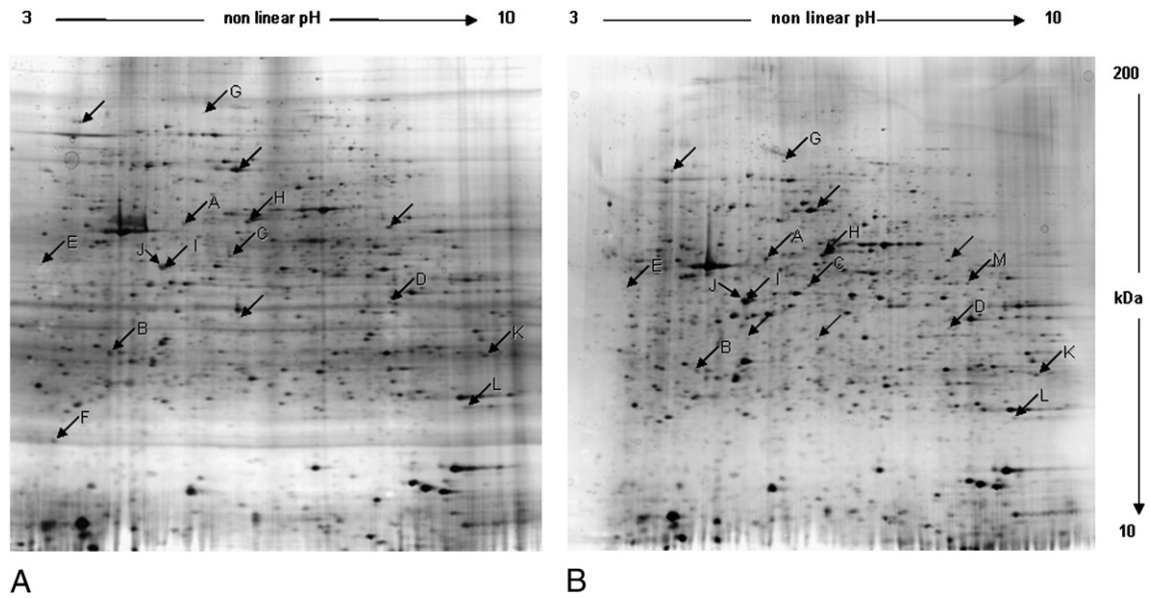


Fig. 1. 2D images of silver stained gels of total proteins extracted from control and HypF-N prefibrillar aggregate treated cells for 5 h. Arrows indicate variations between control (panel A) and treated cells (panel B). Letters indicate the identified proteins. Controls were performed by exposing the cells to native, soluble HypF-N.

Coomassie. The spots indicated by arrows in representative gels shown in **Figs. 1 and 2** were selected for mass spectral identification after merging the images of preparative and analytical gels. The proteins excised from the gels were reduced, alkylated and in situ digested with trypsin. The resulting peptide mixtures were directly analyzed by MALDI/MS according to the peptide mass fingerprinting procedure. The peaks detected in the MALDI spectra were used to search for a non redundant sequence database using the in house MASCOT software, taking advantage of the specificity of trypsin and the taxonomic category of the samples. The number of measured masses that matched within the given mass accuracy of 200 ppm was recorded and the proteins that received the highest number of peptide matches were examined. Some spots could not be identified unambiguously either due to the low protein content of the spot or to the presence of more than one protein per spot. Among the 19 spots differentially expressed after 5 h of cell exposure to the aggregates, 13 spots were

successfully identified and are indicated by arrows and letters in **Fig. 1**. Among the 9 spots differentially expressed in cells treated for 24 h with 2 μ M HypF-N prefibrillar aggregates, 8 spots were identified and are indicated by arrows and numbers in **Fig. 2**. Some spots gave no confident identification by the peptide mass fingerprinting procedure. Additional data were then provided by nanoLC/MS/MS experiments. The peptide mixtures were fractionated by nanoHPLC and sequenced by tandem mass spectrometry leading to the unambiguous identification of the protein candidate. The lists of proteins identified by these approaches are reported in **Table 1** (5 h treatment) and in **Table 2** (24 h treatment). The identified proteins included cytoskeleton elements (actin, tubulin alpha 1C chain, microtubule-actin cross-linking factor 1), enzymes involved in energy metabolism and transcriptional regulation (Gapdh, enolase), proteins involved in stress response (Hsp60 and 78 kDa glucose-regulated protein) and the focal adhesion kinase, Fak1. Among these proteins only Fak1 and Hsp60 showed an

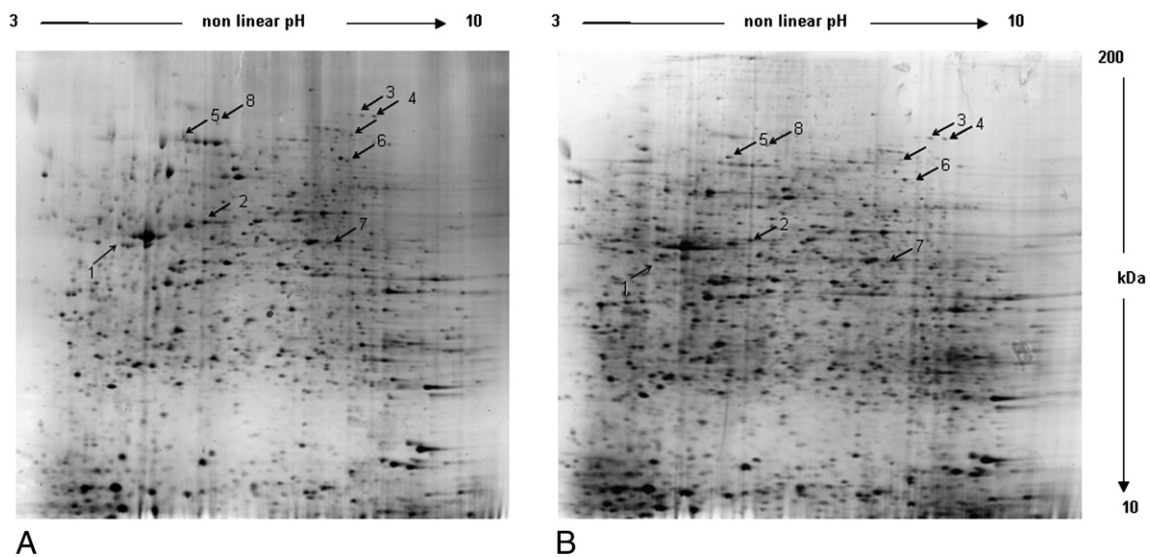


Fig. 2. 2D images of silver stained gels of total proteins extracted from control and HypF-N prefibrillar aggregate treated cells for 24 h. Arrows indicate variations between control (panel A) and treated cells (panel B). Numbers indicate the identified proteins. Controls were performed by exposing the cells to native, soluble HypF-N.

Table 1

Relative change in protein expression in cells treated for 5 h with HypF-N prefibrillar aggregates versus control cells.

Spot	Protein name	Accession number	MASCOT score	No. of matched peptides	Sequence coverage (%)	Fold change	p-value	Functional categorization
<i>Up-regulated by HypF-N</i>								
A	Hsp60	P63038	162	5	12%	+ 1.58	0.0005	Protein folding and stress response
B	Actin	P60710	79	5	22%	Only in treated cells.	N.D.	Cytoskeletal organization
C	Actin	P60710	241	7	18%	+ 9.8	0.007	Cytoskeletal organization
D	Glyceraldehyde-3-phosphate dehydrogenase	P16858	120	7	33%	+ 9.7	0.0081	Energy related (glycolysis)
E	78 kDa glucose-regulated protein	P20029	78	2	5%	+ 4.8	0.1	Protein folding and stress response (response to unfolded proteins)
F	Nucleophosmin	Q61937	165	5	15%	Only in treated cells.	N.D.	Associated with nucleolar ribonucleoprotein structures and bind single-stranded nucleic acids
<i>Down-regulated by HypF-N</i>								
G	Focal Adhesion kinase-1	P34152	77	17	22%	- 1.6	0.01	Non-receptor protein-tyrosine kinase involved in cell motility, Proliferation and apoptosis.
H	Enolasi-1	P17182	118	12	32%	- 2.4	0.09	Energy related (glycolysis)
I	Tubulin alpha-1C chain	P68373	80	9	31%	- 8.3	0.015	Major constituent of microtubules
J	Annexin-A3	O35639	111	13	43%	- 2	0.019	Inhibitor of phospholipase A2
K	Hsp60	P63038	85	2	5%	- 1.6	0.032	Protein folding and stress response
L	Transgelin-2	Q9WVA4	171	7	41%	Only in control cells	N.D.	Belongs to the calponin family
M	Heparan sulfate glucosamine 3-O-sulfotransferase 2	Q673U1	47	2	4%	Only in control cells	N.D.	Catalyzes the O-sulfation of glucosamine

expression variation persisting over time, since it was observed both after 5 and after 24 h of cell exposure to the aggregates.

3.3. Validation of proteomics results

In order to validate the proteomic results, the amounts of Fak1 and Hsp60 were evaluated by Western blot analysis with specific antibodies as shown in Fig. 3, panels A and B, respectively. Thirty µg of proteins was loaded on 12% SDS-PAGE and transferred onto a PVDF membrane. For quantification, the intensities of the immunostained bands were normalized to the total protein intensities in the same blot, as measured by Coomassie brilliant blue. In Fig. 3, panel C the histograms representing the variation of the expression of Fak1 and Hsp60 are also reported. Such analysis confirmed the decrease of Fak1 expression and the increase of Hsp60 expression both after 5.0 and 24 h.

3.4. Protein expression changes in cells exposed for 5 h to HypF-N prefibrillar aggregates

In cells exposed to HypF-N prefibrillar aggregates for 5 h, 6 spots (A to F in Table 1) appeared up-regulated, whereas 7 spots (G to M in

Table 1) appeared down-regulated. Among the up-regulated proteins, we found the heat shock protein Hsp60 (spot A). Hsp60 belongs to a family of highly homologous chaperone proteins that are induced in response to environmental, physical and chemical stresses, including accumulations of misfolded proteins and reactive oxygen species [24, 25]. The increase of Hsp60 expression limits the consequences of damage facilitating cell recovery. Hsp60 was also identified in the spot K, which was down-regulated upon cell treatment with the aggregates. However, the position in the gel and the peptide coverage, indicates that this spot is probably a fragment arising from a proteolytic cleavage of Hsp60.

A further indication of a stress condition induced in cells exposed to the HypF-N aggregates is the marked increase in the expression of the key glycolytic enzyme Gapdh (spot D). This protein plays a central role in glycolysis, catalyzing the reversible conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate. More recent studies have highlighted unexpected non-glycolytic functions of Gapdh in physiological and pathological processes, including transcriptional regulation of cell-cycle [19]. In addition two spots corresponding to actin were up-regulated. Among the proteins whose expression appeared decreased upon exposure to HypF-N prefibrillar aggregates, the focal adhesion kinase (Fak1; spot G) is particularly interesting.

Table 2

Relative change in protein expression in cells treated for 24 h with HypF-N prefibrillar aggregates versus control cells.

Spot	Protein name	Accession number	MASCOT score	No. of matched peptides	Sequence coverage (%)	Fold change	p-value	Functional categorization
<i>Up-regulated by HypF-N</i>								
1	Ovostatin homolog	Q3UU35	48	4	15%	+ 1.4	0.028	Proteases inhibitor
2	Hsp60	P63038	162	5	12%	+ 1.58	0.0005	Protein folding and stress response
<i>Down-regulated by HypF-N</i>								
3	Microtubule-actin cross-linking factor 1	Q9QXZ0	82	14	10%	- 1.66	0.035	F-actin-binding protein which may play a role in cross-linking actin to other cytoskeletal proteins
4	Microtubule-actin cross-linking factor	Q9QXZ0	88	10	8%	- 2.24	0.023	
5	Pol protein	Q7M6W3	73	10	17%	- 1.67	0.1	Unknown
6	Pdia2 protein	Q14AV9	43	2	5%	- 3.21	0.004	Protein disulfide isomerase
7	Poly(rC)-binding protein 2	Q61990	71	2	6%	- 3.78	0.014	Single-stranded nucleic acid binding protein
8	Fak1 focal adhesion kinase	P34152	77	17	22%	- 1.53	0.015	Non-receptor protein-tyrosine kinase involved in cell motility, proliferation and apoptosis

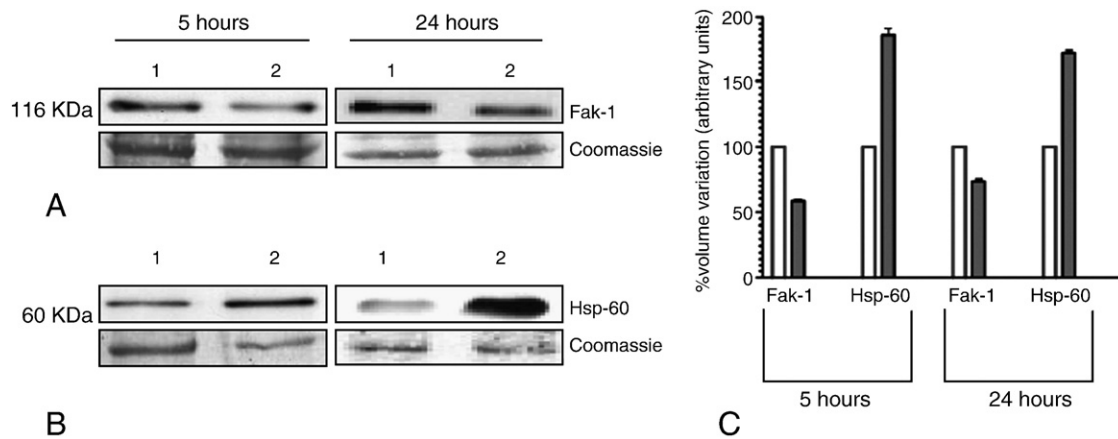


Fig. 3. Validation of proteomic results by western blot analysis. Western blot were probed with antibodies against Fak1 and Hsp60 proteins identified by proteomic screening. The intensity of immunostained bands was normalized with the total protein intensities measured from the same blot stained with Coomassie brilliant blue (in panel A and panel B a representative band of the lane is reported). (A) Aggregate-induced reduced expression of Fak1 after 5 and 24 h of treatment: lane 1, cells exposed to native HypF-N; lane 2, cells exposed to HypF-N prefibrillar aggregates. (B) Aggregate-induced increased expression of Hsp60 after 5 and 24 h of treatment: lane 1, cells exposed to native HypF-N; lane 2, cells exposed to HypF-N prefibrillar aggregates. (C) Histograms representing Fak1 and Hsp60 protein expression variation. The two-tailed non paired Student's *t*-test was performed using ORIGIN 6.0. ($p < 0.05$).

Fak1 is a non-receptor cytoplasmic tyrosine kinase that plays a key role in the regulation of proliferation and migration of normal and tumour cells [26, 27]. Interestingly, Fak1 and Hsp60 are the only two proteins, among those differentially expressed after 5 h of cell exposure, that do not recover a normal expression level after 24 h.

3.5. Protein expression changes in cells exposed for 24 h to HypF-N prefibrillar aggregates

It was previously shown that NIH-3T3 cells treated for 24 h with 10 μ M HypF-N prefibrillar aggregates die with necrotic features, including cytoplasmic vacuolization and nuclear swelling after an initial apoptotic activation [10]. The amount of HypF-N aggregates (2 μ M, soluble protein concentration) used in this work is not so high to induce cell death, thus allowing the cells to overcome damage. After 24 h of cell exposure to HypF-N aggregates, few other proteins, besides Fak1 and Hsp60, displayed altered expression. In particular, we identified 6 new spots (1 up-regulated and 5 down-regulated). Among the proteins down-regulated, the microtubule-actin cross-linking factor 1 (MACF1, spot 3) belongs to the Plakin family, that includes proteins involved in the linkage of cytoskeletal elements and the junctional complex. MACF1 was found to regulate microtubule remodelling in response to the activation of signal transduction

pathways, although its function has not yet been fully explored [28, 29].

4. Discussion

To our knowledge, this study is the first proteomic investigation focused on highlighting the alterations of the protein expression profiles in a cultured cell model exposed to toxic amyloid aggregates of a protein not involved in any amyloid disease. Our analysis was performed using a non lethal dose of HypF-N prefibrillar aggregates, allowing the detection of fine variations more directly implicated in a response to the cell injury given by the aggregates, instead of the complex pattern of changes arising during the process of cell death.

Our approach led us to identify a subset of cell proteins whose levels were significantly altered upon cell exposure to the aggregates for 5 h or 24 h. Some of the proteins detected in our investigation, including Hsp60, actin, enolase-1 and Gapdh had previously been identified in other proteomic studies carried out on cells exposed to A β 42 or α -synuclein. In Table 3 a comparison between the protein identified in our study and the proteins identified in other studies is shown, indicating that there is a general response of cells to toxic aggregates that is not sequence specific [12,13,30,31,32]. Actually, changes in the expression levels of Gapdh, actin, tubulin and heat shock proteins have frequently

Table 3

Protein expression changes observed in this study in comparison to previous protein reported in other studies.

Protein	Amyloid aggregates	Cell type or animal model	
GAPDH	HypF-N Amyloid β -peptide	NIH3T3 fibroblast cells	This study
		Mitochondria from primary neuron	[30]
		Tg2576 transgenic mice	[31]
Tubulin α -chain	β -amyloid (1–42) HypF-N β -amyloid (1–42)	3xTgAD Alzheimer's mice (hippocampal protein)	[12]
		P301L tau overexpressing SH-SY5Y	[32]
		NIH3T3 fibroblast cells	This study
		SN56.B5.G4 cholinergic cells	[13]
		3xTgAD Alzheimer's mice (hippocampal protein)	[12]
78 kDa glucose-regulated protein	β -amyloid (1–42) HypF-N	Amygdalae from P301L tau transgenic mice	[32]
		NIH3T3 fibroblast cells	This study
		3xTgAD Alzheimer's mice (hippocampal protein)	[12]
Enolase	β -amyloid (1–42) HypF-N	P301L tau overexpressing SH-SY5Y	[32]
		NIH3T3 fibroblast cells	This study
		Tg2576 transgenic mice	[31]
		3xTgAD Alzheimer's mice (hippocampal protein)	[12]
		3xTgAD Alzheimer's mice	[12]
Hsp60	HypF-N	NIH3T3 fibroblast cells	This study
		Amygdalae from P301L tau transgenic mice	[32]
Actin	HypF-N β -amyloid (1–42)	NIH3T3 fibroblast cells	This study
		Amygdalae from P301L tau transgenic mice	[32]

been reported in amyloid-linked proteomic studies possibly because they are related to a generic response to stress conditions [33] such as that associated with the growth of amyloid aggregates. In our experiments, cell exposure to toxic amyloid aggregates induced an increase of the expression levels of several proteins such as Hsp60. In addition to its chaperone activity, Hsp60 has been suggested to perform complex functions, producing both anti- and pro-apoptotic effects. In fact, cytosolic Hsp60, can promote either cell survival or caspase-mediated cell death by preventing the translocation of the pro-apoptotic protein Bax into the mitochondria or by favouring the maturation of procaspase-3, respectively [34–38]. In a recent study, Hsp60, Hsp70, and Hsp90 were shown to provide differential protection against intracellular stress caused by β -amyloid by maintaining the efficiency of the mitochondrial oxidative phosphorylation and the tricarboxylic acid cycle enzymes. In particular, Hsp60 was shown to prevent the inhibition of complex IV activity by β -amyloid, thus preventing apoptosis [39].

We also found significantly increased levels of Gapdh and actin. As far as Gapdh is concerned, several recent studies have shown that, in addition to glycolysis, it is involved in several glycolysis-unrelated activities; these include a role in vesicle fusion and transport [40], microtubule bundling [41], nuclear RNA transport [42], and transcription [43]. Furthermore, increased expression and nuclear translocation of Gapdh have recently been reported to participate to the apoptotic pathway in different cell types [44–47]. Finally, Gapdh has also been reported to bind to a variety of proteins involved in neuronal diseases, including the amyloid precursor protein and huntingtin [18]. The increased levels of actin expression in cells exposed to prefibrillar HypF-N aggregates are similar to the effects previously reported by Muller et al.; these authors found up-regulation of the actin gene expression in cells harbouring the cytoplasmic domain of the amyloid precursor protein [20].

Finally, we found a significant decrease of Fak1 in cells exposed both 5 h and 24 h to the HypF-N aggregates. The repression of Fak1 synthesis in exposed cells, confirmed by Western blot analysis, is one of the major results of this proteomic analysis. In vivo animal studies have shown that Fak1 expression is increased in a number of human cancers, thus contributing to tumour development and malignancy [48]. Moreover, Fak1 has recently been shown to be a critical protein in survival signalling, since it blocks apoptosis induced by several stimuli [49]. Fak1 expression decrease following proteolytic cleavage in various cell types has been associated with various cell dysfunctions including c-Myc-induced apoptosis of chicken embryo fibroblasts (CEF) [49], growth factor deprivation-induced apoptosis of human umbilical vein endothelial cells [50], and detachment-induced cell death (anoikis) of intestinal epithelial cells [51]. In a recent study, different epithelial cell lines treated with thimerosal displayed increased levels of hydrogen peroxide resulting in caspase activation, Fak1 cleavage and apoptosis [52]. All these observations suggest that the decreased Fak1 expression in NIH-3T3 cells treated with HypF-N aggregates could be related to the apoptotic process: in particular its decreased intracellular levels could be the consequence of a proteolytic cleavage making the cells more vulnerable to death.

The reported changes in protein expression profiles in exposed cells suggest some alterations in specific signalling pathways involved in the control of gene transcription and translation and/or in protein degradation pathways. These alterations could be triggered, at least in part, by modifications of signalling pathways following the reported interaction of amyloid aggregates with the cell membrane [10, 11, 53]. Overall, our results can provide useful information on the crucial events underlying cytotoxicity induced by amyloid aggregates of different peptides or proteins both related and unrelated to disease.

Acknowledgements

This work was supported by grants from the INBB, the Italian Ministero dell'Università e della Ricerca Scientifica (PRIN 2006 and 2007) and the Ente Cassa di Risparmio di Firenze. Support from the

National Center of Excellence in Molecular Medicine (MIUR – Rome) and from the Regional Center of Competence (CRdC ATIBB, Regione Campania – Naples) is gratefully acknowledged.

References

- [1] M. Stefani, C.M. Dobson, Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution, *J. Mol. Med.* 81 (2003) 678–699.
- [2] M. Stefani, Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world, *Biochim. Biophys. Acta* 1739 (2004) 5–25.
- [3] F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* 75 (2006) 333–366.
- [4] J.I. Guijarro, M. Sunde, J.A. Jones, I.D. Campbell, C.M. Dobson, Amyloid fibril formation by an SH3 domain, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 4224–4228.
- [5] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, C.M. Dobson, Designing conditions for in vitro formation of amyloid protofilaments and fibrils, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 3590–3594.
- [6] C.G. Glabe, Common mechanisms of amyloid oligomer pathogenesis in degenerative diseases, *Neurobiol. Aging* 27 (2006) 570–575.
- [7] C.G. Glabe, R. Kaye, Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis, *Neurology* 66 (2006) 74–78.
- [8] M. Bucciantini, G. Calloni, F. Chiti, L. Formigli, D. Nosi, C.M. Dobson, M. Stefani, Prefibrillar amyloid protein aggregates share common features of cytotoxicity, *J. Biol. Chem.* 279 (2004) 31374–31382.
- [9] M. Zhu, P.O. Souillac, C. Ionescu-Zanetti, S.A. Carter, A.I. Fink, Surface-catalyzed amyloid fibril formation, *J. Biol. Chem.* 277 (2002) 50914–50922.
- [10] M. Bucciantini, S. Rigacci, A. Berti, L. Pieri, C. Cecchi, D. Nosi, L. Formigli, F. Chiti, M. Stefani, Patterns of cell death triggered in two different cell lines by HypF-N prefibrillar aggregates, *FASEB J.* 19 (2005) 437–439.
- [11] J.I. Kourie, Mechanisms of amyloid beta protein-induced modification in ion transport systems: implications for neurodegenerative diseases, *Cell. Mol. Neurobiol.* 21 (2001) 173–213.
- [12] B. Martin, R. Breneman, K.G. Becker, M. Gucek, R.N. Cole, S. Maudsley, iTRAQ analysis of complex proteome alterations in 3xTgAD Alzheimer's mice: understanding the interface between physiology and disease, *PLoS ONE* 23 (2008) e2750.
- [13] S. Joerchel, M. Raap, M. Bigl, K. Eschrich, R. Schliebs, Oligomeric beta-amyloid (1–42) induces the expression of Alzheimer disease-relevant proteins in cholinergic SN56.B5.G4 cells as revealed by proteomic analysis, *Int. J. Dev. Neurosci.* 26 (2008) 301–308.
- [14] Y. Hu, J.P. Malone, A.M. Fagan, R.R. Townsend, D.M. Holtzman, Comparative proteomic analysis of intra- and interindividual variation in human cerebrospinal fluid, *Mol. Cell. Proteomics* 4 (2005) 2000–2009.
- [15] Z. Xun, R.A. Sowell, T.C. Kaufman, D.E. Clemmer, Quantitative proteomics of a presymptomatic A53T alpha-synuclein *Drosophila* model of Parkinson disease, *Mol. Cell. Proteomics* 7 (2008) 1191–1203.
- [16] M.H. Chin, W.J. Qian, H. Wang, A. Petyuk, J.S. Bloom, D.M. Sforza, G. Lačan, D. Liu, A.H. Khan, R.M. Cantor, D.J. Bigelow, W.P. Melega, D.G. Camp 2nd, R.D. Smith, D.J. Smith, Mitochondrial dysfunction, oxidative stress, and apoptosis revealed by proteomic and transcriptomic analyses of the striata in two mouse models of Parkinson's disease, *J. Proteome Res.* 7 (2008) 666–677.
- [17] S. Campioni, M.F. Mossuto, S. Torrassa, G. Calloni, P.P. de Laureto, A. Relini, A. Fontana, F. Chiti, Conformational properties of the aggregation precursor state of HypF-N, *J. Mol. Biol.* 379 (2008) 554–567.
- [18] J.L. Mazzola, M.A. Sirover, Alteration of intracellular structure and function of glyceraldehyde-3-phosphate dehydrogenase: a common phenotype of neurodegenerative disorders? *Neurotoxicology* 23 (2002) 603–609.
- [19] S. Carujo, J.M. Estanyol, A. Ejarque, N. Agell, O. Bachs, M.J. Pujol, Glyceraldehyde 3-phosphate dehydrogenase is a SET-binding protein and regulates cyclin B-cdk1 activity, *Oncogene* 25 (2006) 4033–4042.
- [20] T. Müller, C.G. Concannon, M.W. Ward, C.M. Walsh, A.L. Tirniceriu, F. Tribi, D. Kögel, J.H. Prehn, R. Egersperger, Modulation of gene expression and cytoskeletal dynamics by the amyloid precursor protein intracellular domain (AICD), *Mol. Biol. Cell.* 18 (2007) 201–210.
- [21] D. Wessel, A. Flüggé, Method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, *Anal. Biochem.* 138 (1984) 141–143.
- [22] D.F. Hochstrasser, A. Patchornik, C.R. Merrill, Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining, *Anal. Biochem.* 173 (1998) 412–423.
- [23] V. Neuhoﬀ, N. Arold, D. Taube, W. Ehrhardt, Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250, *Electrophoresis* 9 (1988) 255–262.
- [24] H.R. Saibil, Chaperone machines in action, *Curr. Opin. Struct. Biol.* 18 (2008) 35–42.
- [25] R. Arya, M. Mallik, S.C. Lakhota, Heat shock genes – integrating cell survival and death, *J. Biosci.* 32 (2007) 595–610.
- [26] M.D. Schaller, C.A. Borgman, B.S. Cobb, R.R. Vines, A.B. Reynolds, J.T. Parsons, pp125FAK a structurally distinctive protein-tyrosine kinase associated with focal adhesions, *Proc. Natl. Acad. Sci. U. S. A.* 89 (1992) 5192–5196.
- [27] M.D. Schaller, Biochemical signals and biological responses elicited by the focal adhesion kinase, *Biochim. Biophys. Acta* 1540 (2001) 1–21.

- [28] I. Karakesisoglou, Y. Yang, E. Fuchs, An epidermal plakin that integrates actin and microtubule networks at cellular junctions, *J. Cell Biol.* 149 (2000) 195–208.
- [29] H.J. Chen, C.M. Lin, C.S. Lin, R. Perez-Olle, C.L. Leung, R.K. Liem, The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway, *Genes Dev.* 20 (2006) 1933–1945.
- [30] M.A. Lovell, S. Xiong, W.R. Markesbery, B.C. Lynn, Quantitative proteomic analysis of mitochondria from primary neuron cultures treated with amyloid beta peptide, *Neurochem. Res.* 30 (2005) 113–122.
- [31] S.J. Shin, S.E. Lee, J.H. Boo, M. Kim, Y.D. Yoon, S.I. Kim, I. Mook-Jung, Profiling proteins related to amyloid deposited brain of Tg2576 mice, *Proteomics* 4 (2004) 3359–3368.
- [32] D.C. David, L.M. Ittner, P. Gehrig, D. Nergenu, C. Shepherd, G. Halliday, J. Götz, Beta-amyloid treatment of two complementary P301L tau-expressing Alzheimer's disease models reveals similar deregulated cellular processes, *Proteomics* 6 (2006) 6566–6577.
- [33] J. Petrak, R. Ivanek, O. Toman, R. Cmejla, J. Cmejlova, D. Vyoral, J. Zivny, C.D. Vulpe, Déjà vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins, *Proteomics* 8 (2008) 1744–1749.
- [34] J.C. Ghosh, T. Dohi, B.H. Kang, D.C. Altieri, Hsp60 regulation of tumor cell apoptosis, *J. Biol. Chem.* 283 (2008) 5188–5194.
- [35] S.R. Kirchhoff, S. Gupta, A.A. Knowlton, Cytosolic heat shock protein 60, apoptosis, and myocardial injury, *Circulation* 105 (2002) 2899–2904.
- [36] S. Xanthoudakis, S. Roy, D. Rasper, T. Hennessey, Y. Aubin, R. Cassady, P. Tawa, R. Ruel, A. Rosen, D.W. Nicholson, Hsp60 accelerates the maturation of pro-caspase-3 by upstream activator proteases during apoptosis, *EMBO J.* 18 (1999) 2049–2056.
- [37] A. Samali, J. Cai, B. Zhivotovsky, D.P. Jones, S. Orrenius, Presence of a pre-apoptotic complex of pro-caspase-3, Hsp60 and Hsp10 in the mitochondrial fraction of Jurkat cells, *EMBO J.* 18 (1999) 2040–2048.
- [38] D. Chandra, G. Choy, D.G. Tang, Cytosolic accumulation of HSP60 during apoptosis with or without apparent mitochondrial release: evidence that its pro-apoptotic or pro-survival functions involve differential interactions with caspase-3, *J. Biol. Chem.* 282 (2007) 31289–31301.
- [39] V. Veereshwarayya, P. Kumar, K.M. Rosen, R. Mestrlil, H.W. Querfurth, Differential effects of mitochondrial heat shock protein 60 and related molecular chaperones to prevent intracellular beta-amyloid-induced inhibition of complex IV and limit apoptosis, *J. Biol. Chem.* 281 (2006) 29468–29478.
- [40] E.J. Tisdale, Glyceraldehyde-3-phosphate dehydrogenase is required for vesicular transport in the early secretory pathway, *J. Biol. Chem.* 275 (2001) 2480–2486.
- [41] P. Huitorel, D. Pantaloni, Bundling of microtubules by glyceraldehyde-3-phosphate dehydrogenase and its modulation by ATP, *Eur. J. Biochem.* 150 (1985) 265–269.
- [42] R. Singh, M.R. Green, Sequence-specific binding of transfer-RNA by glyceraldehyde-3-phosphate dehydrogenase, *Science* 259 (1993) 365–368.
- [43] G. Morgenegg, G.C. Winkler, U. Hübscher, C.W. Heizmann, J. Mous, C.C. Kuenzle, Glyceraldehyde-3-phosphate dehydrogenase is a nonhistone protein and a possible activator of transcription in neurons, *J. Neurochem.* 47 (1986) 54–62.
- [44] M.R. Hara, N. Agrawal, S.F. Kim, M.B. Cascio, M. Fujimuro, Y. Ozeki, M. Takahashi, J.H. Cheah, S.K. Tankou, L.D. Hester, C.D. Ferris, S.D. Hayward, S.H. Snyder, A. Sawa, S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding, *Nat. Cell Biol.* 7 (2005) 645–646.
- [45] D.M. Chuang, C. Hough, V.V. Senatorov, Glyceraldehyde-3-phosphate dehydrogenase, apoptosis, and neurodegenerative diseases, *Annu. Rev. Pharmacol. Toxicol.* 45 (2005) 269–290.
- [46] Z. Dastoor, J.L. Dreyer, Potential role of nuclear translocation of glyceraldehyde-3-phosphate dehydrogenase in apoptosis and oxidative stress, *J. Cell Sci.* 114 (2001) 1643–1653.
- [47] F. Magherini, C. Tani, T. Gamberi, A. Caselli, L. Bianchi, L. Bini, A. Modesti, Protein expression profiles in *Saccharomyces cerevisiae* during apoptosis induced by H₂O₂, *Proteomics* 7 (2007) 1434–1445.
- [48] M.D. Basson, An intracellular signal pathway that regulates cancer cell adhesion in response to extracellular forces, *Cancer Res.* 68 (2008) 2–4.
- [49] D.H. Crouch, V.J. Fincham, M.C. Frame, Targeted proteolysis of the focal adhesion kinase pp125 FAK during c-MYC-induced apoptosis is suppressed by integrin signalling, *Oncogene* 12 (1996) 2689–2696.
- [50] B. Levkau, B. Herren, H. Koyama, R. Ross, E.W. Raines, Caspase-mediated cleavage of focal adhesion kinase pp125FAK and disassembly of focal adhesions in human endothelial cell apoptosis, *J. Exp. Med.* 187 (1998) 579–586.
- [51] J. Grossmann, M. Artinger, A.W. Grasso, H.J. Kung, J. Scholmerich, C. Flocchi, A.D. Levine, Hierarchical cleavage of focal adhesion kinase by caspases alters signal transduction during apoptosis of intestinal epithelial cells, *Gastroenterology* 120 (2001) 79–88.
- [52] M.F. Mian, C. Kang, S. Lee, J.H. Choi, S.S. Bae, S.H. Kim, Y.H. Kim, S.H. Ryu, P.G. Suh, J.S. Kim, E. Kim, Cleavage of focal adhesion kinase is an early marker and modulator of oxidative stress-induced apoptosis, *Chem. Biol. Interact.* 171 (2008) 57–66.
- [53] L. Zhang, G.Q. Xing, J.L. Barker, Y. Chang, D. Maric, W. Ma, B.-S. Li, D.R. Rubinow, α -Lipoic acid protects rat cortical neurons against cell death induced by amyloid and hydrogen peroxide through the Akt signalling pathway, *Neurosci. Lett.* 312 (2001) 125–128.

Review

Technical advances in proteomics mass spectrometry: identification of post-translational modifications

Angela Amoresano, Andrea Carpentieri, Chiara Giangrande, Angelo Palmese, Giovanni Chiappetta, Gennaro Marino and Piero Pucci*

Department of Organic Chemistry and Biochemistry, University of Naples Federico II, Naples, Italy

Abstract

The importance of post-translational modifications (PTMs) of proteins has become evident in the proteomic era as it plays a critical role in modulating cellular function, and can vary in response to different stimuli thereby tuning cellular mechanisms. Assessment of PTMs on a proteomic scale is a challenging task since they are substoichiometric, transient and reversible. Moreover, the amount of post-translationally modified proteins is generally very small when compared to their unmodified counterparts. Existing methodologies for identification of PTMs essentially relies on enrichment procedure to selectively increase the amount of modified peptides. These procedures need to be integrated with sophisticated mass spectrometric methods to enable the identifications of PTMs. Although the strategies developed so far are not optimal, a number of examples will be given where the combination of innovative separation methods along with advanced mass spectrometric analyses provide positive results. These experiences are leading the way for the next generation of proteomic approaches for identification of a wide range of PTMs.

Clin Chem Lab Med 2009;47

Keywords: glycosylation; nitration; phosphorylation; post-translational modifications (PTMs); proteomics; tandem mass spectrometry (MS/MS).

Introduction

Over the past few years following the conclusion of the Human Genome Project, the challenge has again shifted from the gene to protein, giving rise to what has been termed the "proteomic era". The aim of this new era is to identify and characterize proteins located within a given organelle, cell or organism, as well as to unravel the *in vivo* protein pathways (1–3).

*Corresponding author: Prof. P. Pucci, Dipartimento di Chimica Organica e Biochimica, Università degli Studi Federico II, Complesso Universitario Monte S. Angelo, via Cynthia 4, 80126 Napoli, Italy
Phone: +39081674318, Fax: +39081674313,
E-mail: pucci@unina.it

Received October 11, 2008; accepted March 23, 2009

These new goals are not easily achieved because the challenge increases by several orders of magnitude when moving from genome to proteome research. The static nature of the genome cannot be compared to the dynamic properties of the proteome. Expression profiles of proteins change several times during the cell cycle and are significantly influenced by a number of intra- and extra-cellular stimuli including temperature, stress, apoptotic signals, and others (4). In addition, alternative splicing and post-translational modifications (PTMs) has resulted in re-thinking of the paradigm "one gene-one protein" that no longer reflects the true nature of the cellular proteome.

Modern proteomics research is focused on two major areas, expression proteomics and functional proteomics. Expression, or differential, proteomics aims to investigate changes that occur in the expression of protein profiles following specific stimuli and/or pathophysiological conditions. The goal here is to assess up regulation and down regulation of protein concentrations in abnormal cells compared with normal cells. Thus, expression proteomic studies are strictly concerned with transcriptomics experiments. However, mRNA concentrations do not necessarily correlate with the amount of corresponding protein due to translation regulation mechanisms. Thus, expression proteomics complements microarray investigations. This comparative approach is usually employed in the identification of proteins that are up regulated or down regulated in specific disease processes and therefore, may be used as diagnostic markers or therapeutic targets (5). It is not enough to map and identify all the proteins in cells; they must also be categorized with respect to their function and interactions with other proteins or genes. Therefore, the proteome must be studied *in vivo*, and the dynamic properties of the proteome used to obtain this information. This is the target of functional proteomics, an emerging area of proteomic study to investigate protein function *in vivo* through the isolation of functional complexes within the cell and the identification of their protein components. It is well understood that many cellular processes are governed by the rapid and transient association of protein complexes that fulfill specific biologic functions. The association of an unknown protein with others belonging to a specific complex would be strongly suggestive of its biological function.

In this scenario, the importance of PTMs is notably increased. Protein modification can affect biological function, thereby playing a critical role in cellular function. Also, they can vary in response to environ-

mental stimuli allowing cellular mechanisms to be finely tuned. Dereglulation might be involved in the development of disease.

Several mass spectrometry (MS)-based strategies have previously been developed to identify and characterize PTMs. With these techniques, a single, largely homogeneous protein was digested, with the resulting peptide mixture analyzed using MS procedures. The site and nature of the PTM could be easily inferred from the anomalous mass value detected for the modified peptide(s). The analytical problems faced today by proteomic laboratories are far more complicated. First, the sample very often consists of a complex mixture of proteins, some of which are modified. Second, the PTMs involved in the regulation of important cellular processes, such as phosphorylation, methylation, acetylation, nitration, etc., are substoichiometric, affect only a minor portion of the protein, are transient and are reversible. Therefore, the amount of modified peptides within a proteomic digest will generally be very small. Also, they will always occur in the presence of a much larger concentration of the corresponding unmodified fragments. Thus, an understanding of cellular mechanisms and processes requires a comprehensive assessment of PTMs and the changes that occur following signaling events (6). New strategies will have to be employed based on the integration of specific enrichment procedures with sophisticated mass spectrometric experiments.

This paper will review current existing methodologies to identify PTMs in proteomic experiments, with a focus on some specific examples.

Mass spectrometry

MS had been the method of choice to elucidate PTMs in protein chemistry. Today, MS plays a key role in proteomics studies since it is the only method with the sensitivity, accuracy, reproducibility and speed required for identifying unknown proteins and characterizing transient PTMs (7–9).

MS is based on the production of ions in a gaseous phase that allow for the determination of the molecular mass of analytes. A wide variety of mass spectrometers are now available; all sharing the ability to assign mass-to-charge values to ions, even though, the principles of operation and the types of experiments that can be done on these instruments differ greatly. The sample has to be introduced into the ion source of the instrument where molecules are ionized. The ions are extracted into the analyzer region where they are separated according to their mass-to-charge ratio (m/z) and then detected.

Mass spectrometers can be classified according to its ionization system; an essential component that defines the measurable mass range, sensitivity and resolution. Many ionization methods are available. However, the most widespread ion sources in biochemical analyses are electrospray ionization (ESI) and matrix assisted laser desorption ionization (MAL-

DI) (10, 11). Over the last few years, ESI (12, 13) has had a tremendous impact on the use of MS in biological research because it is well suited to the analysis of polar molecules. During standard ESI, the sample is dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary (12–14). High voltage, 3–5 kV, is applied to the tip of the capillary within the ion source. Sample emerging from the capillary tip is dispersed into a spray of highly charged droplets. This process is aided by a nebulizing gas flowing around the outside of the capillary. The charged droplets decrease in size by solvent evaporation which is assisted by a flow of warm nitrogen gas. This is also known as the drying gas and it passes across the front of the ionization source. Eventually, charged sample ions that are free of solvent are released from the droplets and pass into the analyzer of the mass spectrometer. One peculiar aspect of this technique is that it gives rise to multiple, charged molecular-related ions such as $(M+nH)^{n+}$ in positive ionization mode and $(M-nH)^{n-}$ in negative ionization mode.

MALDI-MS is based on bombardment of sample molecules with a laser to induce sample ionization. The sample is pre-mixed on a stainless steel plate with a highly absorbing matrix such as a weak aromatic acid. Once excited, the matrix is able to transfer energy and protons to the sample. After drying, the matrix molecules crystallize and solid crystals consisting of sample and matrix are formed. A laser, usually a pulsed nitrogen laser at 337 nm, is fired into the sample. The absorbance of the laser energy by the matrix molecules results in a desorption event. This energy is used to transfer a proton from the matrix to the analyte resulting in the formation of MH^+ ions.

Mass analyzers may incorporate quadrupoles (Q), ion traps (IT), time-of-flight (TOF), or some combination of these in “hybrid instruments” in order to obtain good resolution and sensitivity. The ESI source is usually employed with quadrupole and ion trap analyzers (15, 16). The advantages of the ion-trap mass spectrometer includes compact size, and the ability to trap and accumulate ions in order to increase the signal-to-noise ratio. It is also possible to use this type of mass spectrometer in fragmentation experiments (see below). Conventional ion trap mass spectrometers operate with a three-dimensional (3D) quadrupole field. This confers very high efficiency with respect to the time required to fill the ion trap and generate a complete mass spectrum. However, there are some problems with respect to trapping efficiency, primarily due to the small volume. These problems have been overcome by the introduction of linear ion traps (LIT) (17–19), characterized by a larger ion storage capacity and greater trapping efficiency (see below).

TOF analyzers are typically used in combination with MALDI sources. Following MALDI ionization, ions pass into a TOF tube for mass analysis. The TOF tube is under high vacuum (10^{-6} – 10^{-8} mbar) and consists of a field-free drift region. All ions entering the TOF tube have a fixed kinetic energy that is propor-

tional to the applied voltage and the charge. The higher the mass of the ion, the lower its velocity and the longer it takes before the ion reaches the detector. Ions of different mass can be separated based on their different velocities in the TOF-tube. Their m/z values are determined by measuring the time each ion takes to travel through the field free region.

MALDI and electrospray sources are known as soft ionization methods because the sample is ionized by the addition or removal of a proton, with very little energy remaining to cause fragmentation of the sample ions. Tandem mass spectrometry (MS/MS) must then be used to produce structural information about the compound by fragmenting specific sample ions inside the mass spectrometer, and then identifying the fragment ions. MS/MS uses two stages of mass analysis; the first to pre-select an ion and the second to analyze fragments. These are separated using a collision cell into which an inert gas, such as argon or xenon, is introduced to produce fragmentation. MS/MS also enables specific compounds to be detected in complex mixtures based on their characteristic fragmentation patterns.

Peptides fragment in a well-described manner (20). There are three different types of bonds that can fragment along the amino acid backbone: NH-CH, CH-CO, and CO-NH bonds. Each bond break gives rise to two species, one neutral and the other one charged, with only the charged species monitored by the mass spectrometer. Thus, there are six possible fragment ions for each amino acid residue as shown in Figure 1. The a, b, and c ions carry the charge on the N-terminal fragment, and the x, y and z ions retain the charge on the C-terminal fragment. The most common cleavage sites are the CO-NH bonds which give rise to the b and/or the y series of ions.

Protein identification by mass spectrometry

The key step in any proteomic study consists in the identification of proteins that have been separated either by gel electrophoresis and digested, or cleaved directly using enzymatic procedures in the so-called "bottom-up" approach (21). Protein identification from 2D gel electrophoresis is usually obtained by MALDI-MS analyses (22, 23). Following colloidal Coomassie staining, protein components are excised from the gel, reduced and alkylated with iodoacet-

amide to irreversibly block the cysteine residues and digested with trypsin. The resulting peptide mixture is extracted from the gel and analyzed directly using MALDI-MS (24, 25).

Identification of the different proteins is performed using the peptide mass fingerprinting procedure (26). This procedure uses the mass values, together with other information such as the protease used for the hydrolysis. This information is input into the Mascot mass search software or analogous programs available on the internet (27). The experimental mass values are compared to those obtained from a theoretical digestion of all proteins present in the database, allowing for identification of the protein(s).

Alternatively, if the fingerprinting procedure is unable to identify the proteins, combined liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods can be used. Peptide mixtures are fractionated by capillary HPLC. The fractions that elute from the column are applied directly into the electrospray MS and mass values determined. Peptide ions will be simultaneously isolated and fragmented within the mass spectrometer, producing daughter ion spectra from which information on sequence of individual peptides can be obtained. This information, together with the peptide mass values, is then used to search protein databases enabling identification of the protein components. It has been shown that sequence information from just two peptides is sufficient to identify a protein by searching protein and expressed sequence tag databases.

Recently, "top-down" proteomics has become popular in the assessment of PTMs. In this approach, the multicharged ions of intact proteins are used as precursor ions and fragmented to provide sequence information and protein identification. Technically demanding, sophisticated instrument setup constitutes the main challenge of top-down proteomics (28). However, recent developments in instrumentation have been introduced, such as specifically designed Fourier transform ion cyclotron resonance (FTICR) mass spectrometers, making the top-down approach suitable even for PTM studies (29).

Post-translational modifications (PTMs)

Protein PTMs are thought to be one of the major determinants for the complexity of life. These modifications increase the molecular heterogeneity of gene products, adding a further dimension to the proteome as compared to the genome. Greater than 5% of the genes in the human genome encode for enzymes that perform such modifications. These include hundreds of protein kinases and phosphatases, ubiquitinyl ligases, acetylases and deacetylases, methyl transferases and glycosyl transferases (30). The repertoire of PTMs is very complex considering that no < 15 out of the 20 common amino acids may be modified. These modifications range from addition of a single group to the linkage of a complex oligosaccharide moiety or glycosyl phosphatidyl

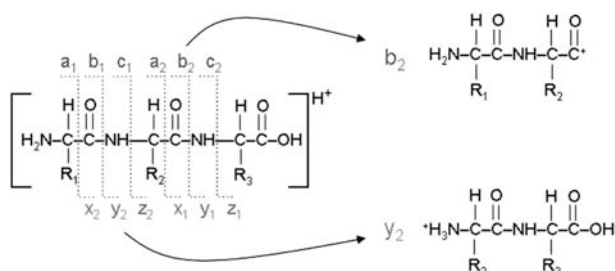


Figure 1 MS/MS peptide fragmentation pattern. Fragment ions are labeled with a, b, and c ions carrying the charge on the N-terminal fragment. The x, y and z ions having the charge retained on the C-terminal fragment.

inositol (GPI) anchor. More than 250 PTMs have been described to date (31). Genomic studies have not been able to elucidate complex mechanisms such as development of disease, aging, differentiation, development, and others that are finely regulated by dynamic changes in protein function mediated through PTMs.

The pattern of protein PTMs constitute a molecular code that dictates protein conformation and folding, cellular localization, cellular interactions and cell activities, tissue and environmental conditions. Understanding this code constitutes a major challenge of proteomics. It should be emphasized that detection of PTMs, while important, does not distinguish the number and type of PTMs that may be present on any one intact protein. This disconnection from the parent protein modifications may lead to the loss of information on the synergies between PTMs. New methods and strategies have been developed, and many more are needed to fully characterize PTMs in a specific proteome and to describe their changes.

Glycosylation

Glycosylation is a co- and post-translational modification where oligosaccharide chains are covalently linked to proteins. Glycoproteins are ubiquitous, being found in all living organisms from eubacteria to eukaryotes (32). They fulfill their biological roles as a complex mixture of glycosylated variants called glycoforms. This microheterogeneity is primarily due to the fact that oligosaccharide biosynthesis is not template driven. Assembly of the glycan moieties is governed by the relative amount of different glycosidic enzymes (33), the cell type, and physiological status. In addition, the reactions involved in glycosylation and deglycosylation are frequently incomplete, resulting in an increase in the heterogeneity of glycoproteins.

A new discipline, glycobiology, emerged in the 1980s, with a focus on the structural and functional characterization of glycoproteins and glycoconjugates (34). Recently, glycobiology has evolved into glycomics, the investigation of the complete set of glycoconjugates and carbohydrates that occur in an organism. Glycomics show a higher level of complexity than genomics, or even proteomics, primarily due to the microheterogeneity of glycoforms described above. Furthermore, glycans have a larger number of building blocks compared with nucleic acids and proteins. These can be assembled in complex branched structures in which monomeric units can adopt various conformations and binding positions.

Biosynthesis of glycopeptide linkages

The enzymes responsible for glycosylation are membrane-bound glycosyltransferases and glycosidases that are active in the endoplasmic reticulum (ER) and Golgi compartments. Their activity can gen-

erate two large classes of glycoforms; N-linked and O-linked glycans. O-linked oligosaccharides have the glycosidic component linked to the hydroxyl group of serine or threonine residues. N-glycosylation is characterized by the formation of an N-glycosidic linkage on the amide of asparagine residues. This glycosylation requires a strict consensus sequence *Asn-X-Ser/Thr* (35). Although this sequence frequently occurs, it is not always a site that is glycosylated. This is most probably due to conformational factors. N-glycosylation begins in the ER (36) where a biosynthetic precursor, a dolichol-P-linked oligosaccharide ($\text{Glu}_3\text{Man}_9\text{GlcNAc}_2$), is transferred to Asn residues in nascent polypeptide chains. This glycan is a well-conserved structure in evolution; the same compound is seen in plant, fungal and mammalian cells.

Further processing of the oligosaccharide may occur in the Golgi complex. The structural diversity found in Asn-linked oligosaccharides from different organisms, cell types or tissues, originates from processing reactions occurring in the Golgi apparatus (37, 38).

O-glycosylation also occurs in the Golgi and consists of the step-wise addition of sugar residues directly onto the polypeptide chain. This process starts from an N-acetylgalactosamine residue linked to the hydroxyl group of Ser or Thr residues (39). A specific glycosyltransferase then adds the galactose residue to the GalNAc. The last steps in biosynthesis of typical O-glycans are the additions of two N-acetylneuramic acid (sialic acid) residues in the trans-Golgi reticulum. O-oligosaccharides can vary in size from a single GalNAc residue, referred to as the Tn antigen, to oligosaccharides comparable in size to complex N-oligosaccharides.

The role of glycosylation

The oligosaccharide moiety confers specific physicochemical properties to glycoproteins, and fulfills several biological functions. Several studies dealing with the importance of N-glycosylation in the folding process have been reported. Glycans constitute bulky hydrophilic moieties that reduce aggregation and promote solubility. This feature helps the protein follow a specific folding pathway. However, the principal role in the folding process is played by deglycosylation and reglycosylation in the ER, which occurs repeatedly during the biosynthesis of glycoproteins. This system is recognized as a central component of a quality control and chaperone pathway in the ER lumen (37, 40–42).

Some oligosaccharides play an important role in the coating of many glycoproteins, protecting them from proteases and antibodies and affecting their turnover and half-life (43, 44). Oligosaccharides can act as specific receptors for pathogens, being responsible for interactions with vertebrate cells through lectins. Lectins are specialized glycoproteins that can recognize vertebrate glycan structures with high specificity (45). The variable nature of surface glycans may be a way to escape pathogens' recognition systems.

Recognition by ligands or receptors is a constant characteristic of glycoproteins. Most are cell-surface receptors, whose glycosylation affects binding and activity. Cells are covered with a high-density coating of oligosaccharides, indicating that oligosaccharides are responsible for cell-cell and cell-matrix interactions and adhesion (43).

Glycoproteome characterization

Understanding the structure, enzymology and cellular biology of glycosylation is the main goal of glycomics with the ultimate goal being the comparison of organisms in normal and pathological conditions (44). However, difficulties exist in the analysis of complex mixture of glycoconjugates, primarily due to their microheterogeneity as stated above. These difficulties largely prevent the simple and direct approach of analyzing the entire set of glycans following their release from glycoproteins or glycoconjugates. Several protocols (46–49) have been developed to release the oligosaccharides from both O-linked and N-linked glycoproteins. However, the ensemble of all these glycans constitute a far too complex mixture to be analyzed directly by any mass spectrometric method. These analyses can provide information only on the most abundant glycoforms, leaving the vast majority of minor glycans undetected.

Various methods have been developed to enrich glycoproteins in a complex mixture. One approach relies on capturing N-linked glycoproteins using hydrazine chemistry. The oligosaccharides of the glycoproteins are oxidized to form the corresponding di-aldehydes that can be covalently linked to hydrazide beads. The captured glycoproteins are then digested and the unmodified peptides washed out. Captured glycopeptides can be deglycosylated with PNGase F and quantified by isotopic labeling (50).

Chemical derivatization presents several side reactions. Thus, lectin affinity chromatography has recently been developed as an analytical tool in glycoproteomics. Several lectin-derivatized media are commercially available for enrichment of glycoproteins and glycopeptides. A single lectin recognizes specific glycoforms, whereas the use of a set of lectins allows the complete set of glycoproteins to be captured in a single step. Multilectin affinity columns have been developed combining different lectins: Hancock and co-workers combined concanavalin A (ConA), wheat germ agglutinin (WGA) and Jacalin for the analysis of the serum glycoproteins (51). Comparative serum glycoproteomics has been applied to the study of pathological samples using a multilectin affinity approach to identify oligosaccharide markers in such diseases (33, 52–54).

Alternative methods rely on the chemical and/or enzymatic hydrolysis of glycoproteins, followed by affinity capture of the entire ensemble of glycopeptides with multiple lectins. Selective isolation of glycopeptides allows larger recovery of glycoforms since the steric hindrance of glycoproteins severely affect the interaction with lectins.

The structural characterization of glycans is still a difficult task due to the presence, in most cases, of several different structures at the same glycosylation site (46). Unfortunately, although MS cannot differentiate between the different hexoses and hexose-amines, it has become a valuable tool in their characterization (55).

The determination of the entire structure of oligosaccharide chains requires that the intact sugar moieties be released from the glycoproteins/glycopeptides pool using either PNGase F or chemical methods (56–61). Glycan profiling is then achieved by MALDI-MS analysis of the intact glycan mixture. The MALDI-MS spectra show the presence of intense quasi-molecular ions originating from the whole oligosaccharide moieties; assignment of the different glycoforms is then performed on the basis of the molecular weight of the oligosaccharides and their known biosynthetic pattern (46).

This analysis is not sufficient to provide structural information about the sequence and linkages of glycans. Combination of MALDI-profiling with post-source decay (PSD) or collision-induced dissociation (CID) analyses is generally required (62). Glycan fragmentation spectra present some main cleavage sites. A systematic nomenclature for oligosaccharide fragment ions was established by Domon and Costello (63), following that already in use for peptides. However, the structure of carbohydrates necessitates the use of a more complex annotation due to possible breakages occurring within the sugar ring. The primary cleavage sites, termed glycosidic cleavages, involves the breaking of a single bond between adjacent sugars (Figure 2). They provide information on sequence and branching, but little information on linkage. Other cleavage sites, termed cross-ring cleavages involving cleavage of two bonds are useful in determination of the linkage. It is worth noting that ions retaining the charge on the reducing terminus are labeled X (cross-ring), Y ($C_1 \rightarrow O$ -glycosidic) and Z ($O \rightarrow C_x$ glycosidic), while if the charge is retained at the non-reducing end, the produced ions are called A, B and C. The B and Y ions are usually the most abundant species.

Cross-ring cleavage necessitates additional superscript prefixes, denoting the atoms of the two cleaved

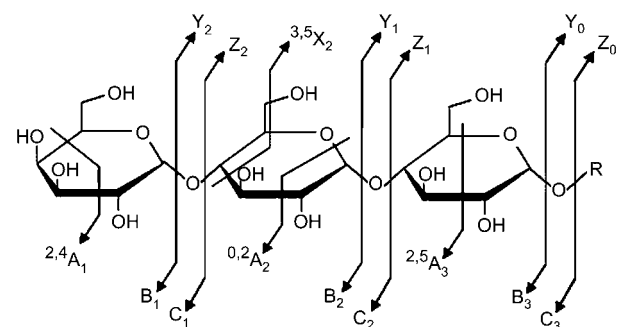


Figure 2 Carbohydrate fragmentation pattern. Ions retaining the charge at the reducing terminus are named X, Y and Z. Complementary ions are labeled A, B and C.

bonds. The situation becomes more complicated in cases involving branched oligosaccharides. The nomenclature is based on the carbohydrate component of glycopeptides and glycoproteins, which consist of a "core" unit and the branches, usually called "antenna". Each antenna is identified by a Greek letter with the largest being given the symbol " α " and the other symbols assigned based on decreasing molecular weight.

CID provides focused ions and nanospray technology provides good sensitivity. It is possible to perform MS/MS or MS³ experiments in order to obtain extensive fragmentation of oligosaccharides or glycopeptides. This analysis helps elucidate the different glycosidic structures present in glycoproteins including the branching pattern, the sequence of antenna and the possible occurrence of modifying groups, such as sulfate, phosphate, acetyl groups, and others (64).

Glycoproteome in tumor and inflammation: a case study

Recently, our group introduced a simple strategy that provides preliminary insight on the comparison of glycoproteomes in sera from healthy and diseased individuals. This technique uses a single affinity chromatography step coupled with MS (Figure 3). The N-linked glycopeptides resulting from digestion of sera samples with trypsin were selectively enriched in order to enhance identification of N-glycosylation sites using LC-MS/MS, and to obtain glycoform profiles using MALDI-MS analysis. Human sera from patients with either myocardial damage or hepatic carcinoma were evaluated since they represent examples of inflammatory and malignant disease.

Myocarditis is a group of diseases caused by infection, toxins, or autoimmune mechanisms characterized by inflammation of the heart, leading to necrosis of myocytes (65, 66). Glycan moieties might be responsible for the increase in leukocytes in damaged tissue. Hepatic carcinoma is a primary malignancy of the liver. There are several hypotheses about the role that glycans play in the development of cancer. C glycan structures are known to contribute to oncogenic signaling, transformation and metastasis (67).

Aliquots of serum from the two different patient groups, and healthy donors, were pooled and digested with trypsin. The three peptide mixtures were subjected to ConA affinity purification. The glycopeptides were then deglycosylated using PNGase F treatment and analyzed by MS. Similar analysis was performed on the unbound ConA fractions which contained primarily non-glycosylated peptides and highly branched glycopeptides. The ConA affinity step specifically retained the high mannose, hybrid and biantenna structures, separating them from the rest of the glycans.

The oligosaccharides retained by ConA were eluted and the glycans analyzed with MALDI-MS. Each signal in the MALDI spectra was tentatively attributed to a glycan arrangement on the basis of the mass value and the known N-glycan biosynthetic pathway. Since desialylation might have occurred in both positive and negative ion mode, MALDI analyses were also performed using 2,4,6 trihydroxyacetophenone as matrix. This has been reported to minimize loss of sialic acids (68).

More than 50 arrangements were detected, belonging primarily to high mannose and biantenna classes, indicating a high degree of glycan heterogeneity. While some qualitative and quantitative differences

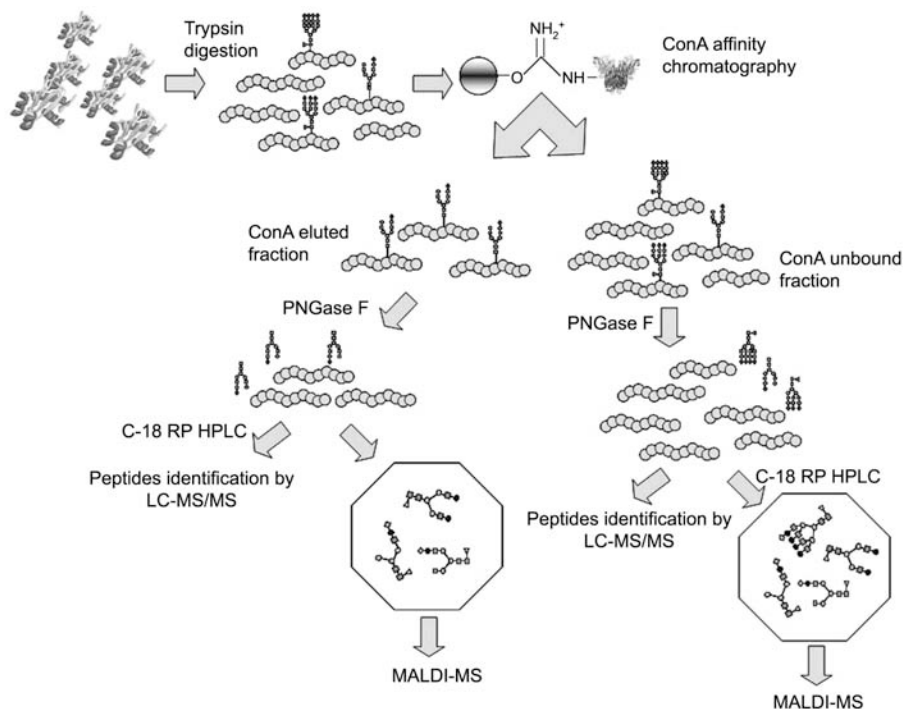


Figure 3 Flow chart showing the strategy to characterize N-linked glycoproteomes in sera from healthy and diseased individuals.

could be recognized in the three glycan profiles, the amounts did not significantly. However, differences in the relative intensities of the most abundant peaks were detected in the samples from patients with disease compared with those from healthy individuals.

Non-glycosylated peptides and peptides containing O-glycans and large complex glycans were not retained on the lectin. These constituted the unbound fraction from the affinity purification. PNGase F was added to this fraction, glycans were separated from peptides by HPLC and analyzed by MALDI-MS. Figure 4A–C shows the negative MALDI spectra of the oligosaccharide unbound fraction of the samples from the diseased and healthy individuals.

The structure of the various glycoforms was inferred on the basis of their mass value and the known biosynthetic pathway. The most common glycans were also selected for PSD fragmentation analysis to confirm oligosaccharide assignments. As an example, Figure 5 shows the PSD spectrum of the glycoform at m/z 2391.10. This glycoform displays the classic fragmentation pattern leading to the reconstruction of the sialylated fucosylated biantenna glycan.

In contrast to the fraction retained with ConA, MALDI-MS analysis of the unbound glycan fraction led to the detection of a plethora of glycoforms, and highlighted differences in the profiling of the three sample types. Glycans from patients with disease were characterized by abnormal glycosylation patterns, with differences in branching, sialylation and fucosylation. These patients showed an excessive expression of peculiar structures, the persistence of incomplete and truncated structures, the accumulation of precursors and the appearance of novel structures. This resulted in a higher degree of heterogeneity compared with healthy patients. The presence of more than 60 different glycoforms was observed in patients with hepatic carcinoma (K-hepatic). The increased size and branching of N-linked glycans created additional sites for terminal sialic acid residues; whose role in promoting metastasis is generally recognized. Myocardial lesions revealed the presence of hyper-fucosylated signals, whereas K-hepatic sera were characterized by hyper-fucosylation and hyper-sialylation.

This study provided an overview of the abnormal glycosylation that occurs in disease. The ConA affinity step was instrumental in depleting commonly occurring glycoforms. Without this step, highly branched, less common glycoforms might have escaped mass spectral analyses. Our data for K-hepatic serum were in perfect agreement with results obtained by Zhao et al. (52) who conducted a similar study on patients with pancreatic cancer.

Selective labeling procedures in PTMs

Identification of post-translationally modified proteins requires direct observation of both the modified peptides and the specific residues involved. Due to the complexity of proteomic samples, identification of

modified proteins often relies on the use of “specific” antibodies able to recognize individual PTMs (70–75). Once the modified proteins have been labeled with the specific antibody, they can be identified by MS procedures such as peptide mass fingerprinting and/or LC-MS/MS. Frequently, however, the actual modified peptides cannot be identified. Moreover, the occurrence of more than one protein in a specific location of the gel makes identification of PTMs problematic.

Several new methods based on selective labeling of modified peptides have recently been proposed (76–79). The strategy behind these methods consist of the direct selective analysis of labeled peptides with the mass spectrometer without the use of 2D gels and enrichment steps needed to reduce the complexity of the peptide mixture prior to MS analysis. Methods specifically tailored for the qualitative and quantitative determination of PTMs have recently emerged (80–83). The use of stable isotope labeled tags (light and heavy tags) (78, 79) can also address the problem of quantitative determination of PTMs. While the tagging method currently described can target some PTMs more specifically than others, the development of suitable selective chemical labeling strategies that can identify a larger number of PTMs should provide increased confidence in the identification of PTMs.

Reporter ion generating tag strategy (RIGhT)

Several groups have reported on the potential of chemical derivatization in conjunction with ESI-MS for the analysis of synthetic and naturally occurring compounds. Treatment with an appropriate derivatizing agent can improve the ionization of a compound and thus the detection of the analyte by ESI-MS.

Recently, our group proposed a novel strategy for identification of PTMs (79, 84). Our strategy has general applicability and is based on selective labeling of phospho-Ser/Thr residues. We proposed that the derivatizing agent may include a chemical moiety capable of fragmenting in a particular, thus allowing for selective mass spectral analysis of modified peptides through use of MS^n techniques.

This strategy was given the acronym reporter ion generating tag (RIGhT) to indicate a common derivatization method based on the labeling of target residues with reagents capable of generating reporter ions in MS^2/MS^3 experiments. Dansyl chloride (DNS-Cl), a very well-known fluorescent derivatizing reagent commonly used in protein chemistry, was chosen as the reporter tag. Approximately 40 years ago we discovered that fragmentation of dansyl derivatives gave rise to the typical m/z 170 and m/z 234 daughter ions fragments in electron impact experiments (85).

The RIGhT method is based on: 1) selective modification of target residues with DNS-Cl or other available dansyl reagents and, 2) selective detection and identification of labeled peptides by their characteristic fragmentation pathway in MS/MS experiments and the diagnostic m/z 234–170 transition in MS^3 mode. Precursor ions scans (PIS) of the m/z 170 frag-

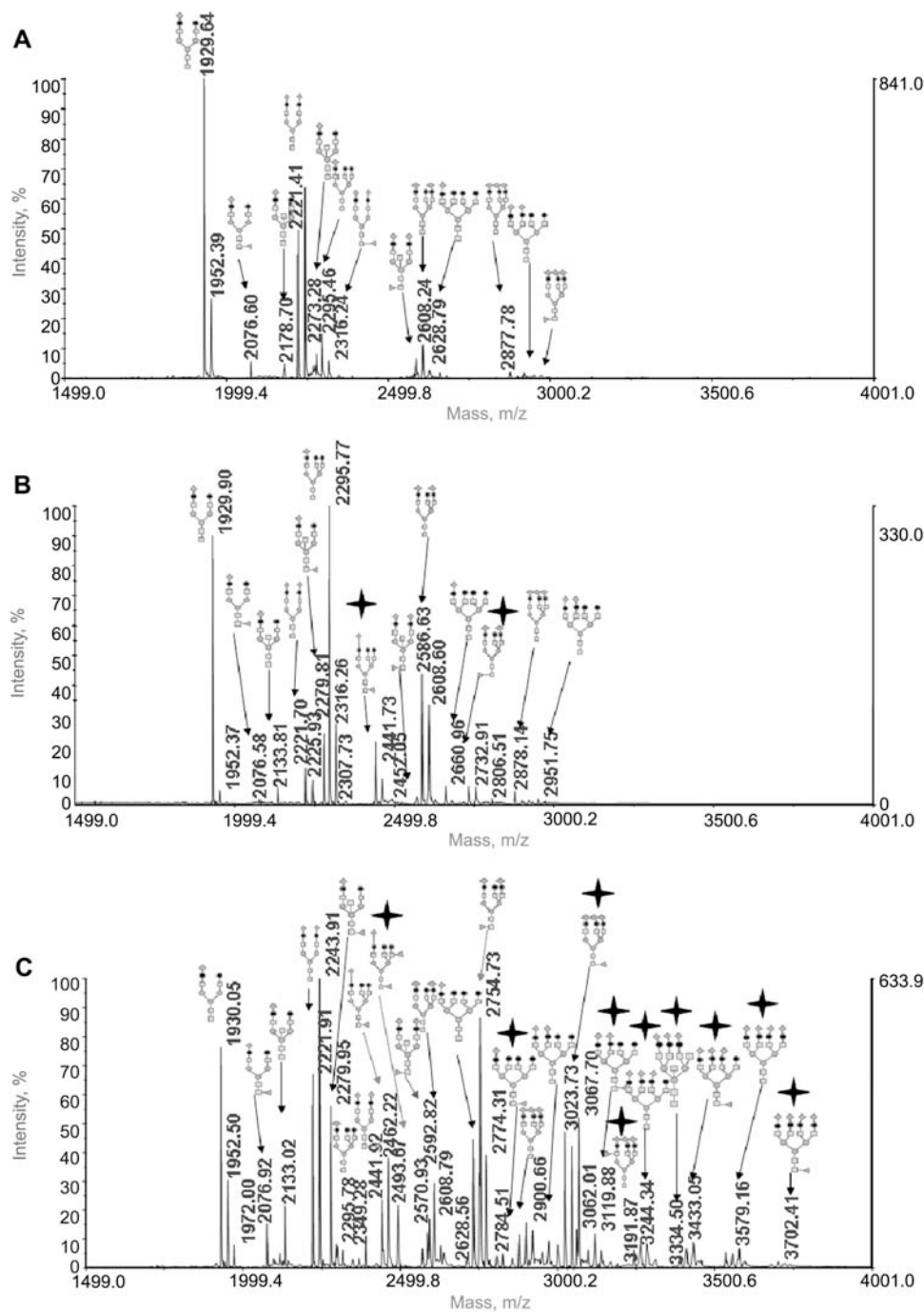


Figure 4 MALDI-MS profiling in negative ion mode of the oligosaccharide extracted from the unbound fraction of ConA affinity chromatography.

(A) Healthy serum, (B) myocardial lesions serum, (C) K-hepatic serum. The most important oligosaccharides are depicted; structures present only in sera from individuals with disease are indicated by a star.

ment can then be combined with MS² and MS³ analysis to specifically detect only the parent ions producing the m/z 234–170 fragmentation (Figure 6).

Application of the RIGhT approach is dependent on the availability of high-performance MS instruments able to perform analysis with high-sensitivity and accuracy. LIT mass spectrometers developed by the combination of triple quadrupole MS with LIT technology is one such device capable of this type of analysis (86). LIT mass spectrometers combine the advantages of very selective scanning of precursor ion with a triple quadrupole along with sensitive MS/

MS capabilities of the ion trap (87). This hybrid instrument is suitable for highly sophisticated MSⁿ analysis.

Phosphorylation

Phosphorylation is a reversible modification that acts as a switch to turn “on” or turn “off” protein activities or cellular pathways. Phosphorylation acts in a reversible manner by promoting the folding and function of proteins, enzyme activities or substrate specificities,

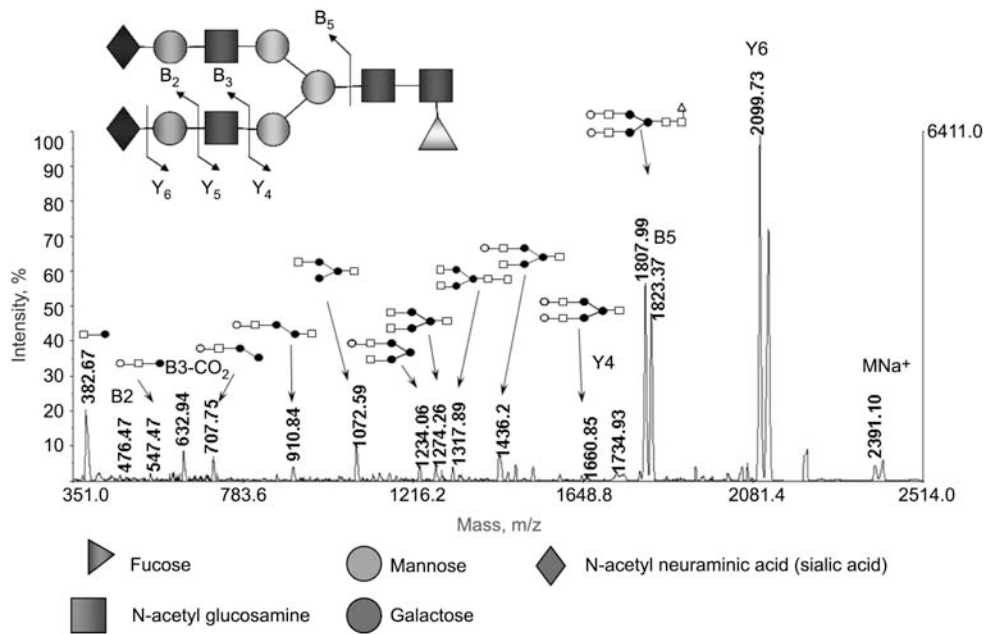


Figure 5 The MNa^+ ion at m/z 2391.10 was selected and submitted to PSD fragmentation. Y and B ions as well as fragments originated by double cleavages are indicated.

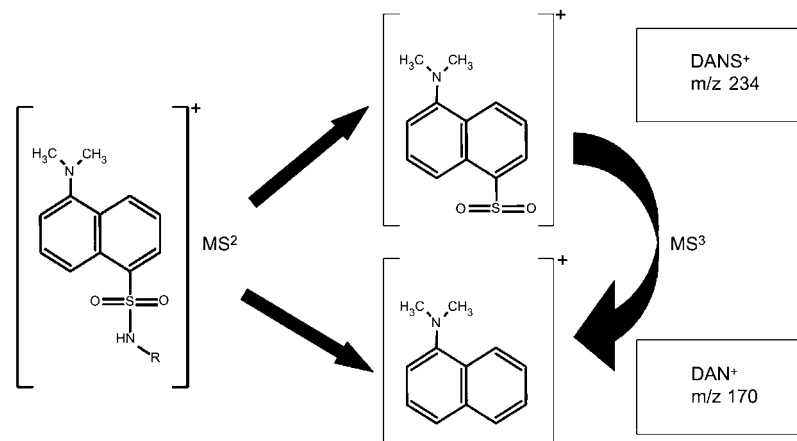


Figure 6 Fragmentation of dansyl derivatives. The typical m/z 170 and m/z 234 daughter ions fragments in MS/MS experiments as well as the diagnostic m/z 234-170 fragmentation in MS^3 mode are indicated.

regulating protein localization, and complex formation and degradation. A wide range of cellular processes including cell cycle progression, differentiation, development, peptide hormone response, signal transduction, metabolic maintenance and adaptation are all regulated by protein phosphorylation (88). The variety of functions in which phosphoproteins are involved necessitates a wide diversity of phosphorylation. Several amino acids can be phosphorylated by four types of phosphorylation. O-phosphates are the most common class and attach most often to serine-, threonine- and tyrosine-residues. However, O-phosphates can also attach to unusual amino acids such as hydroxyl-proline. N-, S- and acyl-phosphorylation are far less common and occur on histidine and lysine (N-phosphates), cysteine (S-phosphates) and aspartic and glutamic acid residues (acyl-phosphates).

Another characteristic of phosphorylation is the ratio of phosphorylated and non-phosphorylated proteins present in the cell. Protein phosphorylation usually occurs at substoichiometric level in the cell; some residues are always phosphorylated, while others are modified only transiently. Phosphorylation rates and phosphoproteins abundance are very low in signaling pathways, accounting for only 1%–2% of the entire protein content present in the phosphorylated form.

Protein phosphorylation is tightly regulated by a complex set of kinases and phosphatases which are responsible for protein phosphorylation and dephosphorylation, respectively. Kinases and phosphatases are highly regulated, and phosphorylation cycles may take place on a very short timescale. In addition to problems concerning regulation, analysis is compli-

cated by the complexity of phosphorylation patterns and the sheer number of phosphorylation sites. In many cases, phosphorylation effects are combinatorial and multiple sites are phosphorylated (89).

It has been estimated that as many as one-third of the proteins within a given eukaryotic proteome undergoes reversible phosphorylation, and there are more than 100,000 estimated phosphorylation sites in the human proteome. Deciphering the phosphoproteome will be a huge and challenging task due to the different types of phosphorylation generating a variety of phosphoproteins that are not easily analyzed.

Phosphoproteome analysis

An important part of phosphorylation analysis is sample preparation that preserves protein phosphorylation. Detection of phosphorylated proteins within complex mixtures is usually not possible without prior separation of the proteins. Although 2D-gel electrophoresis has been used primarily for separation and quantification of proteins from different systems, methods are also being developed to characterize the phosphoproteome. Phosphoproteins can be visualized directly in 2D gels using phospho-specific stains or Western blotting.

Commercially available phospho-specific stains are less sensitive than methods that use radioisotopes, but the handling of these non-isotopic reagents is more convenient (90). As an example, Pro-Q Diamond, (Molecular Probes) a fluorescent dye, allows detection of 1–20 ng of phosphoproteins in polyacrylamide gel, depending on the phosphorylation state of the protein. For individual phosphoproteins, the strength of the signal correlates with the number of phosphate groups. Depending on the principle used for staining, this method is more specific for Ser/Thr phosphorylations.

Antibodies can be used to discriminate between phosphorylation of serine, threonine and tyrosine (91). The specificity and sensitivity of immunostaining is highly dependent on the antibody used. Various antibodies of good specificity toward phospho-tyrosine are available with very little cross-reactivity.

Once phosphorylated, protein can be detected on a gel using either specific staining or immunotechniques. A 2D gel can be prepared and developed using colloidal Coomassie Blue or Silver staining. Following image analysis, phosphorylated protein spots are matched to protein spots on the preparative gel. Proteins of interest can then be excised from the gel, digested and identified by MS. Major advantages of this approach are in the detection of N-, S- and O-phosphorylation and in the possibility of quantitating the modified proteins (92, 93). To measure changes in the state of protein phosphorylation, two distinct proteome samples can be analyzed and immunoblotted, and the intensity of corresponding spots compared.

The main drawback of the method is that indirect identification of phosphorylated proteins by staining methods for direct detection of modified peptides is very unlikely using MS (94, 95). Moreover, when

many proteins are present in the same location on the gel, identification of the actual modified protein is not trivial.

Over the years, several strategies have been developed to increase the sensitivity of phosphoproteome analysis that eliminate the need for specific staining and antibody labeling (96, 97).

Recently, MS based methods have emerged as powerful and preferred tools for the analysis of PTMs, including phosphorylation, due to higher sensitivity, selectively, and speed compared with other biochemical techniques. However, despite advances in MS for analysis of phosphorylation, many difficulties remain. First, the generally low phosphorylation stoichiometry of proteins, including phosphopeptides, results in low amounts of protein in the peptide mixtures. Second, the increased hydrophilicity results in reduced retention of phosphopeptides on reversed-phase columns. Finally, there is selective suppression of the ionization/detection efficiency in the presence of large amounts of unphosphorylated peptides when using MS analysis in positive mode. For these reasons, several technologies such as affinity, liquid reverse phase, ion exchange chromatography and capillary electrophoresis, or some combination of these prior to MS analysis, have been used in proteomics to enrich phosphoproteins that might otherwise not be detected. The enrichment step combined with the high-sensitivity of MS technologies provides great potential for characterization of the phosphoproteome (98).

The use of miniaturized immobilized metal affinity chromatography (IMAC), in which phosphopeptides are non-covalently bound to resins that chelate Fe(III) or other trivalent metals, followed by base elution, has proved to be a potentially valuable method for enrichment of phosphopeptides (99–102). With further refinements, this technique may offer the best potential for large-scale phosphorylation analysis.

Another approach to enrich phosphoproteins is based on the use of specific antibodies to immunoprecipitate proteins from complex lysates. With this approach, the extraction of phosphoproteins leads to substantial simplification of the protein pattern, and enrichment phosphoproteins present in low concentrations (103).

Several methods for selective enrichment of phosphoproteins and phosphopeptides utilize chemical modification of the phosphate group. This approach does not distinguish between O-glycosylated and phosphorylated analytes, thus requiring additional analysis to confirm phosphorylation. Zhou et al. established a multi-step derivatization method capable of enriching not only Ser/Thr-phosphorylated peptides, but all types of phosphorylated peptides as well (6, 104). Phosphopeptides are bound to sulfidryl-containing compounds via phosphoamidate-bonds and can, therefore, be linked covalently to a solid support with immobilized iodoacetyl-groups. The phosphate-groups are not cleaved off the respective residues, so native phosphopeptides are obtained after elution with trifluoroacetic acid (TFA). A drawback of this

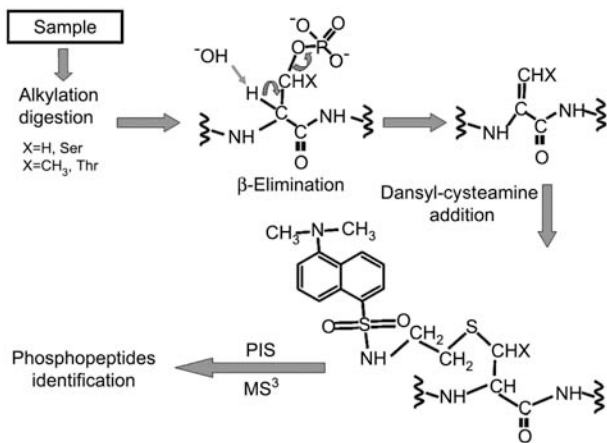


Figure 7 Identification of phosphoSer/Thr by β -elimination of the phosphate group from Ser/Thr residues, selective addition of a dansyl derivative on the newly generated double bond, and detection by precursor ion/ MS^2 / MS^3 mass spectral analysis.

method is the extensive work is required to quantitatively block all amino- and carboxy-groups within the peptides to prevent intramolecular and intermolecular condensation.

Another chemical tagging reaction is based on the β -elimination of phosphate groups in phospho-Serine/phospho-Threonine (pSer/pThr) residues under alkaline conditions, followed by a Michael-type addition of a thiol compound on the resultant double bond (105).

Dansyl labeling of pSer/pThr

In an attempt to improve and simplify the plethora of methods described, we employed the RIGHt strategy (84) to the selective identification of phosphoSer/Thr peptides. The method consists of (a) β -elimination of the phosphate group from Ser/Thr residues; (b) selective addition of a dansyl derivative on the newly formed double bond, and (c) the selective detection

and identification of labeled peptides by precursor ion/ MS^2 / MS^3 mass spectral analyses (Figure 7).

This approach was first developed using bovine α -casein as a model phosphoprotein. The phosphate moieties from pSer and pThr using barium hydroxide ion-mediated β -elimination (79). The peptide mixture was then analyzed using MALDI-MS which showed ions corresponding to the β -eliminated peptides that were 98 Da lower than expected. However, although DNS-Cl is known to react with primary NH_2 groups, it is unable to give a Michael addition to double bonds. A novel dansyl derivative of cysteamine was, therefore, synthesized following the synthetic route depicted in Figure 8. The free SH group of dansyl-cysteamine was then reacted with the β -eliminated peptide mixture via Michael-type addition. The yield of this addition reaction was monitored with MALDI-MS, showing that the extent of the reaction was not quantitative as expected for a typical Michael reaction addition (84).

In a proof-of-principle experiment, a tryptic digest from ten standard proteins was spiked with a minimum amount of an α -casein tryptic mixture modified with dansyl-cysteamine. An aliquot of 1 pmol of the peptide mixture was analyzed by LC-MS/MS with a LIT instrument, combining a PIS (m/z 170) with MS^3 analysis (see above) (Figure 9).

Figure 10 shows a typical reconstructed ion chromatogram for the transition m/z 234/170 in MS^3 mode that is dominated essentially by two signals. The corresponding MS^2 spectra led to reconstruction of the entire sequence of the β -eliminated phosphopeptides 104–119 and 106–119 that carried a dansyl-cysteamine moiety. The modified ion is stable during CID, providing easily interpretable product ion spectra. In fact, the y and b product ions still retain the modifying group linked to the β -eliminated Ser residue, allowing exact localization of the site of phosphorylation.

However, this approach seems amenable only for monophosphorylated peptides since the diphosphorylated and pentaphosphorylated α -casein tryptic

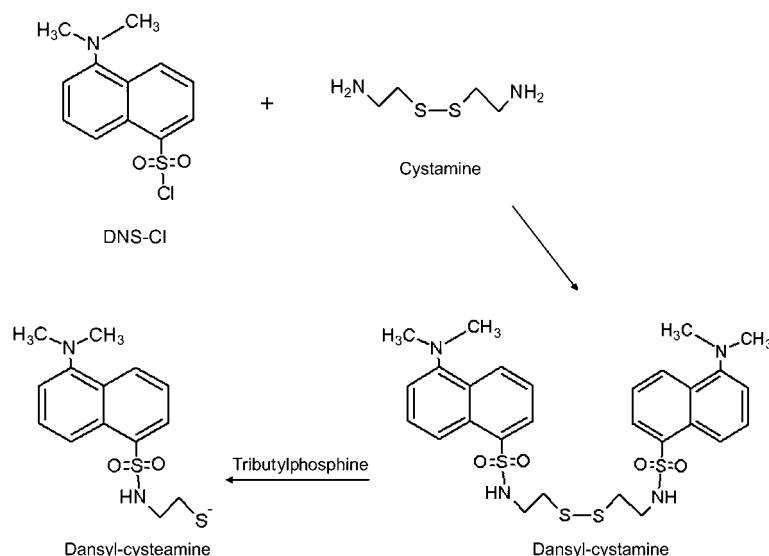


Figure 8 Synthesis of dansyl-cysteamine.

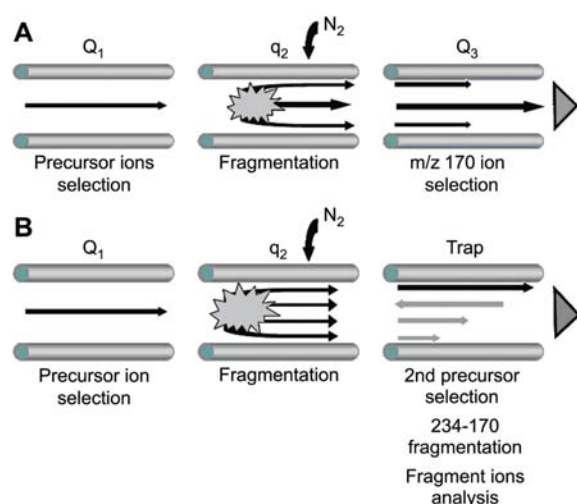


Figure 9 Flow chart showing bidimensional mass spectral analysis.

In PIS mode the Q1 was scanned across the full mass range, and ions are fragmented in the collision cell. The Q3 was set to transmit only the mass of the diagnostic product ion (m/z 170). Therefore, only those precursor ions that originated the ion at m/z 170 by fragmentation were detected (A); these selected precursor ions were then submitted to a combined MS²/MS³ experiment to specifically detect only those ions that produce the transition m/z 234/170 in the MS³ scan mode leading to the detection of the dansyl peptides (B).

peptides could not be detected. This is probably due to steric hindrance and high hydrophobicity of highly modified dansyl peptides. The detection of multiphosphorylated peptides is a difficult challenge. Among the proposed derivatization procedures, only the modification strategy that we have proposed based on the use of dithiothreitol (79) has resulted in mass spectrometric analysis of multiphosphorylated peptides that we have confidence in.

Nitration

Nitration can occur in cells during oxidative stress and inflammation, likely through generation of peroxynitrite (ONOO⁻) which is a potent oxidizing and nitrating agent. Peroxynitrite is produced by the reaction of superoxide anion in the presence of excess nitric oxide (106, 107). Targets of nitration include nucleic acids, sugars, lipids, alcohols, small organic molecules and aromatic amino acids (108). Sites of nitration in proteins are phenylalanine, tryptophan and tyrosine residues, with the phenolic ring being particularly active toward this modification.

Tyrosine nitration is increasingly recognized as the prevalent, functionally significant post-translational protein modification that serves as an indicator of nitric oxide (NO⁻) mediated oxidative inflammatory reactions (109). Tyrosine nitration, in fact, reflects the extent of oxidant production that occurs in both physiological and pathological conditions. Two catalytic mechanisms have been proposed for *in vivo* nitration. Both mechanisms involve the formation of tyrosyl radicals that can react with nitric dioxide to generate NO₂Tyr. In addition to this principal pathway, tyrosyl radicals can also react with another tyrosyl radical to produce the 3-3'-dityrosine specie which can generate intramolecular cross-links or protein multimerization.

Another possible mechanism of protein nitration *in vivo* is the reaction of tyrosyl radicals with nitric oxide (NO), resulting in the formation of 3-nitrosotyrosine (108). Apparently, nitrotyrosine is not stable and is oxidized by two electrons to form 3-nitrotyrosine.

Addition of a NO₂ group to the *ortho* position of tyrosine decreases the pK_a of the phenolic group approximately three pH units. This gives a net negative charge that is approximately half of that of nitrated tyrosine residues at physiological pH.

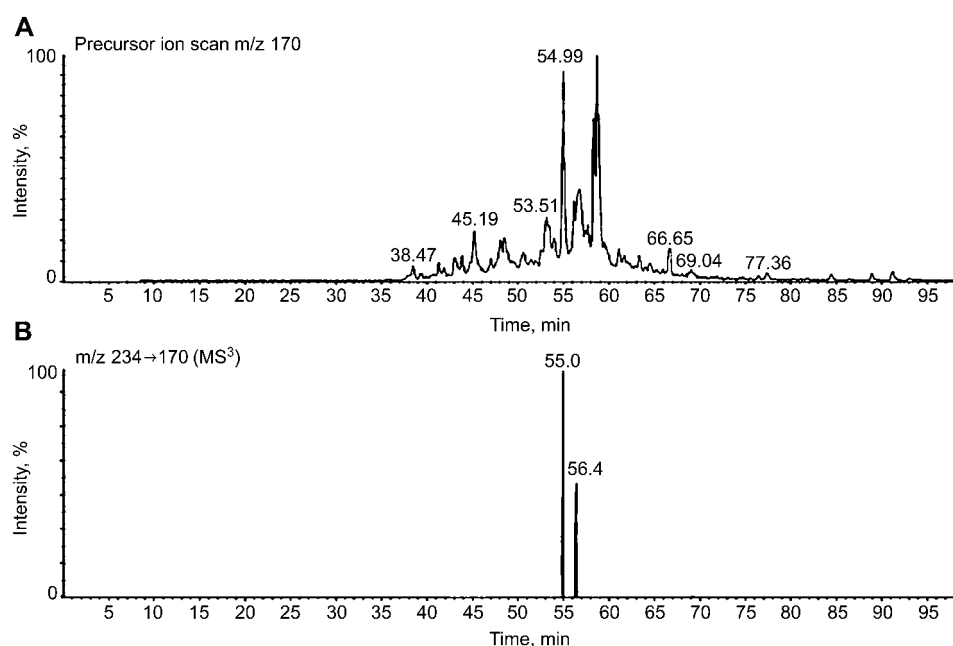


Figure 10 Ion chromatogram for PIS mode analysis (A) and reconstructed ion chromatogram for the transition m/z 234/170 in MS³ mode (B).

The MS³ TIC showed the presence of only two peaks with a large increase in the signal-to-noise ratio.

Nitration of proteins modulates catalytic activity, cell signaling and cytoskeleton organization. This bulky anionic constituent can induce changes in protein conformation, resulting in the generation of antigen epitopes, altered enzyme catalytic activity, modulation of metabolic pathways and inhibition of tyrosine phosphorylation by protein kinases. NO₂Tyr is considered the most important biomarker for identification of cellular processes associated with RNS formation leading to PTMs of proteins. RNS is the acronym for Reactive Nitrogen Species used to indicate all chemicals derived by nitrogen metabolism. These chemicals have a crucial role in regulation of many biological functions, but if their concentration increases too much, these species might initiate various types of cellular damage. It is well-known that some PTMs due to oxidative stress are linked to the pathogenesis of different diseases (109). In this respect, it is important to find biological markers that can provide information on the localization and physiological effects of oxidative stress. Molecules modified by oxidation, nitration and nitrosilation can act as such cellular biomarkers.

For these reasons many research groups have focused their attention on 3-nitrotyrosine residues, developing various methodologies for identification of NO₂Tyr residues. Protein nitration has been observed in connection with more than 60 human disorders including Alzheimer's disease, atherosclerosis, acute lung disease, amyotrophic lateral sclerosis (ALS) and others. Nitration is detected primarily by immunodetection using antibodies raised against nitrotyrosine (110).

From the limited number of experiments simulating *in vivo* conditions, it is apparent that tyrosine nitration is a selective process. Not all proteins or all tyrosine residues are targets for nitration (108). However, neither the abundance of a protein nor the abundance of tyrosine residues in a protein can predict nitration. Several factors such as the nitrating agent, protein folding, and the presence of glutamate in the local environment of the tyrosine residue are expected to contribute to nitration of specific tyrosine residues.

Nitroproteome analysis

Detection and quantification of nitroproteins has been accomplished essentially using 2D gel electrophoresis and Western blotting. In particular, a 2D gel containing an aliquot of the protein sample is stained with anti-3-nitrotyrosine antibody. The spots containing modified protein are selectively detected. The remaining portion of the sample can be loaded onto a preparative 2D gel stained either with colloidal Coomassie or Silver stain. Detection of nitrated protein spots is performed using image analysis by comparing the antibody stained and the preparative gels. The corresponding protein can then be identified using conventional mass spectral techniques. Although widely employed, this procedure has a number of drawbacks. Recognition of nitroproteins is done with an antibody which might show some non-specific detection. Moreover, when two or more proteins are present in the same spot, definitive identification of

the nitrated component cannot be accomplished. Finally, oftentimes localization of the site of modification within the nitrated protein sequence is impossible unless the modified peptides are detected during the identification procedure.

Other authors have employed methods using liquid chromatography separation combined with electrochemical detection or GC-MS techniques (111, 112). An alternative method to examining the overall amount of 3-nitrotyrosine in proteins is based on the use of electrospray MS. It has been shown by Petersson and co-workers that use of ESI-MS in PIS mode for detection of nitrated peptides in complex mixtures can be accomplished by taking advantage of characteristic fragment ions of modified peptides (113). The immonium ion of nitrotyrosine, in fact, provided unequivocal evidence for the presence of this modification, suggesting that mass spectral analyses of intact peptides might be a suitable analytical alternative.

The spectrum of nitrated peptides identified by use of MALDI shows a specific and unique pattern of molecular ions that can be used to identify this modification (113). The mass signal of the modified peptide occurs 45 Da higher than the unmodified specie due to substitution of an aromatic hydrogen with NO₂. This signal is usually accompanied by satellite ions that are 16, 30 and 32 Da lower in mass. This unique molecular ion pattern generated by photodecomposition during MALDI analysis provides a unique signature for peptides containing 3-nitrotyrosine. This fragmentation of nitrated peptides observed with MALDI-MS reduces the abundance of signals for nitrated peptides. This results in failure to detect modified peptides in very complex mixtures. It has been shown by Petersson and co-workers (113) that ESI-MS analysis of nitrated peptides prevents photodecomposition and allows the nitrated peptides to be specifically detected in PIS mode using the immonium ion of nitrotyrosine.

Dansyl labeling for the detection of 3-nitrotyrosine residues

We extended the RIGHt strategy to the selective isolation and identification of o-nitrotyrosine-containing proteins (114). The philosophy of the method consisted of the specific reduction of 3-nitroTyr to the corresponding aminotyrosine (NH₂-Tyr). The amino group of this residue is endowed with a particular low pK_a value, thus allowing selective labeling with DNS-Cl at variance with aliphatic (10.4–11.1) and N-terminal amino groups (6.8–8.0), which are largely protonated.

The methodology was first tested on bovine serum albumin (BSA) nitrated *in vitro* as a model protein and then applied to more complex matrices. BSA was nitrated *in vitro* with tetranitromethane (TNM) and the TNM-induced modifications were determined with MALDI MS. Spectral analyses led to 85% of BSA sequence coverage. More importantly, however, 95% of tyrosine residues could be verified. Two mass signals recorded at m/z 972.5 and m/z 1524.8 occurred 45 Da higher than the signals corresponding to the peptides 137–143 and 396–409, respectively. This

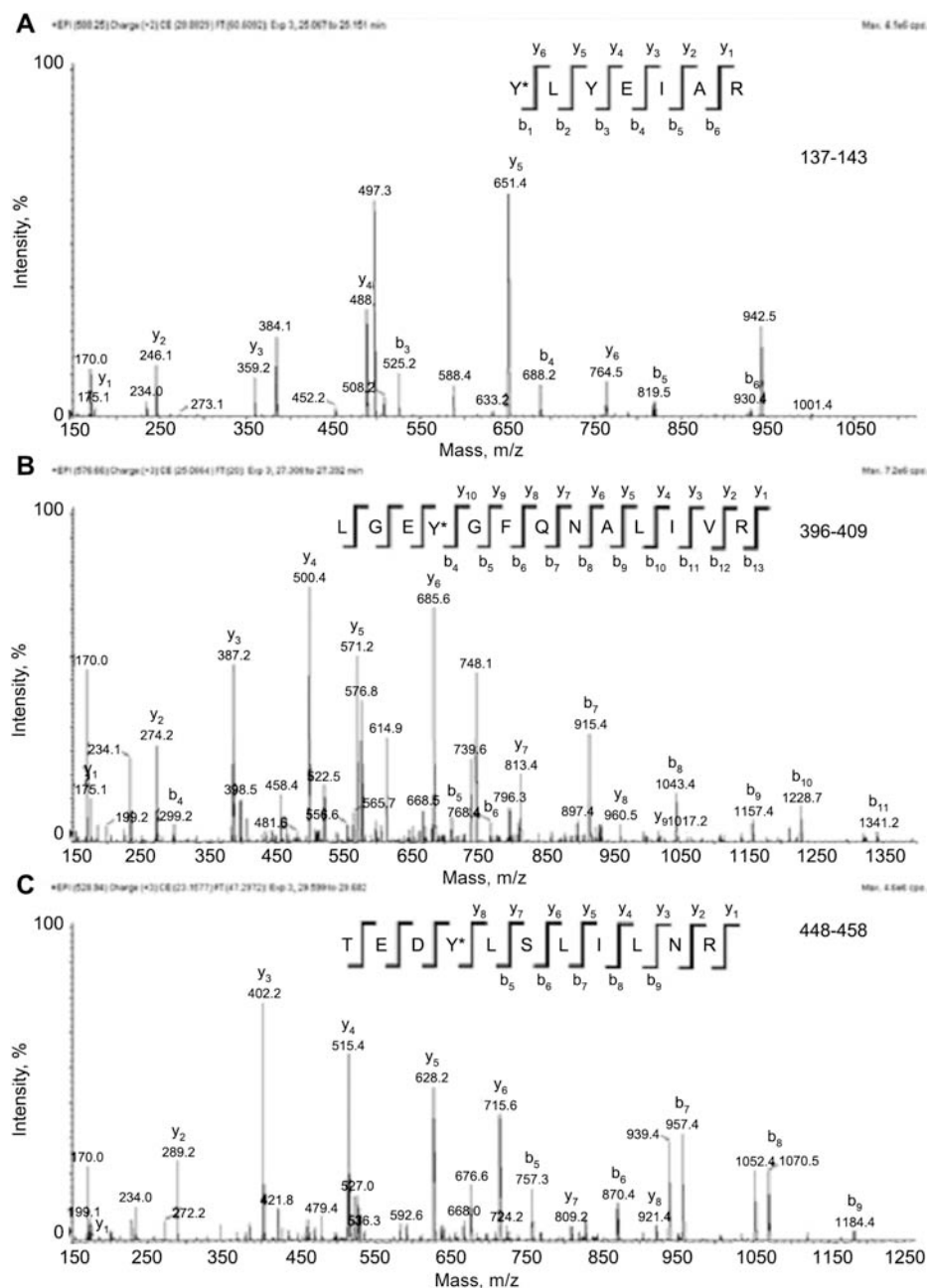


Figure 11 MS/MS spectra of nitrated BSA peptides 137–143 (A), 396–409 (B) and 448–458 (C).

Fragment ions are labeled as “y” or “b” according to the common peptide MS/MS fragmentation rules. Interpretation of the spectra led to the localization of the original nitration sites and to the determination of the peptide sequence. Unlabeled signals originated from secondary fragmentation pathways. Dansylated Tyr residues located at the level of Tyr137, Tyr399, and Tyr451, are highlighted as Y* in the sequence.

suggests that they originated from the nitrated derivatives of these fragments. However, signals corresponding to the unmodified fragments were also detected, indicating that the nitration reaction was not quantitative.

Conversion of nitropeptides into their o-dansyl-amino derivatives was accomplished by treatment with $\text{Na}_2\text{S}_2\text{O}_4$. Aminotyrosine residues were then selectively labeled with DNS-Cl following the procedure reported by Cirulli et al. (115). The reaction was carried out at pH 5.0, exploiting the different pK_a -value of the o-aminotyrosine (4.7). MALDI-MS analysis showed the presence of two new signals occurring

233 Da higher than the o-aminotyrosine-containing peptides, corresponding to the expected dansyl derivatives. Signals corresponding to the unmodified o-aminotyrosine peptides indicated that the extent of dansyl modification was about 60%. No modification of the N-terminus or Lys residues was detected.

The exact location of the original nitro groups was assessed by LC-MS/MS analysis of the peptide mixture using the experimental procedure described above (101, 103). LC-MS/MS full scan analysis showed a highly complex mixture of peaks that prevented identification of the modified peptides. The total ion current (TIC) trace for the m/z 170 PIS dis-

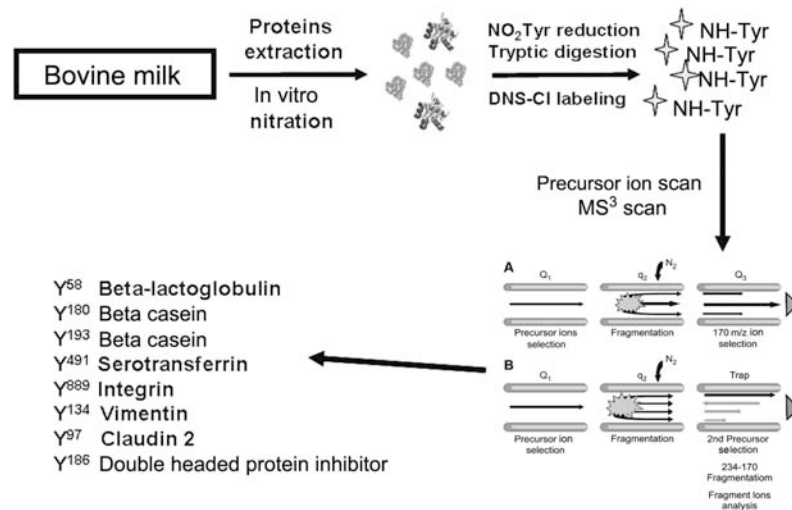


Figure 12 Flow chart of the analytical strategy for detecting labeled nitroproteins in bovine milk protein extract.

played a simplified chromatogram with significantly fewer number of signals. However, too many peaks were still present and it was very unlikely they could be related to dansyl derivatized peptides only. Finally, once the m/z 170 precursor ion and the m/z 234/170 transition scan selectivity criteria were applied, the MS³ TIC showed the presence of only three peaks with a large increase in the signal-to-noise ratio. The corresponding MS² fragmentation spectra enabled the determination of the entire sequence of these species and allowed definitive identification of the peptides. Of greater interest was the fact that the dansylated Tyr residues were located at the level of Tyr137, Tyr399 and Tyr451. The allowed for exact identification of the original nitration sites (Figure 11). It should be noted that LC-MS/MS analysis revealed the occurrence of a further nitration site at level of Tyr451 within the peptide 448–458, which had escaped previous MALDI analysis.

The MS/MS spectra of *o*-dansylamino-Tyr-containing peptides showed additional peculiar features. A stable fragment signal was always observed at m/z 384 corresponding to the immonium ion of *o*-dansylamino-Tyr. This could then be used either as an alternative diagnostic fragment in the PIS, or to confirm the presence of an *o*-dansylamino-Tyr residue in the peptide sequence.

The feasibility of the MS procedure to identify *o*-nitrotyrosine peptides in proteomics was evaluated by mixing 100 μ g of a mixture of BSA and nitrated BSA with 10 mg of the entire cellular extract from *Escherichia coli*. The total protein extract was then submitted to the procedure described above. The selection based on the MS³ scan removed a large number of false positives leading to a simple ion chromatogram that was dominated essentially by three intense signals. The MS² fragmentation spectra of these signals confirmed the occurrence of the same three BSA nitrated peptides detected previously.

The proposed strategy was finally employed to identify *o*-nitrotyrosine residues in a more complex sample, bovine milk (114). Bovine milk was nitrated

in vitro with TNM and an aliquot of milk proteins were fractionated using SDS-PAGE and submitted to Western blot analysis using antibody against *o*-nitrotyrosine to confirm the extent of nitration. The presence of nitrated proteins was detected essentially in the regions between 20–35 and 45–60 kDa. The entire milk protein extract was dissolved in denaturing buffer and cysteine alkylation and reduction of nitro groups were performed LC-MS/MS analysis using the double selectivity criteria led to the identification of nitropeptides from the high-abundant milk proteins, α -casein and α -lactoglobulin. In addition, it was able to assess the nitration sites occurring in proteins present in low concentration, such as vimentin and claudin 2 (Figure 12).

Conclusions

At the turn of the last century, the international scientific community witnessed a tremendous burst of interest in the area of expression proteomics. There was an enormous increase in numbers of papers being published, new journals being launched and large amounts of money invested by the pharmaceutical industry in search of new biomarkers of disease and therapeutic targets. In the last few years, interest has subsided slightly due to the lack of any new effective biomarkers being discovered. Also, new applications in proteomics have begun to emerge. It is easy to foresee that in the near future, most of proteomic research will be addressed toward two main goals; the investigation of cellular mechanisms through isolation of functional complexes, and qualitative and quantitative analysis of PTMs.

As discussed in this review, assessment of substoichiometric and transient PTMs on a proteomic scale is still a challenge. Existing procedures are not endowed with general applicability. The successful identification of PTMs is still dependent on a number of factors including the nature and the extent of the modifications, the amount of sample available, and

other variables. Despite advances in technology, the proteomic approach to PTMs still seem far from being satisfactory and productive.

The recent trend shown by research in this field suggests that scientists are moving along two main paths that might also be combined. One path involves particular attention to areas related to sample preparation and separation technologies, with particular emphasis on enrichment procedures. The goal is to develop selective methodologies able to increase the amount of modified samples when compared to unmodified ones.

The other path involves efforts made toward addressing the utilization of advanced analytical chemistry strategies focused on the development of new instrumentation and technologies in the field of MS. The introduction of sophisticated instrumentation, like the Orbitrap, or totally new concepts of ion separation such as the ion mobility MS might open up ways to utilize "bidimensional" or "multidimensional" mass spectral analysis able to identify modified peptide ions in very complex mixtures.

The integration of innovative enrichment procedures with these new mass spectrometric techniques should enable a new generation of proteomic experiments to be performed, thus providing a more general approach to analysis of PTMs. In this respect, specific chemical labeling of modified residues allowing for both selective enrichment with affinity tags, and sophisticated mass spectral experiments using reporter tag fragmentation and quantitative analysis by labeling with isotopes, will play a pivotal role in the next generation of proteomics.

References

- Godovac-Zimmermann J, Brown LR. Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom Rev* 2001;20:1–57.
- Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat Genet* 2003;33:311–23.
- Taylor SW, Fahy E, Ghosh SS. Global organellar proteomics. *Trends Biotechnol* 2003;21:82–8.
- Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001;10:437–73.
- Banks RE, Dunn MJ, Hochstrasser DF, Sanchez JC, Blackstock W, Pappin DJ, et al. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 2000;356:1749–56.
- Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001;19:379–82.
- Riederer BM. Non-covalent and covalent protein labeling in two-dimensional gel electrophoresis. *J Proteomics* 2008;71:231–44.
- Gesellchen F, Bertinetti O, Herberg FW. Analysis of post-translational modifications exemplified using protein kinase A. *Biochim Biophys Acta* 2006;1764:1788–800.
- Han KK, Martinage A. Post-translational chemical modifications of proteins – III. Current developments in analytical procedures of identification and quantitation of post-translational chemically modified amino acid(s) and its derivatives. *Int J Biochem* 1993;25:957–70.
- Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2299–301.
- Beavis RC, Chait BT. Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc Natl Acad Sci USA* 1990;87:6873–7.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989;246:64–71.
- Chait BT, Kent SB. Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science* 1992;257:1885–94.
- Lee TD, Moore RE, Young MK. Introducing samples directly into electrospray ionization mass spectrometers using a nanospray interface. *Curr Protoc Protein Sci* 2001;5:unit 16.9.
- Dal Piaz F, De Leo M, Braca A, De Simone F, Morelli I, De Tommasi N. Electrospray ionization mass spectrometry for identification and structural characterization of pregnane glycosides. *Rapid Commun Mass Spectrom* 2005;19:1041–52.
- Schwartz JC, Jardine I. Quadrupole ion trap mass spectrometry. *Methods Enzymol* 1996;270:552–86.
- Macek B, Waanders LF, Olsen JV, Mann M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol Cell Proteomics* 2006;5:949–58.
- Hager JW. A new linear ion trap mass spectrometer. *Rapid Commun Mass Spectrom* 2002;16:512–26.
- Hager JW, LeBlanc JC. High-performance liquid chromatography-tandem mass spectrometry with a new quadrupole/linear ion trap instrument. *J Chromatogr A* 2003;1020:3–9.
- Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984;11:601.
- Hanrieder J, Nyakas A, Naessén T, Bergquist J. Proteomic analysis of human follicular fluid using an alternative bottom-up approach. *J Proteome Res* 2008;7:443–9.
- Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep* 2005;25:45–56.
- Monti M, Orrù S, Pagnozzi D, Pucci P. Functional proteomics. *Clin Chim Acta* 2005;357:140–50.
- Zhang K, Wrzesinski K, Fey SJ, Mose Larsen P, Zhang X, Roepstorff P. Assessing CMT cell line stability by two dimensional polyacrylamide gel electrophoresis and mass spectrometry based proteome analysis. *J Proteomics* 2008;71:160–7.
- Li Z, Zhao X, Bai S, Wang Z, Chen L, Wei Y, et al. Proteomic identification of cyclophilin A as a potential prognostic factor and therapeutic target in endometrial carcinoma. *Mol Cell Proteomics* 2008. [Epub ahead of print].
- Pucci P, Carestia C, Fioretti G, Mastrobuoni AM, Pagano L. Protein fingerprint by fast atom bombardment mass spectrometry: characterization of normal and variant human haemoglobins. *Biochem Biophys Res Commun* 1985;130:84–90.
- Mead JA, Shadforth IP, Bessant C. Public proteomic MS repositories and pipelines: available tools and biological applications. *Proteomics* 2007;7:2769–86.
- Macek B, Waanders LF, Olsen JV, Mann M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol Cell Proteomics* 2006;5:949–58.
- Gafken PR, Lampe PD. Methodologies for characterizing phosphoproteins by mass spectrometry. *Cell Commun Adhes* 2006;13:249–62.
- Walsh TC. Posttranslational modification of proteins: expanding nature's Inventory. Roberts and Company Publishers, 2006:490 pp.

31. Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr. Protein post-translational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 2005; 44:7342–72.
32. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 2002;12:43R–56R.
33. Qiu R, Regnier FE. Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum. *Anal Chem* 2005;77:7225–31.
34. Rademacher TW, Parekh RB, Dwek RA. *Glycobiology*. *Annu Rev Biochem* 1988;57:785–838.
35. Marshall RD. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* 1974;17–26.
36. Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 1985;54:631–64.
37. Parodi AJ. Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J* 2000;348:1–13.
38. Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science* 2001;291:2351–6.
39. Perez-Vilar J, Hidalgo J, Velasco A. Presence of terminal N-acetylgalactosamine residues in subregions of the endoplasmic reticulum is influenced by cell differentiation in culture. *J Biol Chem* 1991;266:23967–76.
40. Spiro RG. Glucose residues as key determinants in the biosynthesis and quality control of glycoproteins with N-linked oligosaccharides. *J Biol Chem* 2000;275:35657–60.
41. Helenius A. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell* 1994;5:253–65.
42. Trombetta ES. The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* 2003;13:77R–91R.
43. Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 1993;3:97–130.
44. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* 1999;9:747–55.
45. Karlsson KA. Microbial recognition of target-cell glycoconjugates. *Curr Opin Struct Biol* 1995;5:622–35.
46. Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science* 2001;291:2351–6.
47. Bretthauer RK. Characterization of O-linked saccharides on glycoproteins. *Methods Mol Biol* 2007;389:107–18.
48. Xie Y, Liu J, Zhang J, Hedrick JL, Lebrilla CB. Method for the comparative glycomic analyses of O-linked, mucin-type oligosaccharides. *Anal Chem* 2004;76:5186–97.
49. Freeze HH, Kranz C. Endoglycosidase and glycoamidase release of N-linked oligosaccharides. *Curr Protoc Protein Sci* 2006;12.
50. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 2003;21:660–6.
51. Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A* 2004;1053:79–88.
52. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM, et al. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J Proteome Res* 2006;5:1792–802.
53. An HJ, Miyamoto S, Lancaster KS, Kirmiz C, Li B, Lam KS, et al. Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J Proteome Res* 2006;5:1626–35.
54. Durham M, Regnier FE. Targeted glycoproteomics: serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome. *J Chromatogr A* 2006;1132:165–73.
55. Dalpathado DS, Desaire H. Glycopeptide analysis by mass spectrometry. *Analyst* 2008;133:731–8.
56. Ciardiello MA, D'Avino R, Amoresano A, Tuppo L, Carpentieri A, Carratore V, et al. The peculiar structural features of kiwi fruit pectin methyltransferase: amino acid sequence, oligosaccharides structure, and modeling of the interaction with its natural proteinaceous inhibitor. *Proteins* 2008;71:195–206.
57. Amoresano A, Siciliano R, Orrù S, Napoleoni R, Altarocca V, De Luca E, et al. Structural characterisation of human recombinant glycohormones follitropin, lutropin and choriogonadotropin expressed in Chinese hamster ovary cells. *Eur J Biochem* 1996;242:608–18.
58. De Lorenzo C, Cozzolino R, Carpentieri A, Pucci P, Laccetti P, D'Alessio G. Biological properties of a human compact anti-ErbB2 antibody. *Carcinogenesis* 2005;26:1890–5.
59. Lityńska A, Pocheć E, Hoja-Lukowicz D, Kremser E, Laidler P, Amoresano A, et al. The structure of the oligosaccharides of alpha3beta1 integrin from human ureter epithelium (HCV29) cell line. *Acta Biochim Pol* 2002; 49:491–500.
60. Inforzato A, Peri G, Doni A, Garlanda C, Mantovani A, Bastone A, et al. Structure and function of the long pentraxin PTX3 glycosidic moiety: fine-tuning of the interaction with C1q and complement activation. *Biochemistry* 2006;45:11540–51.
61. Zaia J. Mass spectrometry of oligosaccharides. *Mass Spectrom Rev* 2004;23:161–227.
62. Harvey DJ. Matrix-assisted laser desorption/ionisation mass spectrometry of oligosaccharides and glycoconjugates. *J Chromatogr A* 1996;720:429–46.
63. Domon B, Costello CE. A systematic nomenclature for carbohydrate fragmentations in FAB/MS of glycoconjugates. *Glycoconjugate J* 1988;5:397–409.
64. Sandra K, Devreese B, Van Beeumen J, Stals I, Claeysens M. The Q-Trap mass spectrometer, a novel tool in the study of protein glycosylation. *J Am Soc Mass Spectrom* 2004;15:413–23.
65. Meldrum DR. Tumor necrosis factor in the heart. *Am J Physiol* 1998;274:577–595.
66. Renkonen J, Tynneninen O, Häyry P, Paavonen T, Renkonen R. Glycosylation might provide endothelial zip codes for organ-specific leukocyte traffic into inflammatory sites. *Am J Pathol* 2002;161:543–50.
67. Dube DH, Bertozzi CR. Glycans in cancer and inflammation – potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 2005;4:477–88.
68. Unkenborg J, Pilch BJ, Podtelejnikov AV, Wiśniewski JR. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* 2004;4: 454–65.
69. Nikov G, Bhat V, Wishnok JS, Tannenbaum SR. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Anal Biochem* 2003; 320:214–22.
70. Deonarain MP. Recombinant antibodies for cancer therapy. *Expert Opin Biol Ther* 2008;8:1123–41.
71. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX, Sutherland LL, et al. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J Proteome Res* 2008;7:1660–74.
72. Kumar GK, Prabhakar NR. Post-translational modification of proteins during intermittent hypoxia. *Respir Physiol Neurobiol* 2008. [Epub ahead of print].

73. Endoh H, Ishibashi Y, Yamaki E, Yoshida T, Yajima T, Kimura H, et al. Immunohistochemical analysis of phosphorylated epidermal growth factor receptor might provide a surrogate marker of EGFR mutation. *Lung Cancer* 2008. [Epub ahead of print].
74. Sabetkar M, Low SY, Bradley NJ, Jacobs M, Naseem KM, Richard Bruckdorfer K. The nitration of platelet vasodilator stimulated phosphoprotein following exposure to low concentrations of hydrogen peroxide. *Platelets* 2008;19:282–92.
75. Larrainzar E, Urarte E, Auzmendi I, Ariz I, Arrese-Igor C, González EM, et al. Use of recombinant iron-superoxide dismutase as a marker of nitrative stress. *Methods Enzymol* 2008;437:605–18.
76. Baty JW, Hampton MB, Winterbourn CC. Proteomic detection of hydrogen peroxide-sensitive thiol proteins in Jurkat cells. *Biochem J* 2005;390:791–2.
77. Härtig W, Seeger J, Naumann T, Brauer K, Brückner G. Selective in vivo fluorescence labelling of cholinergic neurons containing p75(NTR) in the rat basal forebrain. *Brain Res* 1998;808:155–65.
78. Julka S, Regnier F. Quantification in proteomics through stable isotope coding: a review. *J Proteome Res* 2004;3:350–63.
79. Amoresano A, Marino G, Cirulli C, Quemeneur E. Mapping phosphorylation sites: a new strategy based on the use of isotopically labelled DTT and mass spectrometry. *Eur J Mass Spectrom* 2004;10:401–12.
80. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 2002;277:1031–9.
81. Kalume DE, Molina H, Pandey A. Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* 2003;7:64–9.
82. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–61.
83. Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 2004;8:33–41.
84. Amoresano A, Monti G, Cirulli C, Marino G. Selective detection and identification of phosphopeptides by dansyl MS/MS/MS fragmentation. *Rapid Commun Mass Spectrom* 2006;20:1400–4.
85. Marino G, Buonocore V. Mass-spectrometric identification of 1-dimethylaminoaphthalene-5-sulphonyl-amino acids. *Biochem J* 1968;110:603–4.
86. LeBlanc JC, Hager JW, Ilisiu AM, Hunter C, Zhong F, Chu I. Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. *Proteomics* 2003;3:859–69.
87. Hager JW, Yves Le Blanc JC. Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. *Rapid Commun Mass Spectrom* 2003;17:1056–64.
88. Ben-Levy R, Leighton IA, Doza YN, Attwood P, Morrice N, Marshall CJ, et al. Identification of novel phosphorylation sites required for activation of MAPKAP kinase-2. *EMBO J* 1995;14:5920–30.
89. Bouras T, Southey MC, Venter DJ. Overexpression of the steroid receptor coactivator AIB1 in breast cancer correlates with the absence of estrogen and progesterone receptors and positivity for p53 and HER2/neu. *Cancer Res* 2001;61:903–7.
90. Meimoun P, Ambard-Bretteville F, Colas-des Francs-Small C, Valot B, Vidal J. Analysis of plant phosphoproteins. *Anal Biochem* 2007;371:238–46.
91. McLachlin DT, Chait BT. Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr Opin Chem Biol* 2001;5:591–602.
92. Lewandrowski U, Sickmann A, Cesaro L, Brunati AM, Toninello A, Salvi M. Identification of new tyrosine phosphorylated proteins in rat brain mitochondria. *FEBS Lett* 2008;582:1104–10.
93. Paradela A, Albar JP. Advances in the analysis of protein phosphorylation. *J Proteome Res* 2008;7:1809–18.
94. Pandey A, Andersen JS, Mann M. Use of mass spectrometry to study signaling pathways. *Sci STKE* 2000;2000:PL1.
95. Janek K, Wenschuh H, Bienert M, Krause E. Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2001;15:1593–9.
96. Idris N, Carothers Carraway CA, Carraway KL. Differential localization of ErbB2 in different tissues of the rat female reproductive tract: implications for the use of specific antibodies for ErbB2 analysis. *J Cell Physiol* 2001;189:162–70.
97. Bendt AK, Burkovski A, Schaffer S, Bott M, Farwick M, Hermann T. Towards a phosphoproteome map of *Corynebacterium glutamicum*. *Proteomics* 2003;3:1637–46.
98. Gaberc-Porekar V, Menart V. Perspectives of immobilized-metal affinity chromatography. *J Biochem Biophys Methods* 2001;49:335–60.
99. Andersson L, Porath J. "Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography". *Anal Biochem* 1986;154:250–4.
100. Zhou H, Ye M, Dong J, Han G, Jiang X, Wu R, et al. Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J Proteome Res* 2008. [Epub ahead of print].
101. Wilson-Grady JT, Villén J, Gygi SP. Phosphoproteome analysis of fission yeast. *J Proteome Res* 2008;7:1088–97.
102. Zhang X, Ye J, Jensen ON, Roepstorff P. Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. *Mol Cell Proteomics* 2007;6:2032–42.
103. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001;19:379–82.
104. Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* 2001;19:375–8.
105. Mattila K, Siltainsuu J, Balaspiri L, Ora M, Lönnberg H. Derivatization of phosphopeptides with mercapto- and amino-functionalized conjugate groups by phosphate elimination and subsequent Michael addition. *Org Biomol Chem* 2005;3:3039–44.
106. Beckman JS, Koppenol WH. Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly. *Am J Physiol* 1996;271:1424–37.
107. Schopfer FJ, Baker PR, Freeman BA. NO-dependent protein nitration: a cell signaling event or an oxidative inflammatory response? *Trends Biochem Sci* 2003;28:646–54.
108. Ischiropoulos H. Biological tyrosine nitration: a pathophysiological function of nitric oxide and reactive oxygen species. *Arch Biochem Biophys* 1998;356:1–11.
109. Greenacre SA, Ischiropoulos H. Tyrosine nitration: localisation, quantification, consequences for protein function and signal transduction. *Free Radic Res* 2001;34:541–81.
110. Castegna A, Thongboonkerd V, Klein JB, Lynn B, Markesbery WR, Butterfield DA. Proteomic identification of nitrated proteins in Alzheimer's disease brain. *J Neurochem* 2003;85:1394–401.
111. Aulak KS, Koeck T, Crabb JW, Stuehr DJ. Dynamics of protein nitration in cells and mitochondria. *Am J Physiol Heart Circ Physiol* 2004;286:30–38.

112. Nikov G, Bhat V, Wishnok JS, Tannenbaum SR. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Anal Biochem* 2003;320:214–22.
113. Petersson AS, Steen H, Kalume DE, Caidahl K, Roepstorff P. Investigation of tyrosine nitration in proteins by mass spectrometry. *J Mass Spectrom* 2001;36:616–25.
114. Amoresano A, Chiappetta G, Pucci P, D'Ischia M, Marino G. Bidimensional tandem mass spectrometry for selective identification of nitration sites in proteins. *Anal Chem* 2007;79:2109–17.
115. Cirulli C, Marino G, Amoresano A. Membrane proteome in *Escherichia coli* probed by MS3 mass spectrometry: a preliminary report. *Rapid Commun Mass Spectrom* 2007;21:2389–97.



Honouring Andries P. Bruins on the occasion of his 65th birthday

Glycoproteome study in myocardial lesions serum by integrated mass spectrometry approach: preliminary insights

Andrea Carpentieri,^a Chiara Giangrande,^a Piero Pucci and Angela Amoresano

Dipartimento di Chimica Organica e Biochimica, Università degli Studi Federico II, Complesso Universitario Monte S. Angelo, via Cynthia 4, 80126 Napoli Italy. E-mail: angamor@unina.it

Bottom up proteomics requires efficient and selective pre-fractionation procedures to simplify the analysis of the enormous number of peptides resulting from the hydrolysis of a cellular extract enabling the detection, identification and the structural characterization of the post-translational modifications. Glycosylation, a well-known post-translational modification, plays a key role in the enormous complexity, and heterogeneity of the human blood serum proteome. Thereby, characterization of glycosylation from serum is a challenging task, even for the existing sophisticated analytical methodologies. Here we report a glycoproteomics study on the identification of even low abundant glycoproteins, including the localization of N-glycosylation sites and the glycan profiling in human sera from healthy and myocarditis affected donors. The strategy is simply based on proteolytic digestion of total serum proteins followed by a single enrichment step of glycopeptides on ConA lectin affinity chromatography. Glycopeptides were then deglycosylated by PNGaseF treatment and nano-liquid chromatography-electrospray ionization tandem mass spectrometry analyses of the free peptides provided the basis for both identification of the individual proteins and elucidation of their modification sites. Moreover, glycan profilings could be obtained by matrix-assisted laser desorption/ionization mass spectrometry analysis of the released oligosaccharides. Our data led to the identification of 68 different glycosylation sites within 49 different proteins. Moreover, the analyses carried out on glycans represent the first picture of a glycosylation pattern in myocardial lesions. As a whole, several differences in the glycosylation patterns from different sera were observed, thus indicating glycan profiling as a possible tool to discriminate among different diseases.

Keywords: human sera, glycoproteomics, mass spectrometry

Introduction

The importance of investigation of post-translational modifications (PTM) is notably increased in the proteomic era, as they can affect biological functions, thus playing a critical role in cellular functioning. Moreover, they can vary in response to environmental stimuli or signalling modulators, thus finely

tuning cellular mechanisms and their deregulation might be involved in the development of diseases. A huge number of different types of PTMs have been identified. The pattern of PTMs on proteins constitute a molecular code that dictates protein conformation, cellular location, macromolecular interactions and activities, depending on cell type, tissue and environmental conditions.

Glycomics shows a higher level of complexity than genomics or proteomics, as glycans have a greater number of building

^aAndrea Carpentieri and Chiara Giangrande contributed equally to the work

blocks than nucleic acids (four nucleotides) and proteins (20 amino acids). To make the process more complex, oligosaccharide biosynthesis is not template-driven and glycans can adopt complex branched structures of monomeric units linked together, whose structures have different conformations.¹ The microheterogeneity of the oligosaccharides is essentially due to the reactions involved in glycosylation and deglycosylation at specific sites being frequently incomplete² and depends on the cell type and on the physiological status. Anomalous patterns of glycosylation may be caused by developmental events or pathological conditions.³⁻⁷ The problem of glycan heterogeneity has been overcome by mass spectrometry methodologies, which allow qualitative and semi quantitative information on oligosaccharides structures present in the mixture to be achieved with high sensitivity.⁸

The first step in glycoprotein identification is the isolation of glycosylated proteins from the remainder of the proteome. Several procedures have been developed aimed at glycoprotein enrichment and analysis including enzymatic and metabolic methods to chemically tag proteins to enable their isolation. Hydrazide trapping coupled to lectin affinity chromatography has also been proposed⁹ as well as combined multilectin affinity columns.¹⁰

Once isolated, glycoproteins can be identified by mass spectrometry. Additional information can be obtained by using either enzymatic or chemo selective reactions to incorporate isotope labels at specific sites of glycosylation. Isotopic labelling facilitates mass spectrometry-based confirmation of glycoprotein identity, identification of glycosylation sites and quantification of the extent of modification.¹¹

Several reports suggested that the alterations in protein glycosylation on the cell surface and in body fluids might be associated with the development of certain types of human cancer and other diseases (Reference 12 and references therein). Comparative serum glycoproteomics have been applied to the study of pathological samples, using a multilectin affinity approach, in order to identify oligosaccharide markers in such diseases.^{13,14} A multi step method for broad analysis of human plasma N-glycoproteins was proposed by Smith and co-workers,¹⁵ based on a combination of immunaffinity subtraction and glycoprotein capture to reduce both the protein concentration range and the overall sample complexity.

Here, we report preliminary insights on the comparison of the glycoproteomes in healthy and myocarditis human sera, by using a single affinity chromatography (ConA) step coupled with mass spectrometry techniques. This strategy led to both the identification of 68 different glycosylation sites within 49 different proteins and to the definition of the glycosylation patterns. Moreover, to the best of our knowledge, this study represents the first definition of glycoform distribution in myocarditis. The analysis of glycan profiling, once extracted from serum glycopeptides, is essential for comparative studies on different sera samples, thus providing a useful tool for the development of screening procedures.

Materials and methods

Materials

Human serum samples from eight healthy donors and from five patients diagnosed with myocardial lesions have been obtained from the "Servizio Analisi" Policlinico, Napoli, Italy.

Guanidine, dithiothreitol (DTT), trypsin, methyl- α -D-mannopyranoside, α -cyano-4-hydroxycinnamic acid, 2,4,6-trihydroxyacetophenone and sodium chloride were purchased from Sigma. Iodoacetamide (IAM), tris(hydroxymethyl)aminomethane, calcium chloride and ammonium bicarbonate (AMBIC) were purchased from Fluka as well as the MALDI matrix 2,5-dihydroxybenzoic acid (DHB). Methanol, trifluoroacetic acid (TFA) and acetonitrile (ACN) are HPLC grade type from Carlo Erba, whereas the other solvents are from Baker. PD-10 gel filtration columns are from Pharmacia, the HPLC columns are from Phenomenex, whereas the pre-packed columns Sep-pak C-18 are from Waters.

Comassie Brilliant Blue was from Bio-Rad. PNGase F were purchased from Boehringer. Ion exchange resins Dowex H+ (50W-X8 50-100 mesh) were provided by BDH. Concanavalin A sepharose resin was purchased from Amersham Biosciences.

Protein concentration determination

Sera protein concentration was determined by the Bradford assay method, using bovine serum albumin (BSA) as standard. Known amounts of BSA were diluted in 800 μ L of H₂O and then mixed to 200 μ L of Comassie Brilliant Blue. Five different BSA concentrations were determined by measuring absorbance at 595 nm and used to obtain a linear calibration curve. Three different sera dilutions were measured at 595 nm. Absorbance data were interpolated on the calibration curve, allowing the determination of protein concentration in the different samples.

Reduction and carbamidomethylation

An aliquot (1 mg) of each serum was lyophilized and dissolved in 200 μ L of denaturation buffer (guanidine 6 M, tris 0.3 M, EDTA 10 mM, pH 8). Reduction was carried out by using a 10:1 DTT:cysteins molar ratio. After incubation at 37°C for 2 h, carboxyamidomethylation was performed by adding iodoacetamide in an excess of alkylating agent of 5:1 with respect to the moles of thiol groups. The mixture was then incubated in the dark at room temperature for 30 min. The alkylation reaction was stopped by the addition of formic acid, in order to achieve an acidic pH.

The excess of salts and reagents was then removed by gel filtration on PD-10 columns. Elution was performed using ammonium bicarbonate (AMBIC) 10 mM buffer. The collected fractions were analyzed by a Beckman DU 7500 spectrophotometer, measuring the absorbance at 220 nm and 280 nm. Protein containing fractions were collected and concentrated in a centrifuge Speed-Vac.

Enzymatic digestions

Trypsin digestion was carried out in AMBIC 10 mM buffer using a 50:1 trypsin:protein mass ratio (w/w). The sample was incubated at 37°C for 12–16 h.

Affinity chromatography

Peptide mixtures were incubated in batches with 50 μ L of ConA sepharose resin, which was activated with calcium ion in 20 mM tris-HCl buffer; 0.5 M NaCl; 1 mM CaCl_2 at pH 7.4, 4°C, for 2 h, as suggested by the manufacturer. Then the ConA containing eppendorf was centrifuged at 3000 rpm at 4°C. The unbound peptides were removed in the supernatant. The resin was then washed six times with the same buffer described before.

The elution phase was based on methyl- α -D-mannopyranoside 0.25 M competition with glycopeptides and was performed for 30 min at 4°C. This procedure was repeated twice.

RP-HPLC

Both the peptides and the glycopeptides needed a desalting step in order to remove salts and methyl- α -D-mannopyranoside.

Both the unbound and eluted peptides were desalted by HPLC on a C-18 resin. HPLC was performed on an Agilent chromatograph, at a flow rate of 0.2 mL min⁻¹ and a linear gradient from 5% B (95% acetonitrile, 5% H₂O and 0.07% trifluoroacetic acid) to 95% B in 1 min, allowing the fast elution of the unfractionated peptides.

Deglycosylation

Glycopeptides were lyophilized, resuspended in 10 mM AMBIC and incubated with PNGase F (5 U), for 12–16 h at 37°C. Deglycosylation was also carried out on unbound peptides, in order to release the glycans not recognized by ConA.

Oligosaccharide purification

For glycan purification, the sample was loaded on a C-18 silica-based bonded phase (Sep-pak). The column was activated with 10 mL methanol and equilibrated with 10 mL isopropanol, 10 mL H₂O and 10 mL acetic acid 5%. The oligosaccharide component was eluted in 3 mL of 5% acetic acid as eluent whereas peptides were collected in 3 mL of 40% isopropanol and 5% acetic acid. Alternatively, a reversed-phase, high-performance liquid chromatography (RP-HPLC) step was performed by using a C-18 column and a gradient profile as follows: 7 min at 5% B, 5 to 95% B in 2 min and 10 min at 95% B.

Glycans matrix-assisted laser desorption/ionization (MALDI)-profiling and post source decay (PSD) analyses

Glycans were collected, lyophilized, resuspended in 50 μ L of 5% acetic acid and submitted to batch ion exchange chromatography using a ready-for-use regenerable mixed bed Dowex MB50 resin, in order to eliminate salts and other contaminants. The sample prepared above was filtrated and mixed with an equal amount of resin solution in a test

tube. The tube was then gently stirred and centrifuged. The supernatant was recovered and used for MALDI-MS analyses.

Both positive and negative MALDI-MS analyses and MALDI PSD experiments were carried out on a Voyager DE-STR-Pro instrument operating in reflectron mode (Applied Biosystems, Framingham, MA, USA). The MALDI matrices were prepared by dissolving 25 mg of DHB (2,5 dihydroxy benzoic acid) or THAP (2,4,6-trihydroxyacetophenone) in 1 mL water/acetonitrile 10:1 or 10 mg α -cyano-4-hydroxycinnamic acid in 1 mL of acetonitrile/0.2% trifluoroacetic acid (70:30 v/v). Typically, 1 μ L of matrix was applied to the metallic sample plate and 1 μ L of analyte was then added. In positive ion mode, the target voltage was set at 20 kV, the first grid at 66% of target voltage and delayed extraction was 200 ns; in negative ion mode the target voltage was -20 kV, the first grid 60% of target voltage and delayed extraction was fixed at 400 ns.

Conditions were optimised to obtain the best signal-to-noise ratios and the best possible isotopic resolution with two-point external calibration using maltoheptose (exact isotopic mass, 1153.02 Da) and tetrantennary sialylated oligosaccharide (exact isotopic mass, 2686.03 Da) as calibrants, resulting in a mass accuracy of 300 ppm. Each spectrum represents the sum of 1500 laser pulses from randomly chosen spots per sample position. All analyses were performed in triplicate.

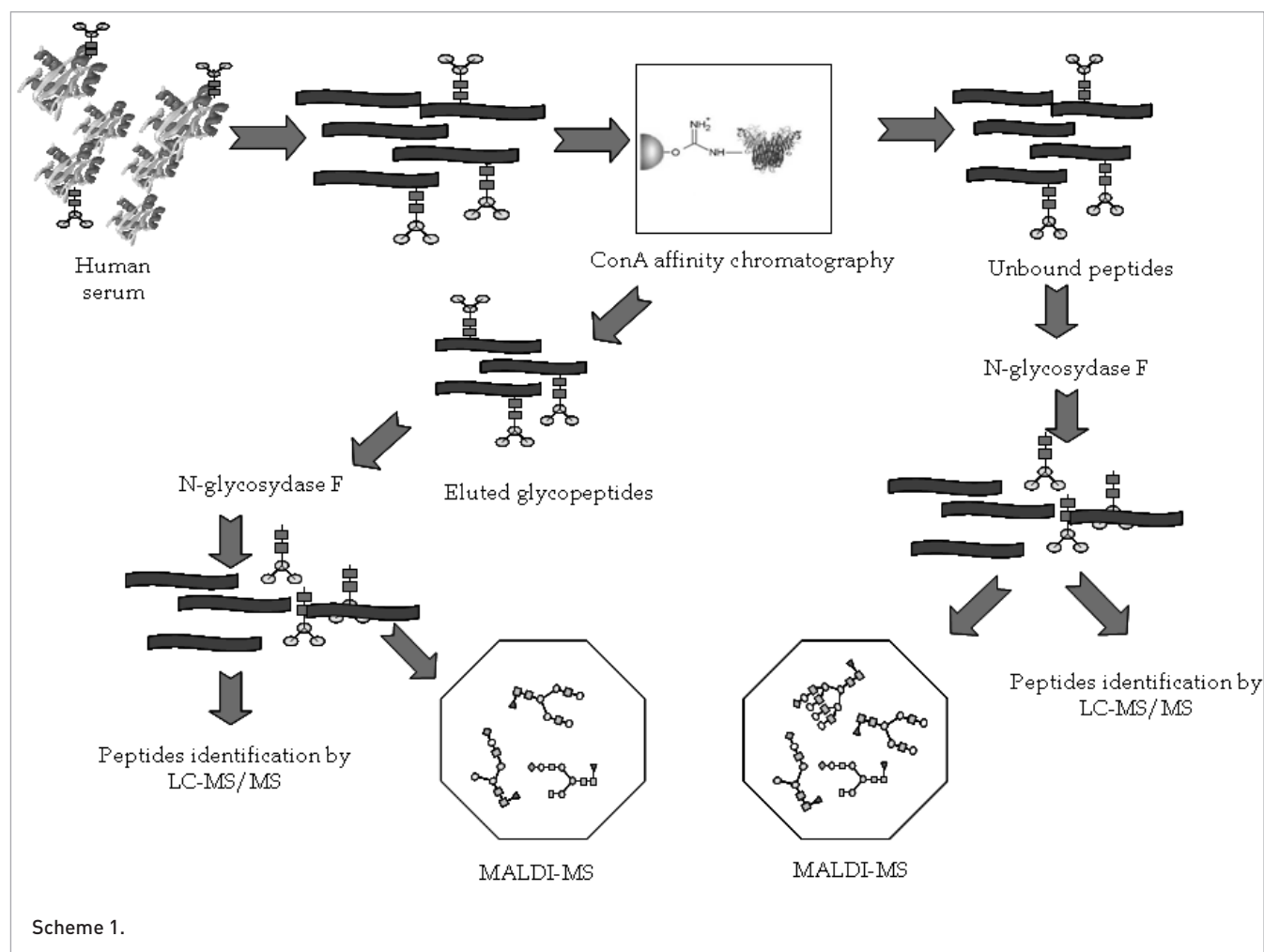
The PSD fragment spectra were acquired after pseudo molecular cation selection using the time ion selector. Fragment ions were refocused by stepwise reducing the reflector voltage by 20%. The individual segments were then stitched together using the Applied Biosystems software. Raw data were analyzed using the computer software provided by the manufacturers and are reported as mono-isotopic masses.

Peptide liquid chromatography tandem mass spectrometry (LC-MS/MS)

The samples were diluted in formic acid 0.1% in order to obtain a final concentration of 20 fmol μ L⁻¹. Eight μ L were loaded and analyzed.

NanoLC-MS² experiments were performed on a 4000 Q-Trap mass spectrometer (Applied Biosystems) coupled to an 1100 nanoHPLC system (Agilent Technologies). Peptides were separated on an Agilent reversed-phase column (Zorbax 300 SB C-18, 150 mm, 3.5 μ m, 0.75 μ m), at a flow rate of 0.2 μ L min⁻¹ using 0.1% formic acid, 2% ACN in water as solvent A 0.1% and formic acid, 2% water in ACN as solvent B. The elution was accomplished by a 5–65% linear gradient of solvent B in 60 min. A micro-ion spray source was used at 2.5 kV with liquid coupling, with a declustering potential of 50 V, using an uncoated silica tip from New Objectives (Ringoos, NJ, USA).

Proteins were identified by searching the MS/MS spectra against the SwissProt or NCBI databases using the Mascot in house search engine (Matrix Science, London, UK).



Results and discussion

This study was focused on a simple and rapid procedure, depicted in Scheme 1, to obtain an overview of the glycosylation profiling in different human sera. This was achieved by enriching for the N-linked glycopeptides resulting from tryptic digestion of sera samples in order to enhance the identification of N-glycosylation sites using LC-MS/MS and obtaining glycans distribution by MALDI-MS analysis. The strategy was applied to the analysis of the glycoproteome from myocarditis sera using sera from healthy donors as control.

Serum samples from patients suffering from myocardial lesions represent a very good example of inflammatory disease. Myocarditis is a collection of diseases of infectious, toxic and autoimmune etiologies, characterized by inflammation of the heart, leading to myocytes necrosis. The initial insult, which causes the death of certain cells, is followed by macrophage and natural killer cells activation, which continue myocyte necrosis.^{16,17} Being an inflammatory disease, the glycan moiety might consist of oligosaccharide epitopes responsible for the leukocytes rolling along the damaged tissues. To the best of our knowledge, no data on the characterization of glycosylation patterns in myocarditis has been reported, yet.

The study was addressed to two main goals, (i) identification of N-glycosylation sites in serum glycoproteins by LC-MS/MS and (ii) glycan profiling of normal and myocarditis serum samples.

Assessment of N-glycosylation sites in serum glycoproteins

Aliquots of serum samples from different donors were pooled in order to obtain an average distribution of glycoproteins. The analyses were performed on intact serum samples without any pre-purification step or removal of most abundant proteins. The two samples obtained corresponding to healthy and myocardial lesions donors were reduced, alkylated and digested with trypsin as described in the Materials and Methods section. According to the procedure developed, glycopeptides were selectively enriched by concanavalin A affinity Chromatography directly performed in batch after tryptic digestion. The recovered glycopeptides were then deglycosylated by PNGase F treatment and the peptide mixtures directly analysed by LC-MS/MS.

The presence of putative N-glycosylation sites was inferred by the increase of 1 Da in the peptide mass values due to

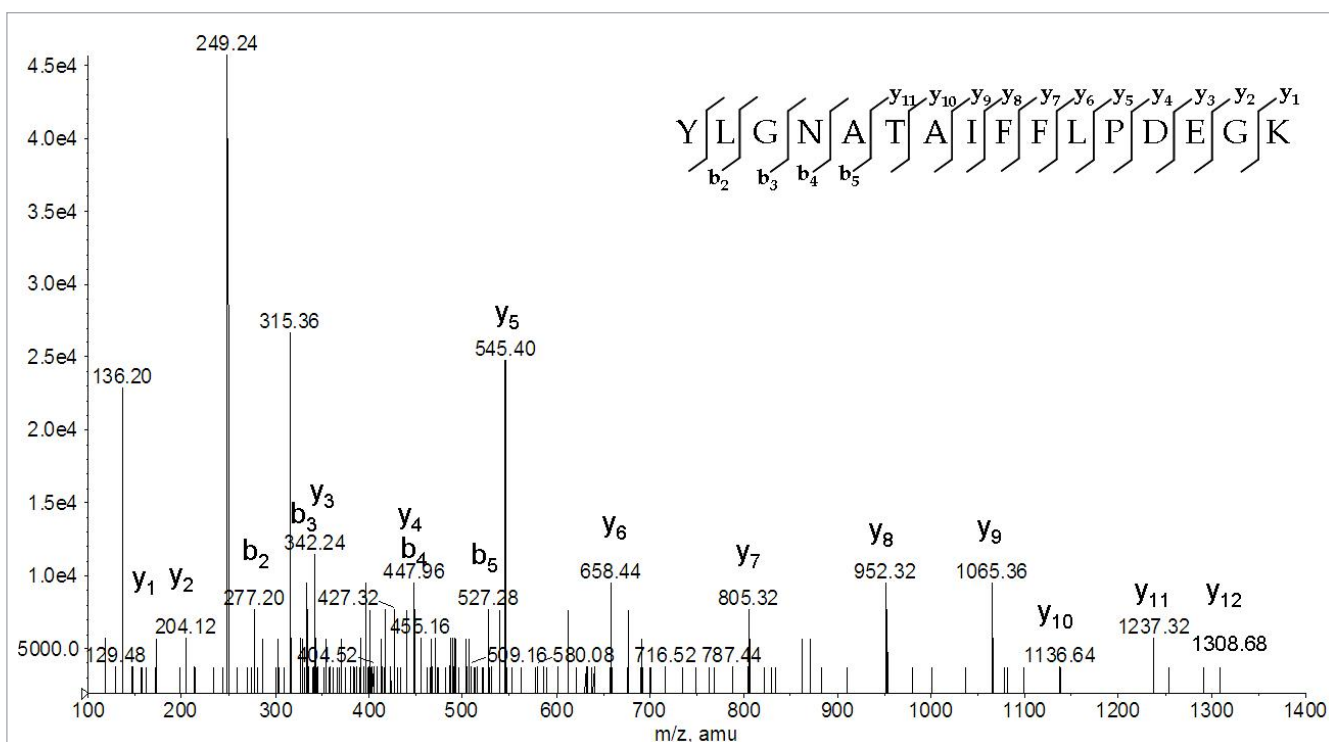


Figure 1. MS/MS spectrum of the doubly charged precursor ion at m/z 878.96 corresponding to the peptide YLGN*ATAIFFLPDEGK of alpha-1-antitrypsin (position 268–283 in the protein sequence). The fragmentation spectrum shows the presence of a N-glycosylation site at Asn271.

the conversion of Asn into Asp after PNGase F incubation. Fine location of the modification sites was then established by the occurrence of key signals in the MS/MS spectra. As an example, Figure 1 shows the daughter ion mass spectrum of a peptide retained by the ConA affinity medium and deglycosylated by PNGase F treatment. The MS/MS spectrum of the precursor ion at m/z 878.96 exhibited a clear series of y -ions from which the sequence of the peptide could easily be inferred as YLGN*ATAIFFLPDEGK. BLAST analysis of this sequence identified the peptide as the fragment 268–283 of the human alpha-1-antitrypsin precursor in which Asn271 shows the consensus sequence for glycosylation (Asn–X–Ser/Thr). The effective occurrence of a N-glycosylation site at Asn-271 was demonstrated by the key y^{12} fragment ion at m/z 1308.68 in the MS/MS spectrum. This fragment occurred 1 Da lower than expected showing the loss of Asp271 instead of Asn thus confirming the conversion of the asparagine to aspartic acid following PNGase F stripping.

Figure 2 shows the MS/MS spectrum of the triply charged precursor ion at m/z 536.52 corresponding to the fragment 3405–3419 of Apolipoprotein B. The series of y ions from which peptide identification was inferred is clearly visible in the daughter ion spectrum. Again, the N-glycosylation site could easily be assigned by the presence of the key y^8 fragment ion that occurred 1 Da lower than expected [836.40 instead of 837.47] demonstrating that Asn3411 was glycosylated and then had been converted to Asp by PNGase F.

Similar analyses were carried out on the unbound ConA fractions, mainly containing non-glycosylated peptides and highly branched glycopeptides. The unbound fraction was also deglycosylated by PNGase F treatment, yielding a mixture of deglycosylated and unmodified peptides that was directly analyzed by LC-MS/MS. Figure 3 shows the daughter ion spectrum of the precursor ion at m/z 628.38 that was assigned to the peptide 1004–1014 from the alpha-2-macroglobulin precursor. The key y^5 ion at m/z 624.48 clearly demonstrated the occurrence of an asparagine residue at position 1009 that resulted in being unmodified following the deglycosylation step.

The overall data from the LC-MS/MS analyses of both the ConA bound and unbound fractions were pooled and summarized in Table 1. A total of 49 glycoproteins were identified in this study which contained 68 N-glycosylation sites. Among these sites, 50 had already been described and were confirmed while 18 were only annotated as potential glycosylation sites due to the presence of the NX(S/T) consensus sequence and were then newly identified.

The results presented here demonstrated that ConA affinity chromatography on serum tryptic digests was effective in entrapping most of the glycopeptides with specific glycan structures and to enhance their detection by LC-MS/MS. A further advantage of the strategy relies on the lectin affinity chromatography being performed on glycopeptides instead of glycoproteins, avoiding fractionation of the extract

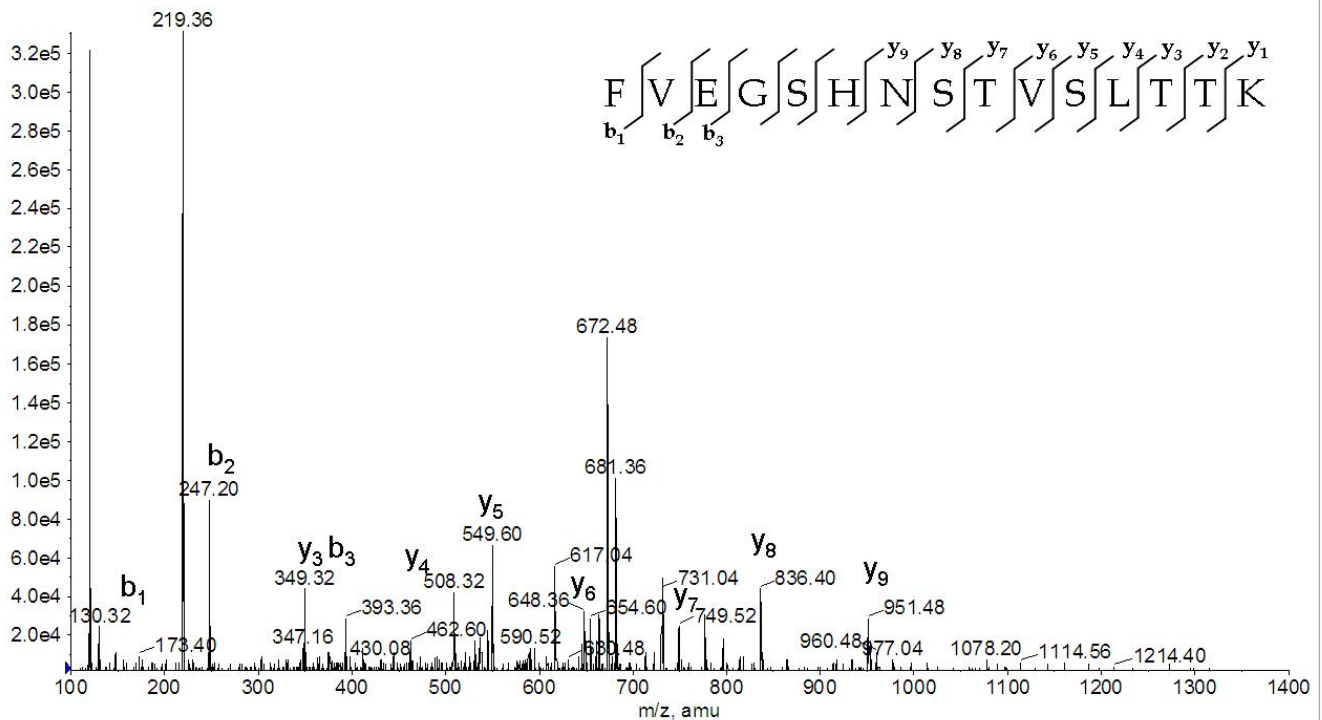


Figure 2. MS/MS spectrum of the triple charged precursor ion at m/z 536.14 corresponding to the peptide FVEGSHNSTVSLTTK of apolipoprotein B (position 3405–3419 in the protein sequence). The fragmentation spectrum shows the increment of 1 Da as Asn3411 was converted into Asp following PNGase F treatment.

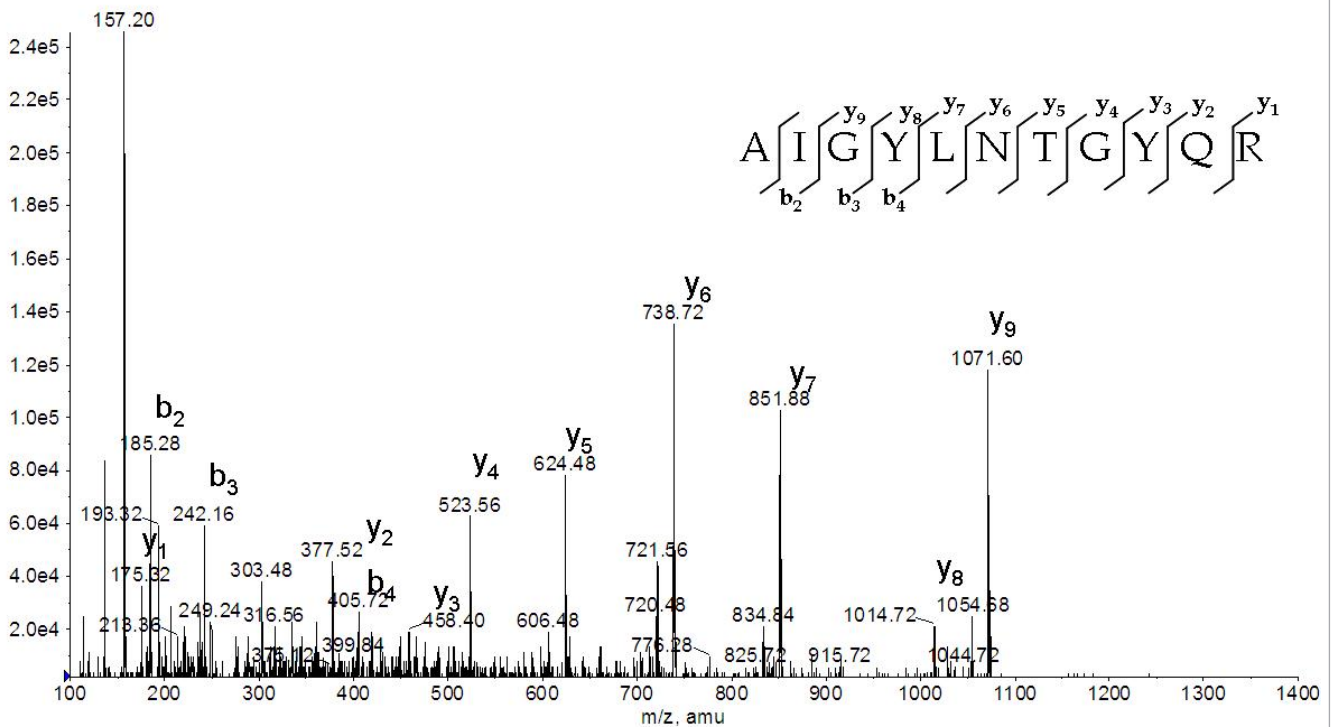


Figure 3. Fragmentation spectrum of the doubly charged precursor ion at m/z 628.38 corresponding to the peptide AIGYLNTGYQR of alpha-2-macroglobulin. The fragmentation pattern showed that the asparagine residue was unmodified in this peptide.

Table 1. Identification of glycosylation sites by LC-MS/MS analysis. (UH=unbound fraction of healthy serum; UL=unbound fraction of myocardial lesions serum; EH=bound fraction of healthy serum serum; EL=bound fraction of myocardial lesions serum).

	Protein	Sequence	Peptide	Charge	
P01011	Alpha-1-antichymotrypsin	YTGN*ASALFILPDQDK ^{UL}	268–283	2+	Confirmed
P02763	Alpha-1-acid glycoprotein	SVQEIQATFFYFTP N *KTEDTIFLR ^{EH}	58–81	3+	Confirmed
		QDQCIY N *TTYLNVQR ^{UH,UL}	87–101	2+	Confirmed
P01009	Alpha-1-antitrypsin	YLG N *ATAIFFLPDEGK ^{EH,UH,EL,UL}	268–283	2+,3+	Confirmed
P01023	Alpha-2-macroglobulin	SLGN VN *FTVSAEALSEQELCGTEVPSVPEHGRK ^{EH}	864–886	3+	Confirmed
P43652	Afamin	DIEN FN *STQK ^{EH,EL}	28–37	2+	Confirmed
		YAEDK FN *ETTEK ^{EH}	396–407	2+	Confirmed
P01008	Antithrombin III	LGAC N *DTLQQLMEVFK ^{EH}	124–139	2+	Confirmed
		SLTF N *ETYQDISELVYGAK ^{EH}	183–201		Confirmed
P04114	Apolipoprotein B-100	FN *SSYLQGTNQUITGR ^{EH,EL}	1522–1536	2+	Confirmed
		FVEGSH N *STVSLTTK ^{EH}	3405–3419	2+	Potential
		AEEMLE N *VSLVCPK ^{EL}	27–41	3+	Potential
		YDF N *SSMLYSTAK ^{EL}	3462–3474	2+	Confirmed
P05090	Apolipoprotein D	ADGTVNQIEGEATPV N *LTEPAK ^{EL}	83–104	3+	Confirmed
P02749	Apolipoprotein H	VYKPSAG N *NSLYR ^{EH}	155–167	3+	Confirmed
		LGN*WSAMP SCK ^{EL}	251–261	2+	Confirmed
075882	Attractin	IDSTG N *VTNELR ^{EH,EL}	411–422	2+	Confirmed
		GPVKMPSQAPTGNFYQP LLN *SSMCLEDSR ^{EH}	1023–1052	2+	Confirmed
P00450	Ceruloplasmin	EHEGAIYPD N *TTDFQR ^{EH,EL}	129–144	2+	Confirmed
		EN *LTAPGSDSAVFFEQGTR ^{EH,EL}	396–414	3+	Confirmed
P08603	Complement factor H	ISEE N *ETTCYMGK ^{EH}	907–919	2+	Confirmed
		MDGAS N *VTCINSR ^{EH,EL}	1024–1036	2+	Potential
		IPCSQPPQIEHGT I *SSR ^{EL}	868–885	3+	Confirmed
		SPDV I *GSPISQK ^{UH}	212–224	2+	Potential
P10909	Clusterin	LAN*LTQGEDQYYLR ^{EH,EL}	372–385	2+,3+	Confirmed
Q8IWW2	Contactin-4	LN *GTDVDTGMDFR ^{EL}	64–76	2+	Confirmed
P05156	Complement factor 1	FL NN *GTCTAEGK ^{EH,EL}	100–111	2+	Confirmed
P01024	Complement C3	TVLTPATNHMG N *VTFTIPANR ^{EH}	74–94	3+	Confirmed
P0C0L4	Complement C4-A	GL N *VTLSSSTGR ^{EH,EL}	1326–1336	2+	Confirmed
P02748	Complement component C9	AV N *ITSENLIDDVVSLIR ^{EH}	413–430	2+	Confirmed
Q14517	Protocadherin-fat 1	QVY N *LTVRAKDK ^{EL}	994–1005	3+	Potential
		FSMDYKTGALTVQ N *TTQLRSR ^{EL}	1930–1950	3+	Potential
P02765	Alpha-2-HS-glycoprotein	VCQDCPLLAP LN *DTR ^{EH}	145–159	2+,3+	Confirmed
		AALAAFNAQ NN *GSNFQLEEISR ^{EH}	166–187	2+,3+	Confirmed
Q03591	Complement factor H related protein 1	LQNN ENN *ISCVER ^{EH}	120–132	2+	Confirmed
Q13439	Golgin subfamily A member 4	HN *STLKQLMREFNTQLAQK ^{EH}	1990–2008	3+	Potential
P02790	Hemopexin	SWPAV GN *CSSALR ^{EH,EL}	181–193	2+	Confirmed
		ALPQPQ N *VTSLLGCTH ^{UH,EH,EL}	447–462	2+,3+	
P00738	Haptoglobin	VVLHP N *YSQVDIGLIK ^{UH,UL,EH,EL}	236–251	2+,3+	Confirmed
		MVSH HN *LTTGATLINEQWLLTTAK ^{EH,EL}	179–202	2+	Confirmed
P04196	Histidine rich glycoprotein	VEN*TTVYYLVLDVQESDCSVLSR ^{EH}	61–83	2+	Confirmed
		VIDF N *CTTSSVSSALANTK ^{EH,EL}	121–139	2+,3+	Confirmed
P05155	Plasma protease C1 inhibitor	VGQLQLSH N *LSLVILVPQNLK ^{EH}	344–364	2+	Confirmed

Table 1. Identification of glycosylation sites by LC-MS/MS analysis. (UH=unbound fraction of healthy serum; UL=unbound fraction of myocardial lesions serum; EH=bound fraction of healthy serum; EL=bound fraction of myocardial lesions serum).

	Protein	Sequence	Peptide	Charge	
P01857	Ig alpha-1 chain C region	LSLHRPALEDLLGSEAN*LTCTLTGLR ^{EH}	127-153	2+,3+	Potential
P01859	Ig gamma-2 chain C region	TKPREEQFN*STFR ^{EH} TPLTAN*ITK ^{EH}	168-180 200-208	2+ 2+	Confirmed Potential
P01860	Ig gamma-1 chain C region	EEQYN*STYR ^{UH,EL}	136-144	2+	Potential
P01861	Ig gamma-4 chain C region	EEQFN*STYR ^{UH,EH,EL}	173-181	2+	Potential
P01591	Immunoglobulin J chain	EN*ISDPTSPLR ^{EH,EL}	48-58	2+	Confirmed
P40189	Interleukin-6 receptor beta	QQYFKQN*CSQHESSPDISHFER ^{EL}	812-833	3+	Potential
P56199	Integrin alpha-1	SYFSSLN*LTIR ^{EL}	1096-1106	2+	Potential
P29622	Kallistatin	DFYVDEN*TTVR ^{EH}	232-242	2+	Confirmed
P01042	Kininogen-1	LNAENN*ATFYFK ^{UH,EL}	389-400	2+	Potential
P11279	Lysosome-associated membrane glycoprotein 1	DPAFKAAN*GSLR ^{EL}	314-325	2+	Confirmed
P01871	Ig mu chain C region	YKN*NSDISSTR ^{EH} GLTFQQN*ASSMCVPDQDPAIR ^{EL}	44-54 204-223	2+ 2+	Confirmed Confirmed
Q9HC10	Otoferlin	NEMLEIQVFN*YSKVFSNK ^{UH,UL,EL}	59-76	3+	Potential
Q5VU65	Nuclear pore membrane glycoprotein 210-like	EVVVN*ASSR ^{EH}	1551-1559	2+	Potential
Q9Y5E7	Protocadherin beta 2	ETRSEYN*ITITVTDFGTPR ^{EL}	414-432	3+	Potential
P36955	Pigment epithelium derived factor	VTQN*LTLIEESLTSEFIHDIDR ^{EH}	282-303	3+	Confirmed
P27169	Serum paraoxanase/arylesterase1	HAN*WTLTPLK ^{EH}	250-259	2+	Confirmed
P49908	Selenoprotein P	EGYSN*ISYIVVNHQGISSR ^{EH}	79-97	3+	Confirmed
Q9Y275	Tumor necrosis factor ligand superfamily member 13B	CIQNPETLPN*NSCYSAGIAK ^{EL}	232-252	3+	Confirmed
P40225	Thrombopoietin	IHELLN*GTRGLFPGPSRR ^{EL}	250-267	2+	Confirmed
P02787	Serotransferrin	QQQHLFGSN*VTDCSGNFCLFR ^{EH,EL}	622-642	2+	Confirmed
P07996	Thrombospondin 1	VVN*STTGPGHELR ^{EH,EL}	1065-1077	2+	Confirmed
P04004	Vitronectin	NN*ATVHEQVGGPSLTSDLQAQSK ^{EL}	85-107	3+	Confirmed
P25311	Zinc-alpha-2-glycoprotein	DIVEYYN*DSN*GSHVLQGR ^{EH,EL}	100-117	3+	Confirmed

by chromatographic and/or electrophoretic methods. In fact, the affinity step allowed us to selectively capture the defined subset of glycopeptides, greatly reducing the influence of the largely abundant non-glycosylated proteome complement.

Moreover, the specific enrichment of glycopeptides allowed both the identification of the individual glycoprotein and the direct assessment of the N-glycosylation site in the same analysis. Finally, the affinity procedure was instrumental for

the detection of the less abundant glycopeptides together with the most abundant ones, such as those deriving from albumins or immunoglobulins.

As shown in Table 1, all the identified peptides contained the conserved N-glycosylation motif Asn-X-Ser/Thr. Most of the retained peptides were effectively glycosylated by specific glycan structures, thus confirming the capability of ConA to bind N-glycosylated peptides carrying biantennary and high mannose-type structures with high selectivity. It should be

underlined that highly branched containing glycopeptides were not retained by the affinity medium according to the specificity of ConA. Moreover, some peptides could be identified in both the ConA bound and unbound fractions, possibly due to different branching structures of the glycan occurring at the same N-glycosylation site. Finally, few non-glycosylated peptides were detected in the fraction retained by ConA; they were identified as belonging to most abundant serum proteins.

Spontaneous deamidation of Asn residues, responsible for conversion of Asn into Asp, seems rather unlikely in generating the results presented, even because it preferentially occurs in Asn-Gly (Reference 18 and references therein) sequences, although it cannot be excluded completely. Most MS/MS spectra giving positive hits were derived from doubly charged precursor ions that predominantly yielded the y -ion fragment series. Although a number of triply charged precursor ions were fragmented generating good daughter spectra from which the modification site could be determined, many others gave poor spectra that prevented the unambiguous assignment of the glycosylation site. Therefore, the majority of the 68 identified sites were assigned by fragmentation of doubly charged precursors.

Almost similar results were obtained by the LC-MS/MS analyses of glycopeptides from both myocardial lesion carriers and healthy donors sera, i.e. the same glycosylated proteins were identified in the two samples. However, some individual glycopeptides were only found in one sample or in the other and, in a few cases, the same glycopeptide was recovered in the ConA bound fraction in one sample and in the unbound fraction in the other. This suggests that differences exist in the glycosylation profile of glycoproteins from healthy or pathological sera with these differences very likely confined to the branching structures of the glycan moieties.

Glycan profiling

To further highlight the differences in N-glycosylation possibly occurring between myocardial lesions and healthy donors sera, the overall glycan profiling of the two samples was investigated. Several papers have demonstrated that N-glycosylation changes in the serum glycome of cancer patients include changes in the levels of highly branched glycan structures, sialyl Lewis X epitopes and agalactosylated bi-antennary glycans. Moreover, several of these glycans have been linked to chronic inflammatory diseases, thus leading to the hypothesis of possible links between inflammation and cancer.¹⁹

Samples were treated as described above and following the ConA affinity selection step, the entire oligosaccharides were released from glycopeptides occurring in both the ConA bound and unbound fractions by incubation with PNGase F to isolate high mannose, hybrid and biantennary structures as well as higher branched glycans.

Oligosaccharides released from enzymatic hydrolysis were desalted by in batch ion exchange chromatography on a Dowex mixed bed resin that retained most of the ions occur-

ring in the sample. Glycans were then directly analyzed by MALDI-MS on a reflectron instrument, both in positive and negative mode. Since significant desialylation might occur during analysis of the samples, MALDI mass spectra were also performed using THAP as matrix that was reported to avoid most of the problems associated with loss of sialic acids by fragmentation.²⁰ Each peak detected in the MALDI spectra was manually interpreted and tentatively attributed to a glycan structure according to the N-glycan biosynthetic pathway and literature data. Manual interpretations were confirmed by using the ExPASy GlycoMod software, an online software for prediction the possible oligosaccharide structures in glycoproteins from their experimentally determined masses. The ultimate definition of the major glycan structures detected in the different mixtures, including the sequence and branching degree of the antennae, was achieved by MS/MS experiments using the MALDI PSD technique. MALDI MS/MS experiments allowed us to confirm glycan structures and to discriminate between isobaric components, as indicated in the Tables 2 and 3.

As a whole, these MALDI analyses constituted the oligosaccharide profiling of the samples allowing us to describe the microheterogeneity of glycosylation within the different serum samples in details and to identify specific changes in glycan structures occurring in the pathological sera.

Figure 4 shows the positive MALDI spectra of glycans deriving from the (a) healthy and (b) pathological glycopeptides retained during the CoA affinity step and incubated with PNGase F. The oligosaccharide composition and structures corresponding to the recorded mass signals are summarized in Table 2. More than 50 structures were detected in the MALDI profilings, as indicated in Table 2, essentially belonging to high mannose and biantennary type, thus indicating a high degree of glycan heterogeneity. These data indicated that the two ConA bound fractions essentially consisted of peptides carrying biantennary and hybrid, as well as high mannose structures. This finding is in perfect agreement with the reported affinity of ConA that generally binds complex biantennary, hybrid type and high-mannose-type N-linked glycans.^{21,22}

The MALDI profiling revealed the occurrence of only three mass signals, at m/z 2028.7, m/z 2174.8 and m/z 2320.8, respectively, that were tentatively attributed to triantennary structures, HexNAc₅Hex₆, FucHexNAc₅Hex₆ and HexNAc₅Hex₆Sia. PSD analysis performed on the precursor ion at m/z 2174.8 confirmed the structure of the triantennary glycan and localized the fucose residue on the 3-antenna. In the fucosylated triantennary structure, the position of the fucose residue plays a very important biological role. As an example, alpha-1-acid glycoprotein, one of the identified glycoproteins in this study, is a non specific marker, whose temporary increase in the serum level is generally associated with many acute phase syndromes (infections, inflammation, cancer etc...)²³ Previous data reported that the fucose residue on the triantennary glycans in alpha-1-acid glycoprotein was generally found on the 3-antenna and not on the terminal core GlcNAc.^{24,25}

Table 2. Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H= healthy serum; L= myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

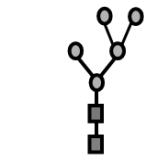
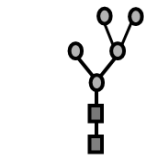
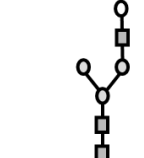
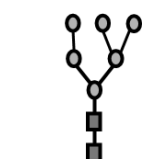
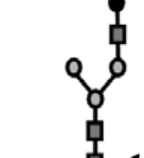
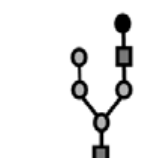
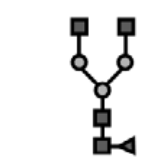
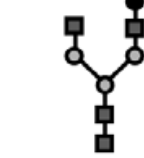
MALDI (m/z [$M + Na$] ⁺)			Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
1	1257.4	1257.4 ^H 1257.3 ^L	HexNAc ₂ Hex ₅	High mannose	[41]		X
2	1298.5	1298.4 ^H 1298.3 ^L	HexNAc ₃ Hex ₄	Complex, monoantennary	[41]		X
3	1419.5	1419.5 ^H 1419.3 ^L	HexNAc ₂ Hex ₆	High mannose	[41]		X
4	1444.5	1444.5 ^H 1444.4 ^L	FucHexNAc ₃ Hex ₄	Complex, fucosylated			X
5	1460.5	1460.4 ^H 1460.4 ^L	HexNAc ₃ Hex ₅	Hybrid	[41]		
6	1485.5	1485.4 ^H 1485.4 ^L	FucHexNAc ₄ Hex ₃	Complex, biantennary, fucosylated	[41]		X
7	1501.5	1501.5 ^H 1501.3 ^L	HexNAc ₄ Hex ₄	Complex, biantennary	[41]		X

Table 2 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H=healthy serum; L=myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

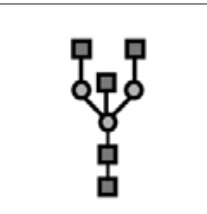
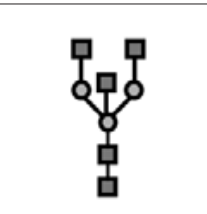
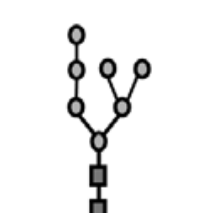
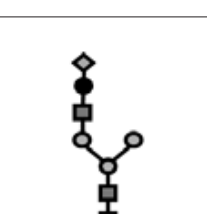
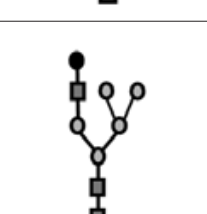
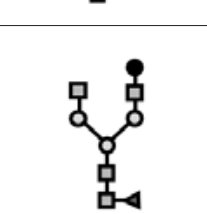
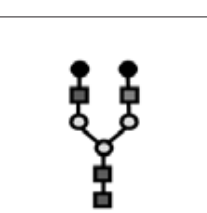
MALDI (m/z [M+Na] ⁺)			Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
8	1542.6	1542.4 ^H 1542.4 ^L	HexNAc ₅ Hex ₃	Complex, bisecting	[41]		X
9	1581.5	1581.5 ^H 1581.4 ^L	HexNAc ₂ Hex ₇	High mannose	[41]		
10	1611.5	1611.4 ^H	HexNAc ₃ Hex ₄ SiaNa	Complex, monoantennary, sialylated	[42]		
11	1622.5	1622.5 ^H 1622.4 ^L	HexNAc ₃ Hex ₆	Hybrid	[41]		
12	1647.6	1647.5 ^H 1647.3 ^L	FucHexNAc ₄ Hex ₄	Complex, biantennary, fucosylated	[41]		X
13	1663.6	1663.5 ^H 1663.3 ^L	HexNAc ₄ Hex ₅	High mannose	[41]		
14	1685.6	1685.8 ^H 1685.4 ^L	HexNAc ₄ Hex ₅ Na	Complex, biantennary		Calculated	

Table 2 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H = healthy serum; L = myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

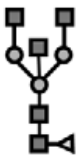
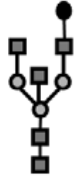
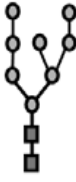
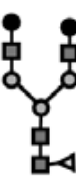

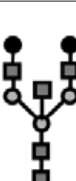
	MALDI (m/z [$M + Na$] $^+$)		Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
15	1688.6	1688.5 ^H 1688.3 ^L	FucHexNAC ₅ Hex ₃	Complex, fucosylated, bisecting	[41]		
16	1704.6	1705.5 ^H	HexNAC ₅ Hex ₄	Complex, bisecting	[41]		
17	1743.6	1743.4 ^H	HexNAC ₂ Hex ₈	High mannose	[41]		
18	1809.6	1809.5 ^H 1809.3 ^L	FucHexNAC ₄ Hex ₅	Complex, biantennary, fucosylated	[41]		X
19	1814.6	1815.4 ^H 1816.3 ^L	HexNAC ₄ Hex ₄ SiaNa	Complex, biantennary, sialylated		Calculated	
20	1825.6	1825.5 ^H 1825.3 ^L	HexNAC ₄ Hex ₆	Hybrid		Calculated	
21	1832.6	1831.3 ^L	FucHexNAC ₄ Hex ₅ Na	Complex, fucosylated, biantennary		Calculated	
22	1850.7	1850.5 ^H 1850.3 ^L	FucHexNAC ₅ Hex ₄	Complex, fucosylated, bisecting	[41]		
23	1866.7	1866.5 ^H 1880.5 ^L	HexNAC ₅ Hex ₅	Complex, bisecting	[41]		

Table 2 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H=healthy serum; L=myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

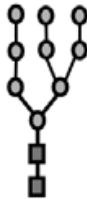

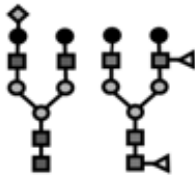
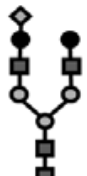
MALDI (m/z [$M+Na$] $^+$)			Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
24	1881.6	1880.8 ^H 1880.5 ^L	Fuc ₂ HexNAC ₃ Hex ₄ Sia	Complex, difucosylated, sialylated, monoantennary		Calculated	
25	1905.6	1906.5 ^H 1905.4 ^L	HexNAC ₂ Hex ₉	High mannose	[41]		
26	1910.7	1910.5 ^H 1910.3 ^L	HexNAC ₅ Hex ₅ Na ₂	Complex, triantennary/bisecting Complex, biantennary, fucosylated, sulfated		Calculated	
27	1935.6	1936.5 ^H 1936.3 ^L	HexNAC ₃ Hex ₆ SiaNa	Hybrid, sialylated	[42]		
28	1955.7	1954.5 ^H 1954.3 ^L	HexNAC ₄ Hex ₅ Sia Fuc ₂ HexNAC ₄ Hex ₅	Complex, biantennary, sialylated Complex, biantennary, difucosylated	[41]		X
29	1960.7	1958.5 ^H 1957.3 ^L	FucHexNAC ₄ Hex ₄ SiaNa	Complex, fucosylated, sialylated, biantennary		Calculated	
30	1971.7	1970.3 ^L	FucHexNAC ₄ Hex ₆	Hybrid, fucosylated		Calculated	
31	1976.7	1975.4 ^H 1976.3 ^L	HexNAC ₄ Hex ₅ SiaNa	Complex, biantennary, sialylated	[42]		
32	1995.7 1996.7	1996.5 ^H 1997.2 ^L	HexNAC ₅ Hex ₄ Sia Fuc ₂ HexNAC ₅ Hex ₄	Sialylated, triantennary/bisecting Difucosylated, triantennary/bisecting		Calculated	

Table 2 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H = healthy serum; L = myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

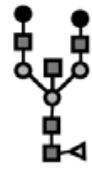
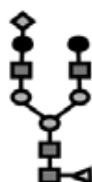
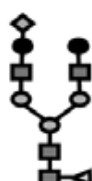
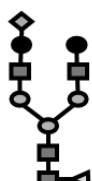
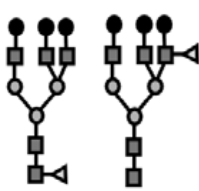
MALDI (m/z [$M + Na$] $^+$)			Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
33	2012.7	2012.5 ^H 2012.3 ^L	FucHexNAC ₅ Hex ₅	Complex, fucosylated, bisecting	[41]		
34	1995.7 1996.7	1996.5 ^H 1997.2 ^L	HexNAC ₅ Hex ₆	Complex, triantennary	[41]	Calculated	
35	2100.7	2100.5 ^H 2101.4 ^L	FucHexNAC ₄ Hex ₅ Sia	Complex, fucosylated, sialylated, biantennary			X
36	2123.7	2122.5 ^H 2122.3 ^L	FucHexNAC ₄ Hex ₅ SiaNa	Complex, fucosylated, sialylated, biantennary		Calculated	
37	2 144.7	2145.4 ^H 2144.2 ^L	FucHexNAC ₄ Hex ₅ SiaNa ₂	Complex, fucosylated, sialylated, biantennary			X
38	2157.1	2158.3 ^L	HexNAC ₅ Hex ₅ Sia	Complex, sialylated, triantennary/bisecting		Calculated	
39	2166.7	2166.3 ^H	FucHexNAC ₄ Hex ₅ SiaNa ₃	Complex, fucosylated, sialylated, biantennary			X
40	2174.8	2175.5 ^H 2175.2 ^L	FucHexNAC ₅ Hex ₆	Complex, fucosylated, triantennary	[41]		
41	2197.7	2193.6 ^H 2194.5 ^L	FucHexNAC ₅ Hex ₆ Na	Complex, fucosylated, triantennary		Calculated	

Table 2 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides retained on the ConA affinity step (H=healthy serum; L= myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.



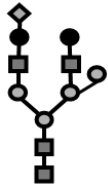

MALDI [m/z [$M + Na$] $^+$]			Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
42	2245.8	2246.5 ^H 2245.3 ^L	HexNAc ₄ Hex ₅ Sia ₂	Complex, disialylated, biantennary	[42]		
43	2262.8	2262.2 ^L	FucHexNAc ₄ Hex ₆ Sia	Hybrid, sialylated, fucosylated		Calculated	
44	2268.8	2267.5 ^H 2267.4 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, disialylated, biantennary	[42]		X
45	2284.8	2283.4 ^H 2283.2 ^L	FucHexNAc ₄ Hex ₆ SiaNa	Hybrid, fucosylated, sialylated			X
46	2290.8	2289.5 ^H 2289.3 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na ₂	Complex, disialylated, biantennary	[42]		
47	2306.8	2305.4 ^H	FucHexNAc ₄ Hex ₆ SiaNa ₂	Hybrid, fucosylated, sialylated		Calculated	
48	2325.8	2326.5 ^H	FucHexNAc ₅ Hex ₅ SiaNa	Complex, fucosylated, sialylated, triantennary/bisecting		Calculated	
49	2434.8	2435.5 ^H	FucHexNAc ₄ Hex ₅ Sia ₂ Na ₂	Complex, fucosylated, disialylated, biantennary		Calculated	
50	2627.9	2625.4 ^H 2625.5 ^L	FucHexNAc ₅ Hex ₇ Sia	Hybrid, sialylated		Calculated	
51	2649.8	2647.0 ^L	FucHexNAc ₅ Hex ₇ SiaNa	Hybrid, fucosylated, sialylated		Calculated	

Table 3. Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H=healthy serum; L=myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

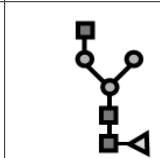
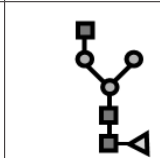
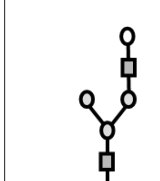
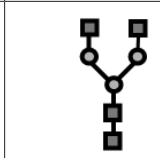
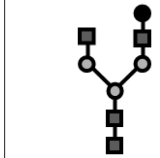
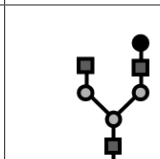
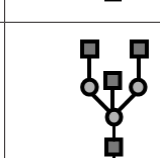
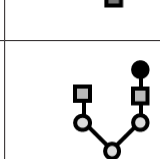
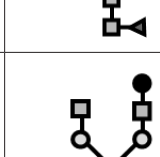
	MALDI (m/z [M+Na] ⁺)		Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW	Found					
1	1282.4	1282.1 ^H	FucHexNAc ₃ Hex ₃	Complex, fucosylated, monoantennary	[41]		
2	1298.5	1298.1 ^H	HexNAc ₃ Hex ₄	Complex, monoantennary	[41]		X
3	1339.5	1339.1 ^H	HexNAc ₄ Hex ₃	Complex, biantennary	[41]		
4	1485.5	1485.1 ^H 1485.0 ^L	FucHexNAc ₄ Hex ₃	Complex, fucosylated, biantennary	[41]		X
5	1501.5	1501.0 ^H 1500.9 ^L	HexNAc ₄ Hex ₄	Complex, biantennary	[41]		X
6	1542.6	1542.0 ^H	HexNAc ₅ Hex ₃	Complex, bisecting	[41]		X
7	1647.6	1647.0 ^H 1646.9 ^L	FucHexNAc ₄ Hex ₄	Complex, fucosylated, biantennary	[41]		X
8	1663.6	1663.0 ^H 1662.9 ^L	HexNAc ₄ Hex ₅	Complex, biantennary	[41]		X

Table 3 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H=healthy serum; L= myocardial lesions serum). \triangle , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

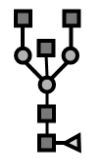
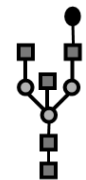
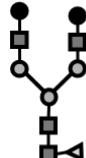
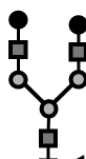
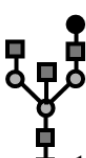
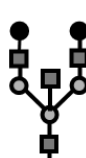
MALDI (m/z [$M+Na$] ⁺)		Monosaccharides composition	Structure	Ref.		PSD	
	Theoretical MW						Found
9	1685.6	1684.6 ^H 1685.1 ^L	HexNAc ₄ Hex ₅ Na	Complex, biantennary		Calculated	
10	1688.6	1688.0 ^H	FucHexNAc ₅ Hex ₃	Complex, fucosylated, bisecting	[41]		
11	1704.6	1705.0 ^H 1703.9 ^L	HexNAc ₅ Hex ₄	Complex, bisecting	[41]		
12	1809.6	1809.0 ^H 1809.9 ^L	FucHexNAc ₄ Hex ₅	Complex, biantennary, fucosylated	[41]		X
13	1831.6	1830.1 ^H 1830.8 ^L	FucHexNAc ₄ Hex ₅ Na	Complex, fucosylated, biantennary			X
14	1850.7	1850.0 ^H 1849.9 ^L	FucHexNAc ₅ Hex ₄	Complex, fucosylated, bisecting	[41]		
15	1866.7	1867.2 ^H 1865.8 ^L	HexNAc ₅ Hex ₅	Complex, bisecting	[41]		
16	1889.7	1889.2 ^H 1886.4 ^L	HexNAc ₅ Hex ₅ Na	Complex, triantennary/bisecting		Calculated	

Table 3 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H=healthy serum; L=myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

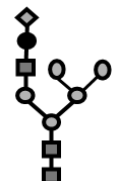
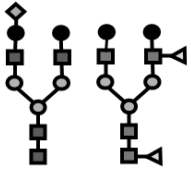
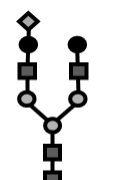
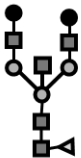
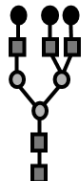
MALDI (m/z [M+Na] ⁺)		Monosaccharides composition	Structure	Ref.		PSD	
Theoretical MW	Found						
17	1913.7	1912.0 ^H 1911.8 ^L	HexNAc ₃ Hex ₆ Sia	Hybrid, sialylated	[42]		X
18	1955.7	1954.0 ^H	HexNAc ₄ Hex ₅ Sia	Complex, sialylated, biantennary	[41]		X
		1953.8 ^L	Fuc ₂ HexNAc ₄ Hex ₅	Complex, biantennary, difucosylated			
19	1977.7	1975.9 ^H 1975.8 ^L	HexNAc ₄ Hex ₅ SiaNa	Complex, sialylated, biantennary	[42]		X
20	1995.7	1996.0 ^H	HexNAc ₅ Hex ₄ Sia	Complex, sialylated, triantennary/bisecting		Calculated	
		1996.7	Fuc ₂ HexNAc ₅ Hex ₄	Complex, difucosylated, triantennary/bisecting			
21	2012.7	2011.9 ^H 2011.8 ^L	FucHexNAc ₅ Hex ₅	Complex, fucosylated, bisecting	[41]		
22	2028.7	2028.0 ^H 2028.7 ^L	HexNAc ₅ Hex ₆	Complex, triantennary	[41]		X
23	2050.7	2049.9 ^H 2047.8 ^L	HexNAc ₅ Hex ₆ Na	Complex, triantennary		Calculated	
24	2081.6	2081.3 ^H 2081.7 ^L	FucHexNAc ₃ Hex ₆ Sia	Hybrid, fucosylated, sialylated		Calculated	

Table 3 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H=healthy serum; L=myocardial lesions serum). \triangle , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

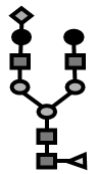
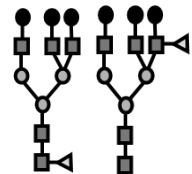
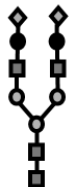
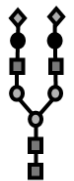
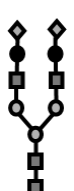
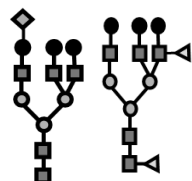
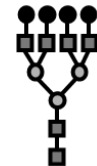
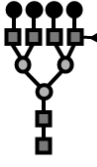
MALDI (m/z [$M+Na$] ⁺)		Monosaccharides composition	Structure	Ref.		PSD
	Theoretical MW					
25	2123.7	2121.9 ^H	FucHexNAC ₄ Hex ₅ SiaNa			X
26	2174.8	2174.7 ^L	FucHexNAC ₅ Hex ₆	[41]		
27	2179.8	2179.6 ^L	HexNAC ₅ Hex ₅ SiaNa		Calculated	
28	2245.8	2244.9 ^H	HexNAC ₄ Hex ₅ Sia ₂	[42]		X
29	2267.8	2266.9 ^H 2267.5 ^L	HexNAC ₄ Hex ₅ Sia ₂ Na	[42]		X
30	2284.8	2282.8 ^H 2283.7 ^L	FucHexNAC ₄ Hex ₆ SiaNa		Calculated	
31	2290.8	2288.9 ^H	HexNAC ₄ Hex ₅ Sia ₂ Na ₂	[42]		
32	2320.8	2318.8 ^H 2319.7 ^L	HexNAC ₅ Hex ₆ Sia Fuc ₂ HexNAC ₅ Hex ₆	[41]		X

Table 3 (continued). Composition and structures of the glycans detected by MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H= healthy serum; L= myocardial lesions serum). Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid. Calculated: attributed on the basis of GlycoMod software only.

MALDI (m/z [$M+Na$] $^+$)		Monosaccharides composition	Structure	Ref.		PSD	
Theoretical MW	Found						
33	2325.8	2324.9 ^H 2324.7 ^L	FucHexNAC ₅ Hex ₅ SiaNa	Complex, fucosylated, sialylated, triantennary/bisecting		Calculated	
34	2342.8	2341.8 ^H 2340.7 ^L	HexNAC ₅ Hex ₆ SiaNa	Complex, sialylated, triantennary		Calculated	
35	2360.9	2359.9 ^H 2359.8 ^L	HexNAC ₆ Hex ₅ Sia	Complex, sialylated, tetraantennary		Calculated	
36	2393.7	2392.9 ^H 2392.6 ^L	HexNAC ₆ Hex ₇	Complex, tetraantennary	[41]		X
37	2413.8	2412.3 ^L	FucHexNAC ₄ Hex ₅ Sia ₂ Na	Complex, fucosylated, disialylated, biantennary		Calculated	
38	2466.8	2465.6 ^L	FucHexNAC ₅ Hex ₆ Sia	Complex, fucosylated, sialylated, triantennary		Calculated	
39	2488.8	2487.7 ^L	FucHexNAC ₅ Hex ₆ SiaNa	Complex, fucosylated, sialylated, triantennary		Calculated	
40	2539.9	2539.4 ^L	FucHexNAC ₆ Hex ₇	Complex, fucosylated, tetraantennary	[41]		
41	2611.9	2610.7 ^H	HexNAC ₅ Hex ₆ Sia ₂	Complex, disialylated, triantennary		Calculated	
42	2633.9	2632.8 ^H 2632.6 ^L	HexNAC ₅ Hex ₆ Sia ₂ Na	Complex, disialylated, triantennary		Calculated	
43	2651.9	2649.7 ^H 2648.6 ^L	HexNAC ₆ Hex ₅ Sia ₂	Complex, disialylated, tetraantennary		Calculated	
44	2654.9	2653.8 ^H	HexNAC ₅ Hex ₆ Sia ₂ Na ₂	Complex, disialylated, triantennary		Calculated	

Negative MALDI spectra performed on the same ConA unbound fractions essentially confirmed the above reported data showing the presence of a few sialylated variants, corresponding to biantennary glycans.

Higher branched glycopeptides were not retained onto the lectin originating the unbound fraction of the affinity purification. Thus, ConA affinity chromatography led to the relative concentration of tri- and tetraantennary oligosac-

charides in this fraction. Without this step, these usually less abundant glycans could have been escaped detection. PNGase F was directly added to the unbound fraction, glycans were separated from peptides by HPLC and directly analysed by MALDI-MS following the procedures described above.

Figure 5 shows the positive MALDI-MS analyses of glycans deriving from the unbound fractions of the (a) healthy and (b)

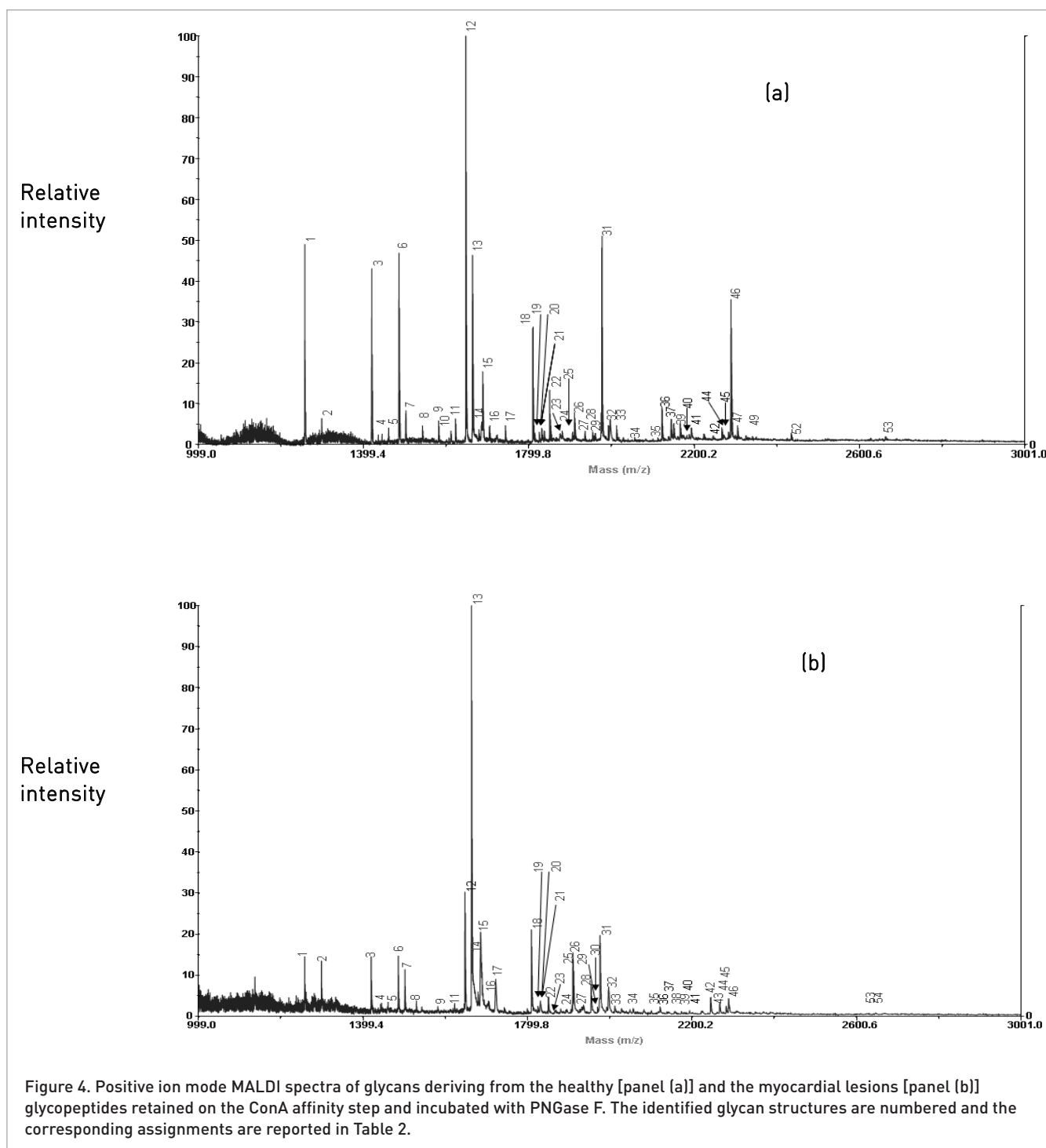
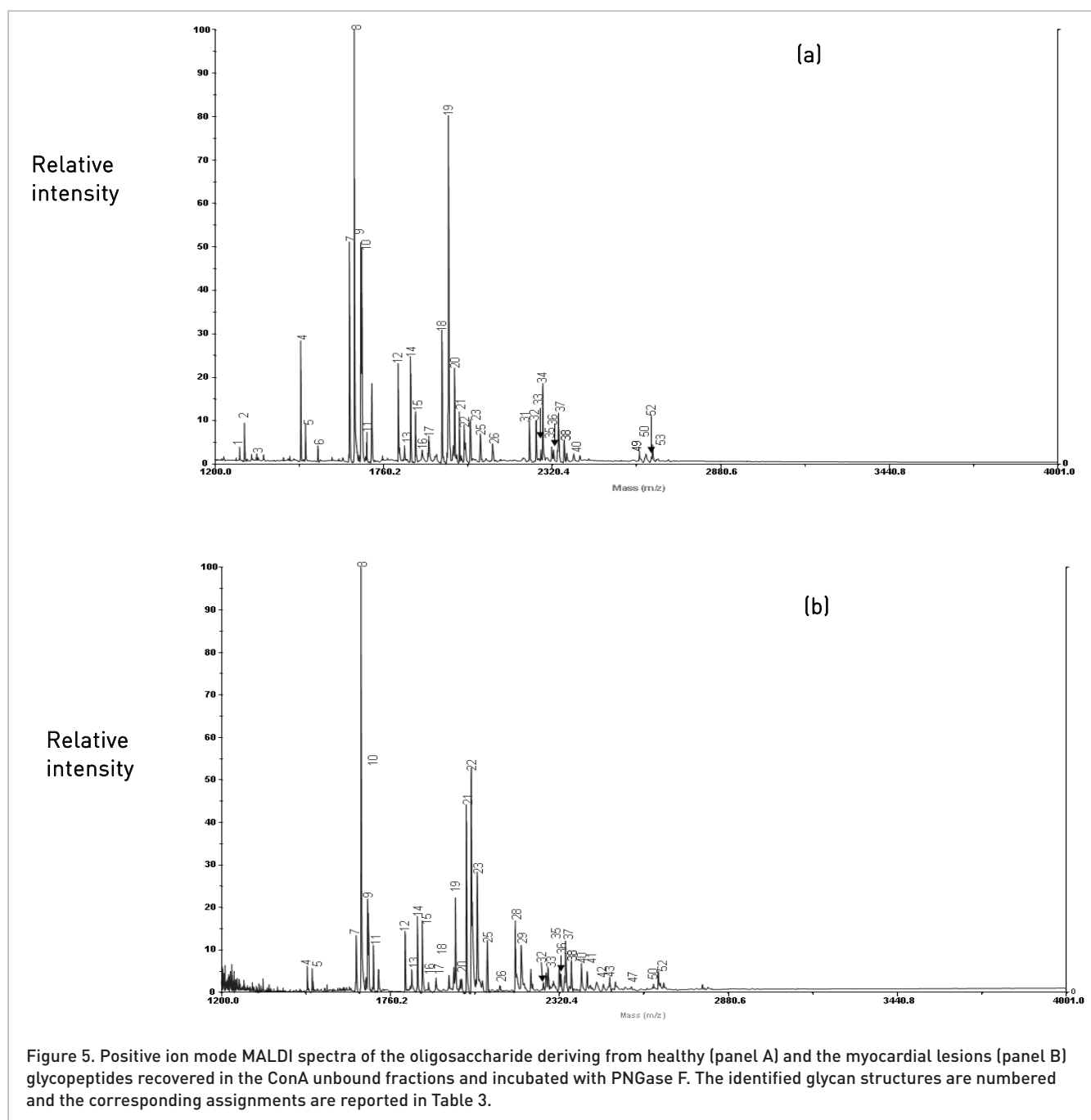


Figure 4. Positive ion mode MALDI spectra of glycans deriving from the healthy [panel (a)] and the myocardial lesions [panel (b)] glycopeptides retained on the ConA affinity step and incubated with PNGase F. The identified glycan structures are numbered and the corresponding assignments are reported in Table 2.

pathological serum samples. Mass signals were tentatively assigned as described before and, when necessary, confirmed by PSD fragmentation experiments. Figures 6(a)–(c) shows the PSD analysis of the non sialylated, monosialylated and disialylated biantennary structure respectively. The PSD spectra were dominated by Y and B ions containing the non reducing or the reducing end of the glycan moiety as well as a large

number of internal fragment ions formed by multiple glycosidic cleavages. Sequence and branching information could then be inferred from Y and B ions and were confirmed by internal fragment ions

As an example, Figure 6(b) shows the PSD spectrum obtained by selecting the precursor ion at m/z 2100.9 corresponding to the putative monosialylated biantennary struc-



ture. The complete series of Y ions from both antennae ($Y\alpha$ and $Y\beta$ ions) could be detected in the spectrum together with a number of B ions. These data led to the complete assignment of the monosialylated fucosylated biantennary structure that was further confirmed by a series of internal fragments. Similar results were obtained for the remaining oligosaccharides.

Negative MALDI analyses were also performed on the same ConA unbound fractions to highlight the presence of sialylated structures. All the data are summarised in Tables 3 and 4.

Compared to the glycans of the bound fraction, these oligosaccharides showed the presence of higher branched, sialylated and fucosylated structures. Table 3 shows the presence of essentially sialylated triantennary glycans with the occurrence of some fucosylated and sialylated tetraantennary structures and few biantennary oligosaccharides. When peaks could not be unambiguously interpreted, the ultimate definition of the correct glycan structure was achieved by MS/MS experiments using the MALDI PSD technique. As an example, the signal at m/z 2392.8 in the MALDI spectrum was tentatively assigned to two different isobaric structures. Following PSD

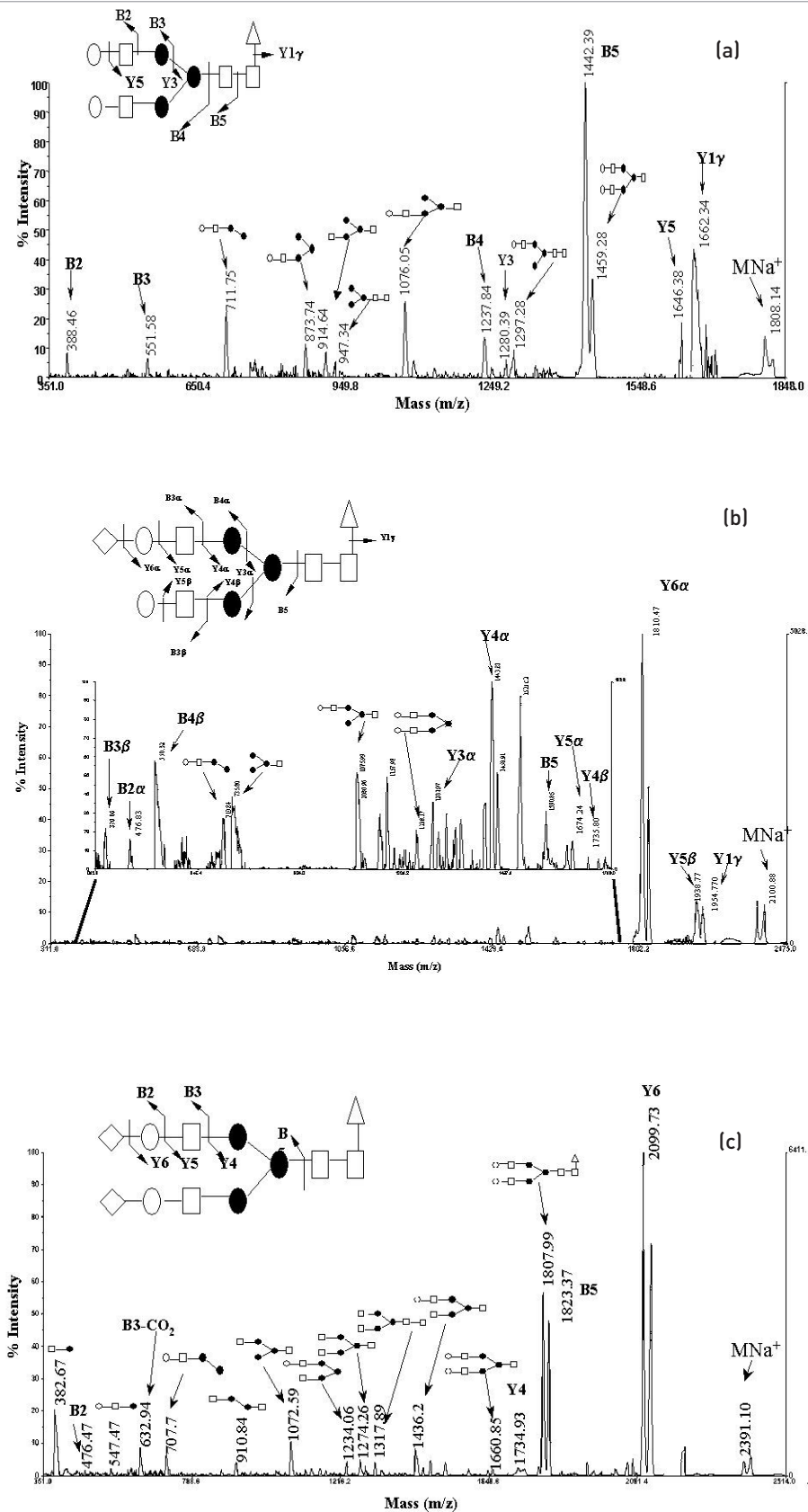


Figure 6. Panel (a) MALDI-PSD analyses of non sialylated, panel (b) monosialylated and panel (c) disialylated biantennary structures. Δ , Fucose; \square , N-acetylglucosamine; \bullet , Hexose; \diamond , N-acetylneuraminic acid.

Table 4. Composition and structures of the glycans detected by negative ion mode MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H = healthy serum; L = myocardial lesions serum).

	MALDI (m/z [M+Na] ⁺)		Monosaccharide composition	Structure
	Theoretical MW	Found		
1	1930.7	1929.6 ^H 1929.9 ^L	HexNAc ₄ Hex ₅ Sia	Complex, sialylated, biantennary
2	1953.7	1952.4 ^H 1952.2 ^L	HexNAc ₄ Hex ₅ SiaNa	Complex, sialylated, biantennary
3	2076.8	2074.5 ^H 2075.8 ^L	FucHexNAc ₄ Hex ₅ Sia	Complex, fucoylated, sialylated, biantennary
4	2133.7	2133.8 ^L	HexNAc ₅ Hex ₅ Sia	Complex, sialylated, triantennary/bisecting
5	2179.7	2178.6 ^H	HexNAc ₅ Hex ₅ SiaNa	Complex, sialylated, triantennary/bisecting
6	2220.8	2220.4 ^H 2221.7 ^L	HexNAc ₄ Hex ₅ Sia ₂	Complex, disialylated, biantennary
	2222.8		Fuc ₂ HexNAc ₄ Hex ₅ Sia	Complex, difucosylated, sialylated, biantennary
7	2243.8	2242.4 ^H 2242.7 ^L	HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, disialylated, biantennary
	2245.8		Fuc ₂ HexNAc ₄ Hex ₅ SiaNa	Complex, difucosylated, sialylated, biantennary
8	2279.8	2279.4 ^H 2278.8 ^L	FucHexNAc ₅ Hex ₅ Sia	Complex, fucosylated, sialylated, triantennary/bisecting
9	2295.7	2294.4 ^H 2294.8 ^L	HexNAc ₅ Hex ₆ Sia	Complex, sialylated, triantennary
10	2318.7	2316.2 ^H 2316.3 ^L	HexNAc ₅ Hex ₆ SiaNa	Complex, sialylated, triantennary
11	2348.8	2349.8 ^H 2348.7 ^L	FucHexNAc ₃ Hex ₆ Sia ₂ Na	Hybrid, trifucosylated, disialylated
12	2389.9	2389.3	FucHexNAc ₄ Hex ₅ Sia ₂ Na	Complex, fucosylated, disialylated, biantennary
13	2441.8	2441.7 ^L	FucHexNAc ₅ Hex ₆ Sia	Complex, fucosylated, sialylated
14	2464.8	2461.8 ^L	FucHexNAc ₅ Hex ₆ SiaNa	Complex, fucosylated, sialylated
15	2536.0	2537.1 ^H	Fuc ₂ HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, difucosylated, disialylated, biantennary
16	2569.9	2570.1 ^H 2570.6 ^L	FucHexNAc ₅ Hex ₅ Sia ₂	Complex, fucosylated, disialylated, triantennary/bisecting
17	2586.8	2586.3 ^H 2585.7 ^L	HexNAc ₅ Hex ₆ Sia ₂	Complex, disialylated, triantennary
18	2608.8	2607.2 ^H 2608.6 ^L	HexNAc ₅ Hex ₆ Sia ₂ Na	Complex, disialylated, triantennary
19	2615.9	2615.0 ^H 2615.2 ^L	FucHexNAc ₅ Hex ₅ Sia ₂ Na ₂	Complex, fucosylated, disialylated, triantennary/bisecting
20	2631.8	2630.2 ^H 2630.3 ^L	HexNAc ₅ Hex ₆ Sia ₂ Na ₂	Complex, disialylated, triantennary
21	2660.9	2660.1 ^H 2660.6 ^L	HexNAc ₆ Hex ₇ Sia	Complex, sialylated, tetraantennary
22	2682.1	2681.9 ^H 2680.6 ^L	Fuc ₃ HexNAc ₄ Hex ₅ Sia ₂ Na	Complex, trifucosylated, disialylated, biantennary
23	2731.8	2731.6 ^L	FucHexNAc ₅ Hex ₆ Sia ₂	Complex, fucosylated, disialylated
24	2754.8	2754.5 ^L	FucHexNAc ₅ Hex ₆ Sia ₂ Na	FucHexNAc ₅ Hex ₆ Sia

Table 4 (continued). Composition and structures of the glycans detected by negative ion mode MALDI MS analysis of the oligosaccharide mixtures derived from glycopeptides recovered in the ConA unbound fraction (H=healthy serum; L=myocardial lesions serum).

	MALDI [<i>m/z</i> [M+Na] ⁺]		Monosaccharide composition	Structure
	Theoretical MW	Found		
25	2806.9	2806.5 ^L	FucHexNAc ₆ Hex ₇ Sia	Complex, fucosylated, sialylated, tetraantennary
26	2875.9	2876.0 ^H 2876.4 ^L	HexNAc ₅ Hex ₆ Sia ₃	Complex, trisialylated, triantennary
27	2898.9	2899.0 ^H 2899.5 ^L	HexNAc ₅ Hex ₆ Sia ₃ Na	Complex, trisialylated, triantennary
28	2921.9	2921.0 ^H 2920.4 ^L	HexNAc ₅ Hex ₆ Sia ₃ Na ₂	Complex, trisialylated, triantennary
29	2951.0	2951.9 ^H 2951.5 ^L	HexNAc ₆ Hex ₇ Sia ₂	Complex, disialylated, tetraantennary
30	2974.0	2973.9 ^H	HexNAc ₆ Hex ₇ Sia ₂ Na	Complex, disialylated, tetraantennary
31	3021.9	3022.7 ^H	FucHexNAc ₅ Hex ₆ Sia ₃	Complex, fucosylated, trisialylated, triantennary

fragmentation, the correct biantennary sialylated and fucosylated glycoform could be inferred.

As shown in Tables 2–4, the microheterogeneity of glycans detected in the healthy sera resulted to be, in general, very similar to the glycan profiling observed for the myocardial lesions serum. However, a number of differences could be detected in the high mass region of the spectra and particularly in the unbound fraction of the myocarditis sample with respect to the healthy glycans.

As serum is a very complex system, this analysis is far from being non-exhaustive. A complete profiling of the glycosylation in the different sera would have needed a detailed quantitative evaluation of all the oligosaccharide

structures detected, which was outside the scope of this paper. However, preliminary semi quantitative data could be obtained on the most abundant glycan structures observed in the pathological sera in comparison with the healthy sera by measuring the ratio of peak areas. As summarized in Tables 5 and 6, most glycoforms displayed quantitative differences when pathological samples were compared to the healthy ones.

The analysis of the ConA eluted glycan fractions (Table 5) showed a three to five times decrease in the relative abundance of most glycan structures in the pathological sample compared to the healthy oligosaccharides. On the contrary, the same analysis carried out on the unbound

Table 5. Semiquantitative analysis of ConA eluted fraction oligosaccharides in pathological samples in comparison to the reference one LM: myocardial lesions.

Calculated <i>m/z</i> [M+Na] ⁺	Structure	LM/healthy serum ratio
1257.4	High mannose	0.172
1419.5	High mannose	0.211
1485.5	Complex, biantennary, fucosylated	0.200
1647.6	Complex, biantennary, fucosylated	0.212
1663.6	Complex, biantennary	0.818
1809.6	Complex, biantennary, fucosylated	0.337
1976.7	Complex, biantennary, sialylated	0.189
2123.7	Complex, fucosylated, sialylated, biantennary	0.0991
2245.8	Complex, disialylated, biantennary	1.16
2290.8	Complex, disialylated, biantennary	0.0604

Table 6. Relative abundance of some selected peaks in pathological samples in comparison to the reference one in ConA unbound fraction. LM: myocardial lesions.

Calculated m/z [M + Na] ⁺	Structure	LM/healthy serum ratio
1688.0	Complex, fucosylated bisecting	0.656
1850.7	Complex, fucosylated bisecting	0.957
1913.7	Hybrid, sialylated	0.842
2012.7	Complex, fucosylated bisecting	3.23
2028.7	Complex, triantennary	4.26
2174.8	Complex, fucosylated, triantennary	Not detected in the healthy serum
2342.8	Complex, sialylated, triantennary	1.05
2466.8	Complex, fucosylated, sialylated, triantennary	Not detected in the healthy serum
2654.9	Complex, disialylated triantennary	Not detected in the myocardial lesions serum

ConA fractions (Table 6) revealed a higher level of heterogeneity within the relative abundance of the glycan structures. In fact, most species remained unchanged in the two samples while the relative abundance of two structures increased three to four times in the pathological glycan profiling. Moreover, the pathological sample showed the occurrence of specific structures completely absent in the normal sample.

As a whole, the myocardial glycan profiling showed the occurrence of peculiar structures, the persistence of incomplete and truncated structures, the accumulation of precursors and the appearance of novel structures, thus leading to a higher degree of heterogeneity, with respect to the healthy one. In particular, the myocardial lesions serum revealed the presence of hyper-fucosylated oligosaccharides. It has been demonstrated that the N-linked glycans attached to proteins generally associated with inflammation, such as haptoglobin, alpha1-acid glycoprotein and alpha1-antichymotrypsin, have increased SLe^x epitopes, during chronic inflammation [Reference 18 and references therein]. Further analyses are currently in progress to deliver assignments of the glycan structure on individual glycosylation sites of the glycoproteins identified.

Conclusions

This report describes an integrated single-step proteomic approach for the assessment of the N-glycosylation profilings occurring in human sera. More interestingly, the glycosylation pattern of glycoproteins occurring in the serum of myocarditis patients was investigated for the first time. This procedure combines a simple affinity chromatography step with sophisticated LC-MS/MS analyses and led to the identification of several N-glycosylation sites in serum proteins,

some of which had never been detected before. The advantages of this protocol was clearly demonstrated through differentiation of glycan profiling of sera from healthy and pathological donors.

A comparative study of the N-glycosylation pattern of the entire human serum glycoproteome, here applied to the study of myocarditis, could be considered an excellent probe, although preliminary, for other specific pathologies. Finally, this glycobiological approach represents a new opportunity for therapeutics and diagnostics, for the screening of oligosaccharide biomarkers in pathological status, but also to biotechnologies, as bio-analytical procedure aimed at the characterization of glycosylation pattern of different complex samples, in order to manufacture correctly glycosylated molecules.

Acknowledgments

This work was supported by grants from the INBB, Ministero dell'Università e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale 2006, 2007) and by the FIRB Project Rete Nazionale per lo studio della Proteomica Umana (Italian Human ProteomeNet) RBRN07BMCT. Support from the National Center of Excellence in Molecular Medicine (MIUR-Rome) and from the Regional Center of Competence (CRdC ATIBB, Regione Campania-Naples) is gratefully acknowledged.

References

1. R.D. Marshall, "The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins", *Biochem. Soc. Symp.* **40**, 17–26 (1974).

2. R. Qiu and F.E. Regnier, "Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum", *Anal. Chem.* **15**, 7725–7731 (2005).
3. A. Dell and H.R. Morris, "Glycoprotein structure determination by mass spectrometry", *Science* **291**, 2351–2356 (2001). doi: [10.1126/science.1058890](https://doi.org/10.1126/science.1058890)
4. Y.J. Kim and A. Varki, "Perspectives on the significance of altered glycosylation of glycoproteins in cancer", *Glycoconj. J.* **14**, 569–576 (1997). doi: [10.1023/A:1018580324971](https://doi.org/10.1023/A:1018580324971)
5. D.H. Dube and C.R. Bertozzi, "Glycans in cancer and inflammation—potential for therapeutics and diagnostics", *Nat. Rev. Drug Discov.* **4**, 477–488 (2005). doi: [10.1038/nrd1751](https://doi.org/10.1038/nrd1751)
6. M. Granovsky, J. Fata, J. Pawling, W.J. Muller, R. Khokha and J.W. Dennis, "Suppression of tumor growth and metastasis in Mgat5-deficient mice", *Nat. Med.* **6**, 306–312 (2000). doi: [10.1038/73163](https://doi.org/10.1038/73163)
7. J. Taylor-Papadimitriou and A.A. Epenetos, "Exploiting altered glycosylation patterns in cancer: progress and challenges in diagnosis and therapy", *Trends Biotechnol.* **12**, 227–233 (1994). doi: [10.1016/0167-7799\(94\)90121-X](https://doi.org/10.1016/0167-7799(94)90121-X)
8. A. Harazono, N. Kawasaki T. Kawanishi and T. Hayakawa, "Site-specific glycosylation analysis of human apolipoprotein B100 using LC/ESI MS/MS", *Glycobiology* **15**(5), 447–462 (2005). doi: [10.1093/glycob/cwi033](https://doi.org/10.1093/glycob/cwi033)
9. M. Geng, X. Zhang, M. Bina and F. Regnier, "Proteomics of glycoproteins based on affinity selection of glycopeptides from tryptic digests", *J. Chromatogr. B.* **752**, 293–306 (2001). doi: [10.1016/S0378-4347\(00\)00550-8](https://doi.org/10.1016/S0378-4347(00)00550-8)
10. Z. Yang and W.S. Hancock, "Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column", *J. Chromatogr. A* **1053**, 79–88 (2004). doi: [10.1016/j.chroma.2004.08.150](https://doi.org/10.1016/j.chroma.2004.08.150)
11. M.R. Bond and J.J. Kohler, "Chemical methods for glycoprotein discovery", *Curr. Opin. Chem. Biol.* 52–58 (2007). doi: [10.1016/j.cbpa.2006.11.032](https://doi.org/10.1016/j.cbpa.2006.11.032)
12. K. Ueda, T. Katagiri, T. Shimada, S. Irie, T.A. Sato, Y. Nakamura and Y. Daigo, "Comparative profiling of serum glycoproteome by sequential purification of glycoproteins and 2-nitrobenzenesulfonyl (NBS) stable isotope labeling: a new approach for the novel biomarker discovery for cancer", *J. Proteome Res.* **6**, 3475–3483 (2007). doi: [10.1021/pr070103h](https://doi.org/10.1021/pr070103h)
13. U. Lewandrowski, J. Moebius, U. Walter and A. Sickmann, "Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach", *Mol. Cell Proteomics* **5**, 226–233 (2006). doi: [10.1074/mcp.M500324-MCP200](https://doi.org/10.1074/mcp.M500324-MCP200)
14. T.H. Patwa, J. Zhao, M.A. Anderson, D.M. Simeone and D.M. Lubman, "Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection", *Anal. Chem.* **78**, 6411–6421 (2006). doi: [10.1021/ac060726z](https://doi.org/10.1021/ac060726z)
15. T. Liu, W.J. Qian, M.A. Gritsenko, D.G. Camp, II, M.E. Monroe, R.J. Moore and R.D. Smith, "Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry", *J. Proteome Res.* **4**, 2070–2080 (2005). doi: [10.1021/pr0502065](https://doi.org/10.1021/pr0502065)
16. D.R. Meldrum, "Tumor necrosis factor in the heart. *Am. J. Physiol.* **274**, 577–595 (1998).
17. J. Renkonen, O. Tynnenen, P. Häyry, T. Paavonen and R. Renkonen, "Glycosylation might provide endothelial zip codes for organ-specific leukocyte traffic into inflammatory sites", *Am. J. Pathol.* **161**, 543–550 (2002).
18. A.A. Wakankar and R.T. Borchardt, "Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization", *J. Pharm. Sci.* **95**, 2321–2336 (2006). doi: [10.1002/jps.20740](https://doi.org/10.1002/jps.20740)
19. J.N. Arnold, R. Saldiva, U.M. Abd Hamid and P.M. Rudd, "Evaluation of the serum n-linked glycome for the diagnosis of cancer and chronic inflammation", *Proteomics* **8**, 3284–3293 (2008). doi: [10.1002/pmic.200800163](https://doi.org/10.1002/pmic.200800163)
20. D.J. Harvey, "Mass matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates", *Spectrom. Rev.* **18**, 349–451 (1999). doi: [10.1002/\(SICI\)1098-2787\(1999\)18:6<349::AID-MAS1>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H)
21. S. Ogata, T. Muramatsu and A. Kobata, "Fractionation of glycopeptides by affinity column chromatography on concanavalin A-sepharose", *J. Biochem. (Tokyo)* **8**, 87–96 (1975).
22. T. Krusius, "A simple method for the isolation of neutral glycopeptides by affinity chromatography", *FEBS Lett.* **66**, 86–89 (1976). doi: [10.1016/0014-5793\(76\)80591-1](https://doi.org/10.1016/0014-5793(76)80591-1)
23. T. Kremmer, É. Szöllösi, M. Boldizsár, B. Vincze, K. Ludányi, T. Imre, G. Schlosser and K. Vékey, "Liquid chromatographic and mass spectrometric analysis of human serum acid alpha-1-glycoprotein", *Biomed. Chromatogr.* **18**(5), 323–9 (2004). doi: [10.1002/bmc.324](https://doi.org/10.1002/bmc.324)
24. W. Van Dijk, E.C.M. Brinkman-Vander Linden and E.C. Navenaar, "Glycosylation of alpha-1-acid glycoprotein (orosomucoid) in healthy and disease: occurrence, regulation and possible functional implications", *TIGG* **10**, 235–245 (1998).
25. S. Hashimoto, T. Asao, J. Takahashi, Y. Yagihashi, T. Nishimura, A.R. Saniabadi, D.C. Poland, W. van Dijk, H. Kuwano, N. Kochibe and S. Yazawa, "alpha1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis", *Cancer* **101**(12), 2825–2836 (2004). doi: [10.1002/cncr.20713](https://doi.org/10.1002/cncr.20713)



hnRNP H1 and intronic G runs in the splicing control of the human rpl3 gene

Annapina Russo^{a,c}, Gabriella Siciliano^a, Morena Catillo^a, Chiara Giangrande^b, Angela Amoresano^b, Pietro Pucci^{b,c}, Concetta Pietropaolo^{a,*}, Giulia Russo^{a,*}

^a Dipartimento di Biochimica e Biotecnologie Mediche, Università Federico II, Via Sergio Pansini 5, Napoli 80131, Italy

^b Dipartimento di Chimica Organica e Biologica, Università Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia 4, Napoli 80126, Italy

^c CEINGE Biotecnologie Avanzate S.C.a.r.l., Via Comunale Margherita 482, Napoli 80145, Italy

ARTICLE INFO

Article history:

Received 7 August 2009

Received in revised form 15 January 2010

Accepted 19 January 2010

Available online 25 January 2010

Keywords:

Alternative splicing

hnRNP H1

NMD

Ribosomal protein

Splicing regulation

ABSTRACT

By generating mRNA containing a premature termination codon (PTC), alternative splicing (AS) can quantitatively regulate the expression of genes that are degraded by nonsense-mediated mRNA decay (NMD). We previously demonstrated that AS-induced retention of part of intron 3 of rpl3 pre-mRNA produces an mRNA isoform that contains a PTC and is targeted for decay by NMD. We also demonstrated that overexpression of rpl3 downregulates canonical splicing and upregulates the alternative splicing of its pre-mRNA. We are currently investigating the molecular mechanism underlying rpl3 autoregulation. Here we report that the heterogeneous nuclear ribonucleoprotein (hnRNP) H1 is a transacting factor able to interact in vitro and in vivo with rpl3 and with intron 3 of the rpl3 gene. We investigated the role played by hnRNP H1 in the regulation of splicing of rpl3 pre-mRNA by manipulating its expression level. Depletion of hnRNP H1 reduced the level of the PTC-containing mRNA isoform, whereas its overexpression favored the selection of the cryptic 3' splice site of intron 3. We also identified and characterized the cis-acting regulatory elements involved in hnRNP H1-mediated regulation of splicing. RNA electromobility shift assay demonstrated that hnRNP H1 specifically recognizes and binds directly to the intron 3 region that contains seven copies of G-rich elements. Site-directed mutagenesis analysis and in vivo studies showed that the G3 and G6 elements are required for hnRNP H1-mediated regulation of rpl3 pre-mRNA splicing. We propose a working model in which rpl3 recruits hnRNP H1 and, through cooperation with other splicing factors, promotes selection of the alternative splice site.

©2010 Elsevier B.V. All rights reserved.

1. Introduction

Alternative splicing (AS) is one of the main regulatory mechanisms of gene expression; it results in a repertoire of mRNAs, and consequently of proteins, much larger than expected from the number of genes. This process contributes substantially to cell-specific and tissue-specific gene expression; it is estimated that over 60% of human genes are alternatively spliced [1].

Splicing is modulated by cis elements in intron or exon sequences that, associated with specific transacting factors, can negatively (intronic or exonic splicing silencer sequences, ISS or ESS) or positively (intronic or exonic splicing enhancing sequences, ISE or ESE) affect splicing [2]. Although ESE elements differ in sequence,

most share a consensus sequence (GAR)_n. The UGCAUG esanucleotide is a frequent element in ISE sequences [3]. Most regulatory sequences are a complex combination of multiple elements that mediate positive or negative effects on gene splicing [4–6]. Intronic stretches of three or four Gs called “G runs” are involved in the splicing of a number of genes [7–13].

Among the factors found to control the selection of the splice site, components of the SR protein family have been shown to antagonize the activity of protein components of heterogeneous nuclear ribonucleoproteins (hnRNPs) [14,15]. hnRNPs are a large group of nuclear RNA-binding proteins that share a common structural domain denoted “qRRM” and are implicated in a variety of processes including RNA stability and translation. Accumulating evidence indicates that these proteins may play an important role in the control of splice site selection [16,17]. hnRNP H1 is a member of the hnRNPH/F family. It is able to bind the splicing regulatory cis elements, i.e. ESE, ISE, ESS and ISS. Depending on the interacting cis element and on the gene context, hnRNP H1 can affect the selection of the splice site and consequently promote either constitutive or alternative splicing [9–13,18].

Data from several laboratories demonstrate that some evolutionarily conserved AS events give rise to aberrant transcripts, which are substrate of nonsense-mediated mRNA decay (NMD). NMD is an RNA

Abbreviations: AS, alternative splicing; hnRNP, heterogeneous nuclear ribonucleoprotein; NMD, nonsense-mediated mRNA decay; PTC, premature termination codon; r-protein, ribosomal protein; rp, ribosomal protein; rp-mRNA, mRNA for ribosomal protein; RT-PCR, reverse transcriptase-PCR; SH-PTP1, SH-protein-tyrosine phosphatase 1; siRNA, small interfering RNA; SR, serine-rich

* Corresponding authors. G. Russo is to be contacted at tel.: +39 0817463061; fax: +39 0817463074. C. Pietropaolo, tel.: +39 0817463065; fax: +39 0817463074.

E-mail addresses: pietropaolo@dbbm.unina.it (C. Pietropaolo), russo@dbbm.unina.it (G. Russo).

surveillance mechanism that selectively degrades aberrant mRNAs that contain a premature termination codon (PTC), thus preventing the production of truncated polypeptides potentially deleterious to the cell. This process is highly conserved in eukaryotes. A PTC may arise through a nonsense or frameshift mutation of DNA, as a consequence of DNA rearrangement, or through splicing errors that produce aberrant mRNAs [19]. A PTC could also arise from alternative splicing that produces an intron-derived nonsense codon or a shift in the ORF that generates a downstream nonsense codon. The resulting aberrant mRNA is targeted for NMD rather than being translated into protein. Thus, the AS–NMD association results in quantitative post-transcriptional regulation of gene expression [20].

Genes that encode ribosomal proteins (rp) are regulated by AS–NMD. This regulatory strategy appears to be evolutionarily conserved in nematodes [21] and mammals [22]. We recently demonstrated that AS of genes for human proteins rpl3 and rpl12 generates alternative RNA isoforms consequent to the removal of part of intron 3 and of intron 1, respectively. The resulting mRNAs include intronic sequences that contain a PTC and are targeted for decay by NMD. We also demonstrated that overexpression of rpl3 results in the down-regulation of canonical splicing and upregulation of the alternative splicing of rpl3 pre-mRNA [22]. Here we report that hnRNP H1 is an rpl3 partner and plays an important role in the splicing regulation of rpl3 pre-mRNA. We also report that G-rich sequences in intron 3 are cis-regulatory elements involved in hnRNP H1-mediated regulation.

2. Materials and methods

2.1. Cell cultures, transfections and drug treatment

Human cell lines HeLa and Calu6 were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with glutamax (Invitrogen, Carlsbad, California) supplemented with 10% fetal bovine serum (FBS). In addition, Calu6 culture medium was supplemented with 0.1 mM non-essential amino acids (Euroclone, West York, UK). siRNA transfections were performed in HeLa cells (1×10^6 cells, 6 mm-well plate) at a concentration of 150 nM by using Oligofectamine Reagent (Invitrogen) according to the manufacturer's instructions.

Plasmids were transfected in HeLa cells (2.5×10^6 cells, 6 mm-well plate) by using Lipofectamine 2000 (Invitrogen) or Fugene6 (Roche Applied Science, Mannheim, Germany) according to the manufacturer's instructions. 24 h after DNA or siRNA transfections, cells were treated with 100 µg/ml cycloheximide for 4 h to block NMD. Then, RNA and proteins were extracted by using the Trizol procedure (Invitrogen) for RT-PCR analysis and western blot, respectively. Transfection efficiency was assessed by co-transfecting a GFP-expressing vector and normalizing RNA levels against GFP mRNA levels (data not shown).

2.2. Constructs, mutagenesis, and RNA interference

The cDNAs of hnRNP H1 and rpl3 were obtained by RT-PCR from HeLa cells using specific primers (Table 1) and cloned into the prokaryotic expression vector pGEX4T3 (GE Healthcare, Waukesha, Wisconsin). The cDNA of hnRNP H1 was also cloned into a version of the eukaryotic expression vector pcDNA4/HisMax C (Invitrogen) containing the HA epitope. The full intron 3 of rpl3 gene, or regions A, B and C (Fig. 3A) were obtained by RT-PCR using specific primers (Table 1) and cloned into the pGEM-4Z vector (Promega, Madison, Wisconsin) to obtain constructs pGEM4Z-Int3, pGEM4Z-IntA, pGEM4Z-IntB and pGEM4Z-IntC, respectively. The wt-rpl3 minigene was kindly provided by Y. Tang (Dr S. Stamm's laboratory, Institute for Biochemistry, University of Erlangen–Nuremberg, Germany). It contains the genomic region of the rpl3 gene spanning from exon 3 to exon 5 flanked by insulin exon sequences (Fig. 5A) cloned in the Exontrap vector (MoBiTec, Gottingen, Germany). The constructs

Table 1
Oligonucleotides sequences.

	5'-Primer (5'–3')	3'-Primer (3'–5')
GST-rpl3	ATGTCTCACAGAAAGTTC	AGCTCCTTCTTCTTTC
Int3	AATTGGTAAGGGAGGAG	TATGCCTTCAGGAGCAGA
rpl3-a	CTCCGCTGGGCTCTGCC	CTTCAGGAGCAGAGCAGA
rpl3-c	GGGCATTGTGGGCTACGT	GTAAGGCCTTCTTCTTAG
IntA	AAGAATTGGTAAGGGAGG	GGAAGCCCACTCAGTGAT
IntB	CCAGGGCAAAGGTTTG	GTAAGGCCTTCTTCTTAG
IntC	AAGAATTGGTAAGGGAGG	TCAGAATGAGGGTGTAGC
GST-hnRNP H1	ATGATGTTGGGCACGGAA	CTATGCAATGTTTGATTG
hnRNP H1-HA	ATGATGTTGGGCACGGAA	CTATGCAATGTTTGATTG
β-Actin	GGCACCCTTCTACA	CAGGAGGACAATGAT
Insulin	GCGAAGTGGAGGACCCACA	ACCCAGCTCCAGTTGTGCCA

carrying the disruption of single or multiple G runs, i.e., GGG/GAT mutation, were obtained by PCR site-directed mutagenesis (primers in Table 1) using the QuickChange® Lightning Site-Directed Mutagenesis Kit (Stratagene, La Jolla, California) and the wt-rpl3 minigene as template. All constructs were verified by DNA sequencing.

The small interfering RNAs (siRNA) targeting hnRNP H1 were: 5'-GGAAATAGCTGAAAAGGCT-3' and 5'-CCACGAAAGCTTATGGCCA-3'. The siRNA and the Silencer® Negative Control #1 siRNA were purchased from Ambion (Foster City, California).

2.3. Nuclear extracts

Nuclear extracts were prepared from Calu6 cells. Briefly, cells from twenty-one 100-mm plates (1.4×10^8 cells) were washed with ice cold PBS and collected in 7 ml of lysis buffer containing 20 mM Hepes–KOH, pH 8.0, 5 mM MgCl₂, 25 mM KCl, 0.5% NP40, 0.25% sodium-deoxycholate, 5 mM DTT and Protease Inhibitor Mix 1X (Roche), and passed through a needle (23 gauge). The cell lysate was incubated for 10 min on ice, then cell nuclei were separated from the cytoplasmic fraction by centrifugation for 5 min at 1000 ×g at 4 °C. Nuclei from twenty-one 100-mm plates were resuspended in 2 ml of nuclear lysis buffer (20 mM Hepes–KOH, 5 mM MgCl₂, 0.5 M NaCl, 20% glycerol, 1 mM EDTA, 5 mM DTT and Protease Inhibitor Mix 1X) and incubated for 10 min on ice, then clarified by centrifugation for 10 min at 10,000 ×g at 4 °C.

2.4. GST pull-down, mass spectrometry and protein identification

The recombinant proteins GST-rpl3 and GST were expressed in *Escherichia coli* and purified by using glutathione-Sepharose 4B beads according to the manufacturer's instructions (GE Healthcare). For GST pull-down assay, 100 µg of GST-rpl3 fusion protein, or GST control, was immobilized on glutathione-Sepharose beads and incubated with 10 mg of nuclear extract from Calu6 cells in pull-down buffer (50 mM Tris–HCl pH 7.5, 0.4 mM EDTA, 150 mM NaCl, 10% glycerol, 1% NP40, 1 mM sodium-ortovanadate, 50 mM NaF, 5 mM DTT and Protease Inhibitor Mix 1X) at 4 °C for 1.5 h. The beads were washed extensively and boiled in SDS sample buffer. The eluted proteins were loaded on 12% SDS-PAGE. Gels were stained with a colloidal blue staining kit (Invitrogen) and protein bands were excised from the gel and destained by repetitive washes with 0.1 M NH₄HCO₃, pH 7.5 and acetonitrile. Samples were reduced by incubation with 50 µl of 10 mM DTT in 0.1 M NH₄HCO₃ buffer, pH 7.5 and carboxyamidomethylated with 50 µl of 55 mM iodoacetamide in the same buffer. Enzymatic digestion was carried out with trypsin (12.5 ng/µl) in 10 mM ammonium bicarbonate, pH 7.8. Gel pieces were incubated at 4 °C for 2 h. Trypsin solution was then removed and a new aliquot of the digestion solution was added; samples were incubated for 18 h at 37 °C. A minimum reaction volume was used to obtain complete rehydration of the gel. Peptides were then extracted by washing the gel particles

with 10 mM ammonium bicarbonate and 1% formic acid in 50% acetonitrile at room temperature.

Liquid chromatography and tandem mass spectrometry (LC/MS/MS) analyses were performed on an LC/MSD Trap XCT Ultra (Agilent Technologies, Palo Alto, CA) equipped with a 1100 nanoHPLC system and a chip cube (Agilent Technologies). Peptides were analyzed using data-dependent acquisition of one MS scan (mass range from 300 to 1800 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. Raw data from nanoLC/MS/MS analyses were used to query a non-redundant protein database using in-house MASCOT software (Matrix Science, Boston, Massachusetts) [23].

2.5. Immunoprecipitation and western blotting

For immunoprecipitation assay, 1 mg of HeLa whole cell lysate was incubated with 30 μ l of protein A/G agarose beads coated with 5 μ g of antibody against rpL3 (Primm Milan, Italy), hnRNP H1 and SH-PTP1 (Santa Cruz Biotechnology, Santa Cruz, California) at 4 °C for 12 h. The beads were washed and boiled in the SDS sample buffer. The eluted proteins were loaded on 12% SDS-PAGE and detected by western blotting. Aliquots of protein samples (30 μ g) were resolved by 12% SDS-gel electrophoresis and transferred into nitrocellulose filters. The membranes were blocked in PBS, 0.2% Tween and 5% dry milk for 2 h, and then challenged with anti-rpL3 (Primm), anti-hnRNP H1, anti- α -

tubulin, anti-SH-PTP1 (Santa Cruz Biotechnology), and anti-HA (Roche). The proteins were visualized with enhanced chemiluminescence detection reagent according to the manufacturer's instructions (Pierce, Rockford, Illinois).

2.6. RNA pull-down assay

To identify proteins that bind specifically to the intron 3 transcript, we carried out an RNA pull-down assay using adipic acid dehydrazide beads. Briefly, 20 μ g of intron 3 RNA, transcribed in vitro from pGEM4Z-Int3, was placed in a 400 μ l reaction mixture containing 100 mM NaOAc, pH 5.2 and 5 mM sodium m-periodate (Sigma, St. Louis, Missouri), incubated for 1 h in the dark at room temperature, ethanol-precipitated, and resuspended in 100 μ l of 100 mM NaOAc, pH 5.2. Then 300 μ l of adipic acid dehydrazide agarose beads 50% slurry (Sigma) equilibrated in 100 mM NaOAc, pH 5.2 was added to this mixture, which was then incubated for 12 h at 4 °C on a rotator. The beads with the bound RNA were pelleted, washed twice with 1 ml of 2 M NaCl, and equilibrated in washing buffer (5 mM Hepes, pH 7.9, 1 mM MgCl₂, and 0.8 mM magnesium acetate). The intron 3 RNA was then incubated with 6 mg of protein extract from HeLa cells for 30 min at room temperature in a final volume of 0.6 ml. Heparin was also added to a final concentration of 7 μ g/ μ l. The beads were then washed four times in 1.5 ml of washing buffer. Bound proteins were

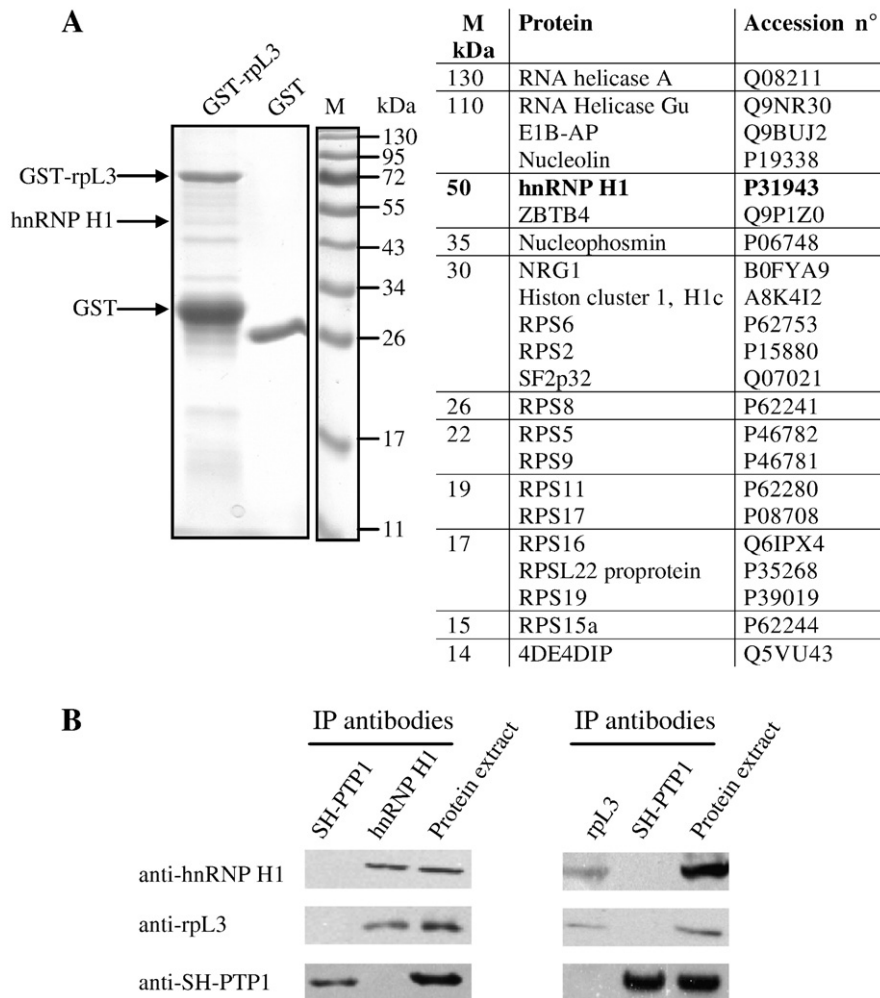


Fig. 1. In vitro and in vivo binding of hnRNP-H1 to rpL3. (A) In vitro binding of hnRNP H1 to rpL3. Coomassie Blue-stained gel of a GST pull-down experiment. The recombinant GST-rpL3 proteins and the GST control were incubated with nuclear extracts from Calu6 cells in the presence of glutathione-Sepharose beads. The eluted proteins were resolved by SDS-PAGE. M indicates the mobility of protein size markers. The arrows indicate the bands corresponding to GST-rpL3, hnRNP H1 and GST. The rpL3 partner proteins are listed. (B) In vivo binding of hnRNP H1 to rpL3. hnRNP H1 or rpL3 were immunoprecipitated from HeLa cells extracts with antibodies against hnRNP H1 and rpL3, respectively. Immunoprecipitates were separated by SDS-PAGE and immunoblotted for the proteins indicated at the left of the panel. These results are representative of three independently performed experiments.

eluted in SDS sample buffer and loaded on a 12% gel for SDS-PAGE to proceed to a MS analysis (see 2.4).

2.7. RNP immunoprecipitation assay (RIPA)

For RNP immunoprecipitation assay, HeLa cells (2×10^6 cells) were lysed in 600 μ l RIPA buffer $1 \times$ (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1 mM Na ortovanadate, 0.05 M NaF, and 0.5% NP40) with protease inhibitors (Roche) for 60 min on ice and then centrifuged at $10,000 \times g$ at 4 °C for 15 min. The supernatant was subjected to a preclearing step in which it was incubated with 50 μ l of protein A/G plus agarose for 1 h at 4 °C. The precleared cell extracts were then incubated with antibodies specific for each protein (Santa Cruz) overnight at 4 °C. Protein G-Sepharose beads (50 μ l of 50% slurry) were then added and the mix was incubated for 1 h at 4 °C with gentle shaking and centrifuged. The immunoprecipitates were suspended in 100 μ l TES buffer (10 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, and 0.5% SDS), incubated at 65 °C for 10 min and centrifuged for 1 min at $10,000 \times g$. Ten microliters was stored as immunoprecipitated

samples and subsequently fractionated by SDS-PAGE to be analyzed by western blotting. RNA was extracted from 90 μ l using the Trizol (Invitrogen) procedure.

2.8. RNA electromobility shift assay (REMSA)

RNA probes were transcribed in vitro using 32 P-NTPs and SP6 polymerase according to the manufacturer's instructions (Roche), and construct pGEM4Z-IntA, pGEM4Z-IntB, or pGEM4Z-IntC as template. The probes were purified using Sephadex G-25 columns (Roche). 5×10^5 counts/min of each RNA probe were incubated in binding buffer containing 20 mM Hepes, pH 7.9, 150 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.1% Triton, 1 mg/ml tRNA, and 0.4 U/ μ l RNase Inhibitor (Roche) in the presence or absence of the recombinant protein GST-hnRNP H1 or GST for 1 h at 4 °C. The complexes formed were then resolved onto a 5% polyacrylamide (37.5:1 acrylamide:bis-acrylamide ratio) RNA native gel. The gel was dried at 80 °C for 30 min and the results visualized by autoradiography.

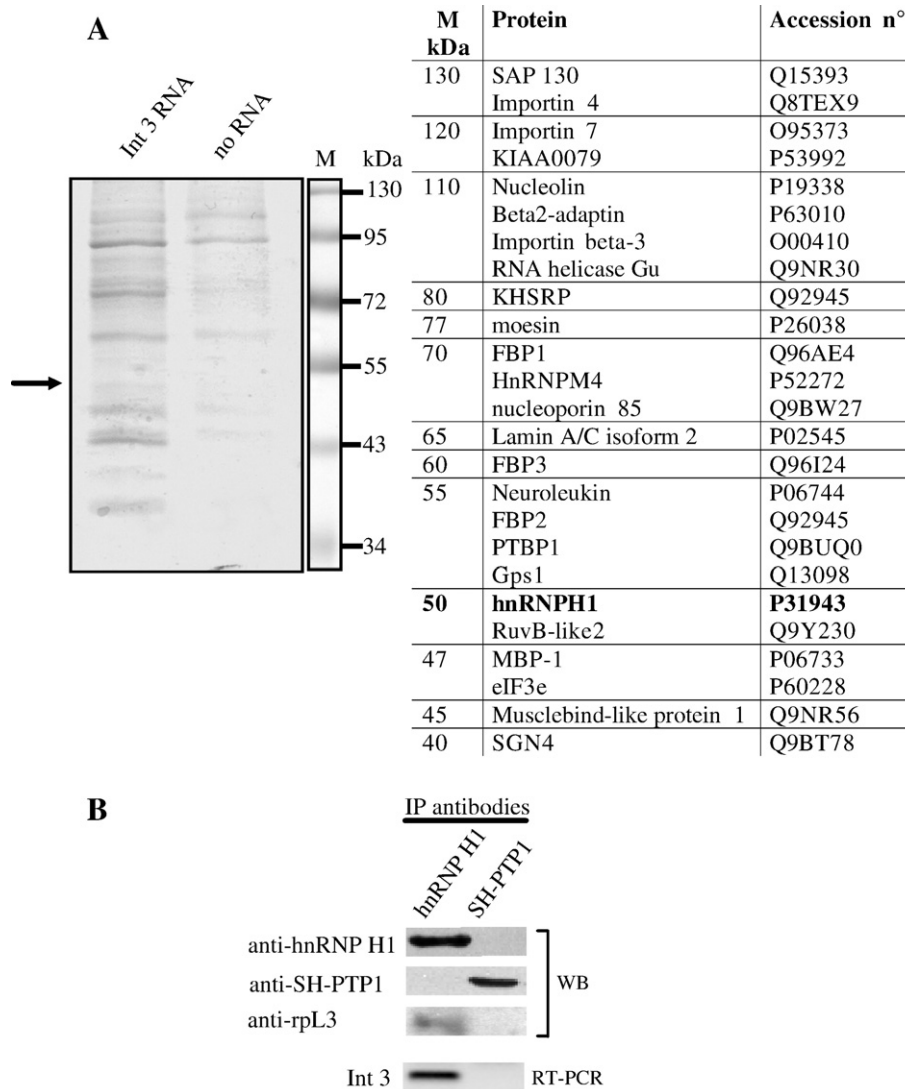


Fig. 2. Interaction of intron 3 of the rpL3 gene with hnRNP H1 in vitro and in vivo. (A) In vitro binding of the intron 3 transcript to hnRNP H1. Coomassie Blue-stained gel of a RNA pull-down experiment, using adipic acid dehydrazide agarose beads coated with intron 3 transcript was incubated with HeLa nuclear protein extract. The mobility of protein size markers is shown (M). The arrow indicates the protein band corresponding to hnRNP H1 identified by mass spectrometry analysis of eluted peptides. The intron 3 RNA interacting proteins identified are listed. (B) In vivo binding of the intron 3 transcript to hnRNP H1. Western blotting (WB) of hnRNP H1 or SH-PTP1 immunoprecipitates from HeLa cells, and RT-PCR of RNA extracted from the same immunoprecipitates by using primers against intron 3 (Table 1). Note the absence of signal in SH-PTP1 immuno-complex. These results are representative of three independently performed experiments.

2.9. RT-PCR

For RT-PCR analysis, 1 µg of total RNA was reverse-transcribed into cDNA with the random hexamer technique using 200 U of Superscript II RNase H⁻ Reverse Transcriptase (Invitrogen). The reaction was carried out at 42 °C for 50 min and was terminated by heating to 75 °C for 15 min. Ten of the 40 µl of reaction mix were PCR-amplified in a final volume of 50 µl, using 5 µM of each specific primer (Table 1), 10 mM dNTPs, and 0.5 U of Taq DNA polymerase (Invitrogen). Typically, 25–30 cycles of amplification were performed. In separate experiments, we ascertained that the cycle number was within the linear range of amplifications. PCR products were visualized on 1% agarose gel containing the fluorescent Vista Green dye (Amersham Pharmacia Biotech). The labeling intensity of the PCR product, which is linear to the amount of DNA, was quantified using the Phosphor-Imager (Bio-Rad, Haercules, California).

3. Results

3.1. rpl3 interacts with hnRNP H1 in vitro and in vivo

An important issue concerning the autoregulation of rpl3 expression via AS–NMD coupling is to identify molecular partners of rpl3 involved in the splicing event. We previously found that rpl3 does not bind its pre-mRNA. Specifically, filter-binding experiments with the recombinant GST-rpl3 and GST alone as control showed that

rpl3 could interact with intron 3 RNA as well as with unrelated RNA (data not shown). These findings favored the hypothesis that rpl3 modulates the splicing of its own gene by interacting with one or more regulatory proteins. Consequently, to search for rpl3 transacting factors, we carried out an in vitro GST pull-down assay. Recombinant GST-rpl3 and GST proteins were purified from *E. coli* cells, immobilized using GSH-Sepharose beads and incubated with a nuclear extract from Calu6 cells. The proteins specifically bound to GST-rpl3 and GST were pulled down, fractionated by SDS-PAGE and then visualized by Coomassie Blue staining.

The whole gel was cut in slices and protein bands, destained by repetitive washings, were excised and in situ-digested with trypsin. We also analyzed the control (proteins bound to GST) to check for nonspecific proteins. The resulting peptide mixtures were extracted from the gel and directly analyzed by LC/MS/MS to obtain sequence information on individual peptides. This information together with the peptide mass values was then used to search protein databases for corresponding proteins. The comparison of GST-rpl3 beads versus control GST beads revealed several specific proteins. Among these, sequence analysis by MS of an excised 50-kDa band yielded several peptides among which is hnRNP H1 (Fig. 1A). hnRNP H1 plays a major role in the regulation of alternative splicing of various genes. Recent studies showed that hnRNP H1 is able to bind specifically G runs on a transcript molecule [9–13,18].

To verify the interaction between rpl3 and hnRNP H1 in vivo, we performed co-immunoprecipitation experiments. hnRNP H1 (Fig. 1B,

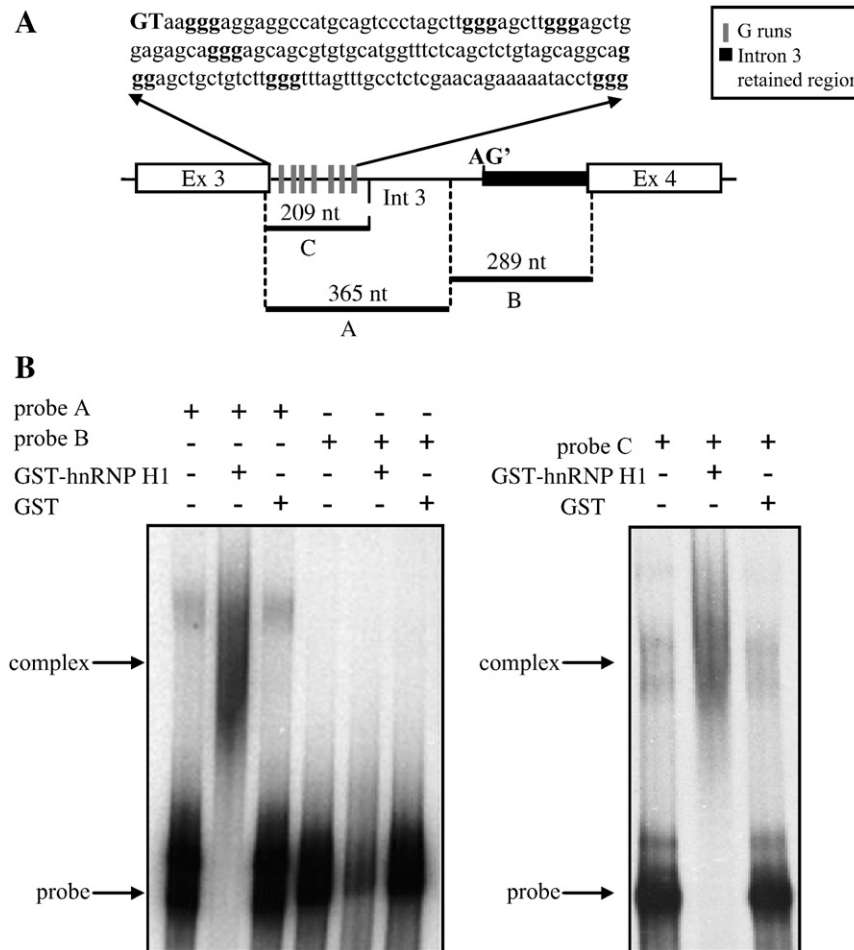


Fig. 3. hnRNP H1 binding to region A within intron 3 in vitro. (A) Schematic representation of regions A (365 nt), B (289 nt) and C (209 nt) of intron 3. The sequence of part of the region C of intron 3 including the G-run motifs is shown (boldface). (B) Gel mobility shift analysis of hnRNP H1 challenged with regions A, and B of intron 3 (left panel), or with region C of intron 3 (right panel). The results are representative of three independently performed experiments.

left panel) and rpL3 (Fig. 1B, right panel) were specifically immunoprecipitated from HeLa cell extracts and the immunoprecipitated complexes were tested by western blotting. A signal for both proteins appeared in immuno-complexes probed with antibodies against rpL3 and hnRNP H1, thus confirming that the two proteins specifically interact in vivo (Fig. 1B). A control immunoprecipitate obtained with an arbitrary antibody, anti-SH-PTP1, did not give any signal when probed with anti-hnRNP H1 or anti-rpL3 (Fig. 1B).

3.2. hnRNP H1 binds to the intron 3 of the rpL3 gene transcript in vitro and in vivo

We next performed an in vitro RNA pull-down assay to identify and isolate the RNA-binding proteins taking part in the alternative splicing event and, specifically, to determine whether hnRNP H1 was able to interact not only with rpL3 but also with its pre-mRNA. A transcript corresponding to the entire intron 3 of rpL3 pre-mRNA was used as bait for pull-down interacting proteins. The beads coated with the intron 3 transcript were incubated with a total protein extract from HeLa cells. After stringent washing, RNA-associated proteins

were eluted and analyzed by SDS-PAGE. The comparison of the proteins bound to the intron 3 RNA-coated beads versus control beads with no RNA showed various specific bands (Fig. 2A). Each protein slice was cut and in situ-digested as previously described. Mass spectral analysis of the protein mixtures revealed several proteins among which is hnRNP H1.

To evaluate whether intron 3 and hnRNP H1 specifically interacted in vivo, we immunoprecipitated hnRNP H1 from HeLa cell extracts and looked for the intron 3 transcript and rpL3 in the RNA–protein immunoprecipitate complex (Fig. 2B). The specific primers, used in the RT-PCR to detect the intron 3 transcript were 18-mer oligonucleotides mapping 5 nt upstream from the 5' splice donor site and 5 nt downstream from the 3' splice acceptor site (Int3 in Table 1). We used western blot and anti-rpL3 antibodies to detect rpL3 in the immuno-complex. Amplification of the signal corresponding to the intron 3 transcript indicated that hnRNP H1 was able to bind rpL3 pre-mRNA in vivo. Moreover, the co-presence of rpL3 in the immuno-complex suggests that rpL3 pre-mRNA is present in a ribonucleoprotein complex that includes rpL3 and hnRNP H1, although the possibility of two independent complexes cannot be ruled out. The absence of

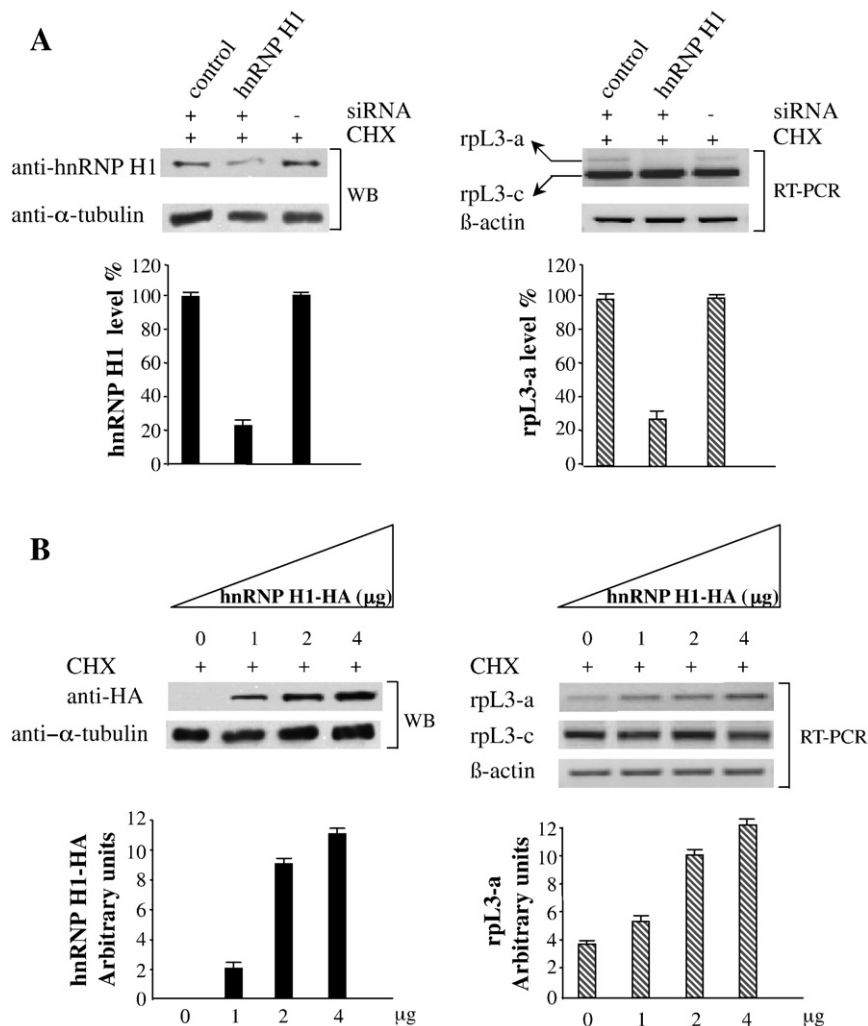


Fig. 4. (A) Effects of RNAi-mediated depletion of hnRNP H1 on rpL3-a mRNA levels. Left panel: protein samples from CHX-treated HeLa cells transfected with hnRNP H1-siRNA (hnRNP H1), with unrelated siRNA (control) or untransfected (–), were analyzed by western blotting (WB) with antibody specific for endogenous hnRNP H1. Loading in the gel lanes was controlled by detection of tubulin protein. Right panel: representative RT-PCR analysis of total RNA from the same samples. β -actin was used as a control of RNA loading. Quantification of hnRNP H1 protein and of the level of rpL3-a mRNA by PhosphorImager (Bio-Rad) is shown. (B) Effects of hnRNP H1 overexpression on rpL3-a mRNA levels. Left panel: protein overexpression detected with an antibody directed against the HA tag. Right panel: representative RT-PCR analysis of total RNA from CHX-treated HeLa cells untransfected or transfected with increasing amounts of hnRNP H1-HA. Quantification of hnRNP H1-HA protein and of the level of rpL3-a mRNA by PhosphorImager (Bio-Rad) is shown. These results are representative of three independently performed experiments.

signal in the immunoprecipitate with anti-SH-PTP1 confirmed the validity of this assay (Fig. 2B).

3.3. hnRNP H1 binds selectively to 5' portion of intron 3

To identify the cis-acting elements of intron 3 that participate in the regulation of rpl3 pre-mRNA splicing, we used a computational analysis of this region to look for hnRNP H1-binding sequence motifs. The search revealed seven G-rich motifs that could be hnRNP H1-binding sites in the 5' moiety on intron 3. Because of experimental evidence that hnRNP H1 activity in the splicing event is mediated by the binding of the protein to G runs, we performed REMSA experiments to assess whether the putative binding sites are necessary for the interaction of hnRNP H1 with intron 3 in the pre-mRNA transcript. We evaluated the hnRNP H1 binding of three intron 3 regions: region A spanning 365 nt downstream from the 5' splice donor splice site containing seven G-run elements; region B (289 nt) constituted by the 3' moiety of intron 3 and 27 nt of exon 4; and region C (209 nt), a more restricted region encompassing the seven G runs (Fig. 3A). The three regions were transcribed in vitro using 32 P-NTPs. The RNA probes were incubated with GST-hnRNP H1 or GST recombinant protein. The resulting RNA–protein complexes were analyzed on a native polyacrylamide gel and the results visualized by autoradiography.

As shown in Fig. 3, GST-hnRNP H1 was able to interact specifically with RNA transcribed on regions A (Fig. 3B, left panel) and C (Fig. 3B, right panel) of intron 3, but failed to do so when challenged with the B region transcript (Fig. 3B). No mobility shift was observed when recombinant GST was challenged with a transcript constituted by the

three regions of intron 3 (Fig. 3B). These results indicate that hnRNP H1 binds in vitro to the region of the intron 3 transcript that contains seven G runs.

3.4. hnRNP H1 regulates the alternative splicing of rpl3 pre-mRNA

To determine the role played by hnRNP H1 in the alternative splicing of rpl3 primary transcript, we evaluated the effects of altered hnRNP H1 expression on the selection of the alternative 3' acceptor splice site in intron 3 of the rpl3 gene. To this aim we switched-off the expression of the gene encoding hnRNP H1 by using RNA interference. siRNAs against hnRNP H1 or unrelated siRNAs were transiently transfected in HeLa cells. After transfection, cells were treated with CHX to stabilize the otherwise labile alternative splice form, and harvested; then RNA and protein were extracted. Lysates from cells transfected with unrelated siRNAs (control), specific siRNAs (hnRNP H1), or untransfected were probed with hnRNP H1 antibodies. As shown in Fig. 4A, the residual level of hnRNP H1 was about 30% of the protein detected in the control lysates. To examine the effects of the impaired production of hnRNP H1 on the splicing pattern of rpl3 pre-mRNA, we examined the rpl3 mRNA isoforms using RT-PCR. Removal of hnRNP H1 resulted in a decrease (about 70%) of the alternative mRNA (rpl3-a) level compared to controls.

To investigate the effects of hnRNP H1 overexpression we fused a cDNA encoding hnRNP H1 to an HA tag in the expression vector pcDNA3. Increasing amounts of this construct were transiently transfected in HeLa cells. 24 h after transfection, the production of hnRNP H1 was measured by western blotting (Fig. 4B). Total RNA was

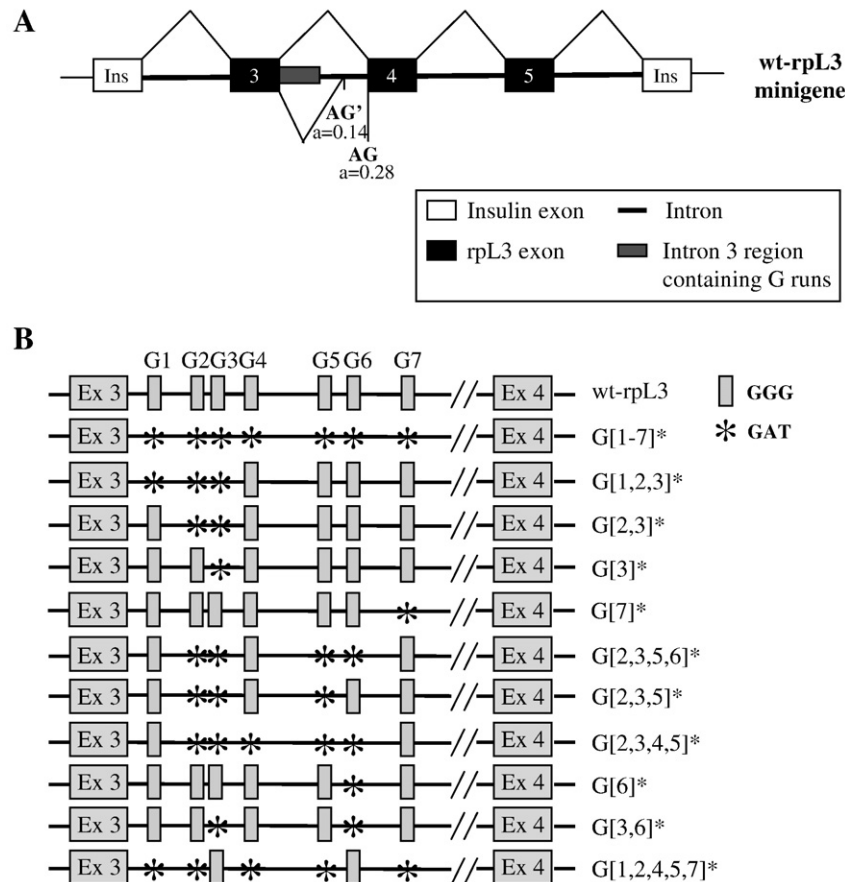


Fig. 5. (A) Schematic representation of wt-rpl3 minigene. The canonic acceptor site AG and the alternative acceptor site AG' are indicated in boldface. The score of acceptor splice sites calculated using the Splice Site Prediction by Neural Network program (SSPNN, http://www.fruitfly.org/seq_tools/splice.html) is indicated. (B) Scheme of G-run mutants. The G runs within rpl3 intron 3 are numbered 1 to 7, and the G-run mutants are indicated by *.

analyzed by RT-PCR by using specific primers (Table 1) to amplify canonical (rpL3-c) or alternative (rpL3-a) isoforms of rpL3 mRNA. Overexpression of hnRNP H1 caused a dose-dependent increase in rpL3-a mRNA level (Fig. 4B).

3.5. G runs are involved in the control of 3' splice site selection within rpL3 intron 3

To investigate whether guanosine stretches within intron 3 play a role in the control of splicing, we exploited a minigene containing a genomic region spanning from exon 3 to exon 5 (Fig. 5A, wt-rpL3 minigene) of the rpL3 gene. Using site-directed mutagenesis we changed GGG to GAT in all the G-run elements (G[1–7]^{*} in Fig. 5B). After transfection of the constructs in HeLa cells, we carried out an RT-PCR analysis to determine the splicing pattern of the corresponding transcripts. Specific primers, designed on the insulin gene sequences flanking rpL3 gene sequences, were used to discriminate the minigene splicing products from the products derived from the endogenous gene (Table 1). As shown in Fig. 6A, the wt-rpL3 minigene reproduced the splicing pattern of the endogenous gene and hence it contained the information necessary for correct rpL3 splicing. Differently, pre-mRNA processing was greatly impaired in the G[1–7]^{*} mutant, with a consequent decrease in the level of the alternative isoform of rpL3 mRNA to 20% versus the control wt-rpL3 minigene. These results

show that the guanosine-rich elements within intron 3 are involved in the control of the 3' splice site selection in rpL3 intron 3.

3.6. Functional mapping of the G motifs within intron 3

To evaluate the role of each G run in the regulation of the splicing of rpL3 pre-mRNA, we produced a series of wt-rpL3 minigene mutants. Mutation of elements G1, G2 and G3 (G[1–3]^{*} in Fig. 5B) reduced the level of rpL3-a mRNA by 50% (Fig. 6A). The same result was obtained with mutants G[2,3]^{*} and G[3]^{*}. Consequently, the alteration of the splicing pattern in these mutants can be ascribed solely to disruption of the G3 element. Differently, the splicing pattern of the G[7]^{*} mutant (Fig. 5B) shows that mutation of a single G run does not alter the level of rpL3-a mRNA. In another set of mutants, we evaluated the role of elements G3, G4, G5 and G6 in the splicing modulation of rpL3 intron 3. Mutants G[2,3,5]^{*} and G[2–5]^{*} yielded levels of rpL3-a mRNA similar to that of the G[2,3]^{*} minigene mutant. On the contrary, the addition of the mutated G6 element in G[2,3,5,6]^{*} reduced the splicing of rpL3-a mRNA to 20% of that of the control, as observed with mutant G[1–7]^{*} (Fig. 6A). The results reported above strongly suggest that elements G3 and G6 are directly involved in the regulation of the splicing of rpL3 transcript intron 3.

To verify the role of G3 and G6 elements, we generated and analyzed another set of minigenes (see Fig. 5B): one with a mutation of element G6 (G[6]^{*}), one with mutation of elements G3 and G6 (G[3,6]^{*}), and one

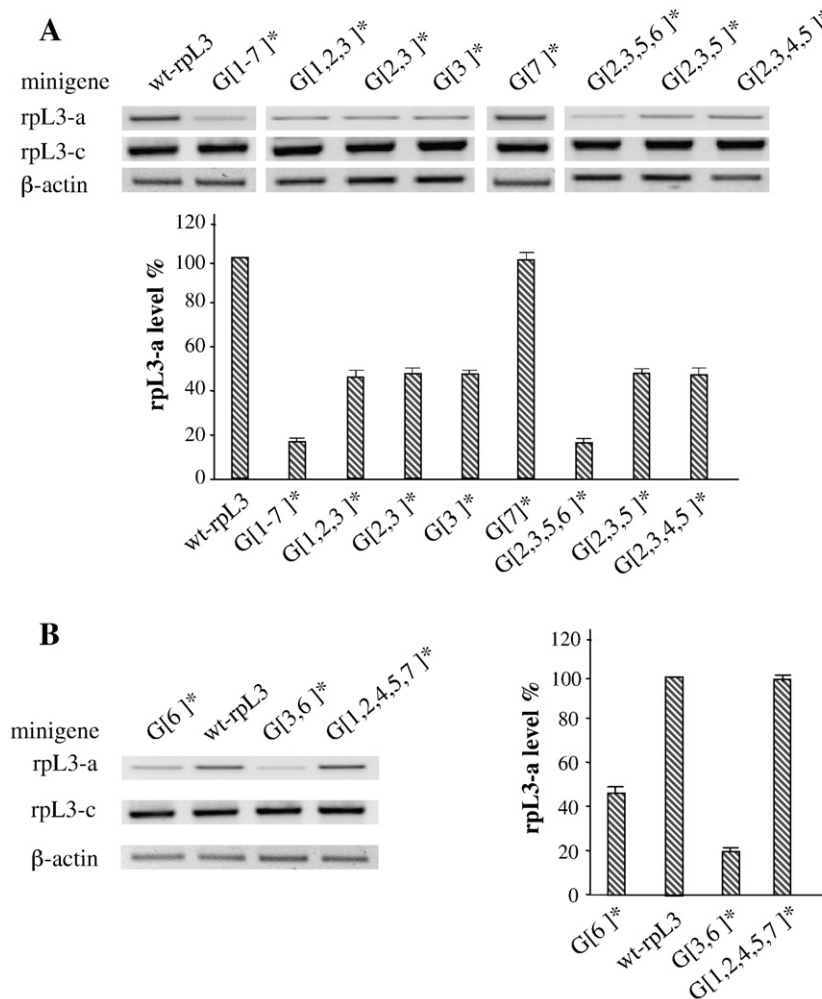


Fig. 6. (A and B) Functional mapping of the G-run elements within intron 3. Representative RT-PCR of total RNA extracted from CHX-treated HeLa cells transfected with the wt-rpL3 minigene or mutated constructs (Fig. 5) by using oligonucleotides that amplify minigene products (Table 1). β-actin was used as a control of RNA loading. Quantification of rpL3-a mRNA level using PhosphorImager (Bio-Rad) is shown.

with mutation of all G runs except elements G3 and G6 (G[1,2,4,5,7]^{*}). Mutation of element G6 alone resulted in a splicing pattern in which the rpL3-a mRNA isoform was about 45% that of the control. A stronger effect (80% reduction of the rpL3-a mRNA isoform) was observed in the splicing pattern of the double mutant G[3,6]^{*} (Fig. 6B); the level of the rpL3-a mRNA isoform was similar to the level obtained with mutant G[2,3,5,6]^{*} or G[1–7]^{*} (Fig. 6A). Thus, the G3 and G6 elements cooperate to control the selection of the alternative 3' splice site in intron 3. Finally, mutant G[1,2,4,5,7]^{*} (see Fig. 5B) had the same splicing pattern as the endogenous gene (Fig. 6B), thereby confirming that elements G3 and G6 play a relevant role in the selection of splicing modality of the rpL3 gene.

These data demonstrate that the G3 and G6 motifs cooperatively drive the recognition of the alternative 3' splice site in intron 3, thereby leading to the inclusion of part of intron 3 in the rpL3-a mRNA. Furthermore, these results together with data from hnRNP H1 silencing provide strong evidence that the effects on the splicing pattern of rpL3 pre-mRNA could be mediated directly by the interaction of hnRNP H1 with G3 and G6 elements. To verify these findings, we overexpressed hnRNP H1 in HeLa cells transfected with the wt-rpL3, G[2,3,5,6]^{*} or G[3,6]^{*} minigenes under conditions that could lead to a larger than three-fold increase of the rpL3-a transcript level. Despite hnRNP H1 overexpression, the splicing pattern of the double mutant G[3,6]^{*} showed that the level of alternative transcript was 60% lower than the level of the control wt-rpL3 minigene (Fig. 7). These data indicate that the G3 and G6 elements represent the consensus binding motifs of hnRNP H1 and are required for the hnRNP H1-mediated regulation of rpL3 pre-mRNA splicing.

4. Discussion

Ribosome biosynthesis requires the coordinated expression of the genes coding for the structural components, rRNA and r-proteins [24]. To maintain ribosome synthesis at the appropriate level, r-protein expression is regulated by multiple control mechanisms mostly at post-transcriptional and translational level. Moreover, r-proteins also

exert a variety of extra-ribosomal functions for which additional and specific regulatory strategies are required [25–29]. In eukaryotes, several post-transcriptional regulation mechanisms have been described that control the level of rp-mRNA in the cell by modulating transcript processing, splicing, and stability. Autoregulation mechanisms appear to be a strategy by which each individual r-protein may control the level of its own expression [21,22,30–32].

We previously demonstrated that, as a consequence of a partial retention of intronic sequences, splicing of the human rpL3 transcript gives rise to a canonical mRNA and to an alternative mRNA isoform containing a PTC that is targeted to NMD. Moreover, our finding that rpL3 overexpression downregulates the level of the canonically spliced mRNA, and upregulates the production of the alternatively spliced isoform demonstrated that production of human rpL3 is regulated via a negative feedback loop [22]. Coupling of AS and NMD appears to be a tool by which to fine tune ribosomal protein levels. However, the molecular mechanism through which the process occurs is unknown.

In an attempt to shed light on this mechanism we looked for molecular partners of rpL3, and identified a number of proteins able to bind in vitro rpL3, the intron 3 of rpL3 gene, or both (Figs. 1 and 2). In vivo experiments confirmed that the splicing factor hnRNP H1 could bind rpL3 as well as its pre-mRNA, and REMSA experiments showed that this interaction occurs through a pre-mRNA region containing seven G-run elements (Fig. 3). In addition, depletion of hnRNP H1 by siRNA resulted in a significant reduction of alternative splicing of rpL3 pre-mRNA (Fig. 4A), whereas its overexpression resulted in increased production of the aberrant rpL3 mRNA (Fig. 4B), which indicates that hnRNP H1 overexpression was associated with a more efficient usage of the alternative 3' acceptor site in intron 3.

G-run motifs are found preferentially in intronic sequences flanking a splicing site and appear to affect the recognition of an adjacent 5' splice site or 3' splice site of the intron. However, they can also function when located in exons [33].

Mutational analysis of the seven G-run elements in rpL3 intron 3 showed that G3 and G6 affected splicing (see Figs. 5B and 6). In fact, when both elements were mutated (G[3,6]^{*} in Figs. 5B and 6B), the level of the alternative mRNA isoform was reduced by 80%, as occurred when all G-run motifs were mutated. When wild-type G3 and G6 were conserved and all other G runs mutated (see G[1,2,4,5,7]^{*} in Figs. 5B and 6B), the amount of alternative rpL3 mRNA isoform was the same as the amount obtained from the wt-rpL3 primary transcript. Moreover, when G3 and G6 motifs were mutated, hnRNP H1 overexpression did not affect the level of the alternatively spliced mRNA isoform (Fig. 7). Given these results we concluded that G3 and G6 form a functional unit, and that its interaction with hnRNP H1 promotes the selection of a weak alternative 3' splice site of intron 3. Cooperation between different G-rich motifs has been widely documented. More copies may increase the affinity of an associated factor and/or recruit more copies of an interacting factor [33]. For instance, multiple intronic G motifs regulate the splicing of the thrombopoietin gene through a complex combinatorial mechanism [11].

The molecular mechanism by which hnRNP H1 regulates AS of the rpL3 transcript remains to be clarified. Using proteomic analysis, we identified several proteins that interacted in vitro with rpL3 and with the pre-mRNA of the rpL3 gene (Figs. 1A and 2A). It is plausible that more complex protein–protein interactions between RNA-binding proteins, hnRNP H1 and rpL3 are involved in the regulation of rpL3 gene splicing. In this context, our results are consistent with a model in which the rpL3 protein directs the splicing reaction towards the alternative mode by specifically interacting with hnRNP H1 (Fig. 8).

In a splicing process leading to the exclusion of exon IIIC of the FGR2 gene, Fox2 recruits silencing factors and assists hnRNP H1 in binding an exonic splicing silencer thereby resulting in exon exclusion [34]. Regarding the rpL3 gene, the process may serve to prevent wasteful production of the protein. Binding to rpL3 could cause conformational changes in hnRNP H1 that favor interaction with the

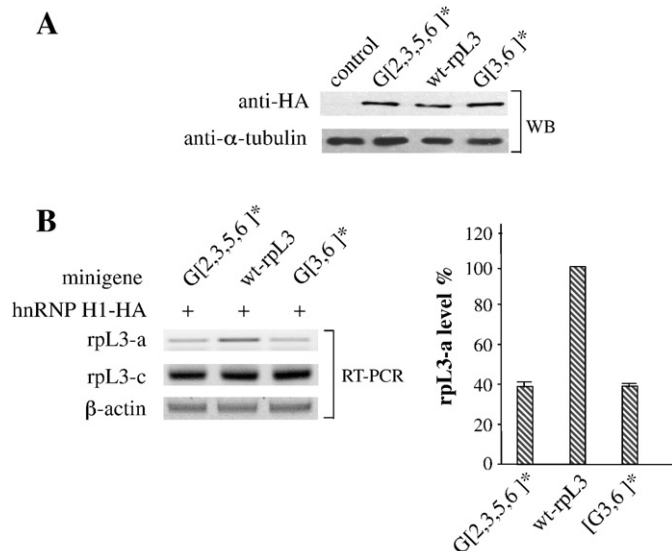


Fig. 7. hnRNP H1 regulates rpL3 pre-mRNA alternative splicing by binding to G3 and G6 elements. (A) hnRNP H1-HA was detected by western blotting (WB) using antibodies against the HA epitope. Loading in the lanes was controlled by detection of tubulin protein. (B) Representative RT-PCR of total RNA extracted from CHX-treated HeLa cells transfected with hnRNP H1-HA (4 μg, control) or co-transfected with G[2,3,5,6]^{*} or G[3,6]^{*} or the wt-rpL3 minigene (Fig. 5) and hnRNP H1-HA (4 μg). β-actin was used to control RNA loading. Quantification of rpL3-a mRNA level using PhosphorImager (Bio-Rad) is shown. These results are representative of three independently performed experiments.

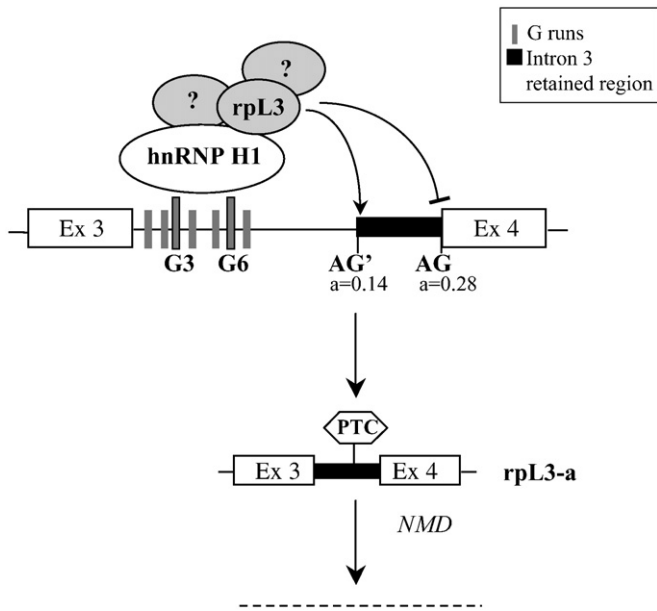


Fig. 8. Schematic representation of rpL3 feedback regulation mediated by hnRNP H1. rpL3 binds to hnRNP H1 and recruits it on intron 3. hnRNP H1 interacts with G3 + G6 units within intron 3, and, through interaction with other transacting proteins, an RNP complex promotes the selection of the weak 3' acceptor site in intron 3.

alternative splicing enhancer unit G3 + G6 within intron 3, thereby allowing correct positioning of the spliceosome, and recruitment of additional protein factors. Ultimately, such a complex could favor the selection of the alternative 3' splice site.

The ongoing analysis of protein components of the RNP complex involved in the selection of the alternative 3' splice site will help to elucidate the fine autoregulatory network of rpL3 expression.

Acknowledgements

This work has been supported by MIUR, Fondo Investimenti Ricerca di Base (FIRB 2001) and Regione Campania, L5/2002. Authors thank Nicoletta Sorvillo for the help at the early stage of the project, and Jean Ann Gilder for editing the text.

References

- [1] B. Modrek, A. Resch, C. Grasso, C. Lee, Genome-wide detection of alternative splicing in expressed sequences of human genes, *Nucleic Acids Res.* 29 (2001) 2850–2859.
- [2] D.L. Black, Mechanisms of alternative pre-messenger RNA splicing, *Annu. Rev. Biochem.* 72 (2003) 291–336.
- [3] L.P. Lim, P.A. Sharp, Alternative splicing of the fibronectin EIIIB exon depends on specific TGATG repeats, *Mol. Cell. Biol.* 18 (1998) 3900–3906.
- [4] R.C. Chan, D.L. Black, Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro, *Mol. Cell. Biol.* 17 (1997) 2970.
- [5] E.F. Modafferi, D.L. Black, A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon, *Mol. Cell. Biol.* 17 (1997) 6537–6545.
- [6] K. Han, G. Yeo, P. An, C.B. Burge, P.J. Grabowski, A combinatorial code for splicing silencing: UAGG and GGGG motifs, *PLoS Biol.* 3 (2005) e158.
- [7] A.J. McCullough, S.M. Berget, G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection, *Mol. Cell. Biol.* 17 (1997) 4562–4571.

- [8] E.M. McCarthy, J.A. Phillips III, Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA, *Hum. Mol. Genet.* 7 (1998) 1491–1496.
- [9] M. Caputi, A.M. Zahler, Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family, *J. Biol. Chem.* 276 (2001) 43850–43859.
- [10] M.C. Schaub, S.R. Lopez, M. Caputi, Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes, *J. Biol. Chem.* 282 (2007) 13617–13626.
- [11] R. Marcucci, F.E. Baralle, M. Romano, Complex splicing control of the human Thrombopoietin gene by intronic G runs, *Nucleic Acids Res.* 35 (2007) 132–142.
- [12] A.J. McCullough, S.M. Berget, An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites, *Mol. Cell. Biol.* 20 (2000) 9225–9235.
- [13] C.D. Chen, R. Kobayashi, D.M. Helfman, Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene, *Genes Dev.* 13 (1999) 593–606.
- [14] M. Caputi, A.M. Zahler, SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D, *Embo J.* 21 (2002) 845–855.
- [15] A. Expert-Bezancon, A. Sureau, P. Durosay, R. Sables, H. Groeneveld, J.P. Lecaer, J. Marie, hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B, *J. Biol. Chem.* 279 (2004) 38249–38259.
- [16] A.M. Krecic, M.S. Swanson, hnRNP complexes: composition, structure, and function, *Curr. Opin. Cell Biol.* 11 (1999) 363–371.
- [17] R. Martinez-Contreras, P. Cloutier, L. Shkreta, J.F. Fiset, T. Revil, B. Chabot, hnRNP proteins and splicing control, *Adv. Exp. Med. Biol.* 623 (2007) 123–147.
- [18] A. Masuda, X.M. Shen, M. Ito, T. Matsuura, A.G. Engel, K. Ohno, hnRNP H enhances skipping of a nonfunctional exon P3A in CHRNA1 and a mutation disrupting its binding causes congenital myasthenic syndrome, *Hum. Mol. Genet.* 17 (2008) 4022–4035.
- [19] Y.F. Chang, J.S. Imam, M.F. Wilkinson, The nonsense-mediated decay RNA surveillance pathway, *Annu. Rev. Biochem.* 76 (2007) 51–74.
- [20] N.J. McGlincy, C.W. Smith, Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* 33 (2008) 385–393.
- [21] Q.M. Mitrovich, P. Anderson, Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*, *Genes Dev.* 14 (2000) 2173–2184.
- [22] M. Cuccurese, G. Russo, A. Russo, C. Pietropaolo, Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression, *Nucleic Acids Res.* 33 (2005) 5965–5977.
- [23] E. Zito, M. Buono, S. Pepe, C. Settembre, I. Annunziata, E.M. Surace, T. Dierks, M. Monti, M. Cozzolino, P. Pucci, A. Ballabio, M.P. Cosma, Sulfatase modifying factor 1 trafficking through the cells: from endoplasmic reticulum to the endoplasmic reticulum, *Embo J.* 26 (2007) 2443–2453.
- [24] F. Loreni, A. Francesconi, F. Amaldi, Coordinate translational regulation in the syntheses of elongation factor 1 alpha and ribosomal proteins in *Xenopus laevis*, *Nucleic Acids Res.* 21 (1993) 4721–4725.
- [25] D.I. Friedman, A.T. Schauer, M.R. Baumann, L.S. Baron, S.L. Adhya, Evidence that ribosomal protein S10 participates in control of transcription termination, *Proc. Natl. Acad. Sci. U. S. A.* 78 (1981) 1115–1118.
- [26] X. Luo, H.H. Hsiao, M. Bubunenko, G. Weber, D.L. Court, M.E. Gottesman, H. Urlaub, M.C. Wahl, Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex, *Mol. Cell* 32 (2008) 791–802.
- [27] D. Singh, S.J. Chang, P.H. Lin, O.V. Averina, V.R. Kabardin, S. Lin-Chao, Regulation of ribonuclease E activity by the L4 ribosomal protein of *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 864–869.
- [28] Z. Wang, J. Cotney, G.S. Shadel, Human mitochondrial ribosomal protein MRPL12 interacts directly with mitochondrial RNA polymerase to modulate mitochondrial gene expression, *J. Biol. Chem.* 282 (2007) 12610–12618.
- [29] S. Chaudhuri, K. Vyas, P. Kapasi, A.A. Komar, J.D. Dinman, S. Barik, B. Mazumder, Human ribosomal protein L13a is dispensable for canonical ribosome function but indispensable for efficient rRNA methylation, *Rna* 13 (2007) 2224–2237.
- [30] M.D. Dabeva, J.R. Warner, Ribosomal protein L32 of *Saccharomyces cerevisiae* regulates both splicing and translation of its own transcript, *J. Biol. Chem.* 268 (1993) 19669–19674.
- [31] A.V. Ivanov, A.A. Malygin, G.G. Karpova, Human ribosomal protein S26 suppresses the splicing of its pre-mRNA, *Biochim. Biophys. Acta* 1727 (2005) 134–140.
- [32] A.A. Malygin, N.M. Parakhnevitch, A.V. Ivanov, I.C. Eperon, G.G. Karpova, Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism, *Nucleic Acids Res.* 35 (2007) 6414–6423.
- [33] Z. Wang, C.B. Burge, Splicing regulation: from a parts list of regulatory elements to an integrated splicing code, *RNA* 14 (2008) 802–813.
- [34] D.M. Mauger, C. Lin, M.A. Garcia-Blanco, hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc, *Mol. Cell. Biol.* 28 (2008) 5403–5419.

Bacteriophage-Resistant *Staphylococcus aureus* Mutant Confers Broad Immunity against Staphylococcal Infection in Mice

Rosanna Capparelli¹, Nunzia Nocerino¹, Rosa Lanzetta², Alba Silipo², Angela Amoresano², Chiara Giangrande², Karsten Becker³, Giuseppe Blaiotta⁴, Antonio Evidente⁴, Alessio Cimmino⁴, Marco Iannaccone¹, Marianna Parlato¹, Chiara Medaglia⁴, Sante Roperto⁵, Franco Roperto⁵, Luigi Ramunno⁴, Domenico Iannelli^{4*}

1 Faculty of Biotechnology, University of Naples, Portici, Naples, Italy, **2** Department of Organic Chemistry and Biochemistry, University of Naples, Naples, Italy, **3** Universitätsklinikum Münster Institut für Medizinische Mikrobiologie, Münster, Germany, **4** School of Agriculture, University of Naples, Portici, Naples, Italy, **5** Department of Pathology and Animal Health, University of Naples, Naples, Italy

Abstract

In the presence of a bacteriophage (a bacteria-attacking virus) resistance is clearly beneficial to the bacteria. As expected in such conditions, resistant bacteria emerge rapidly. However, in the absence of the phage, resistant bacteria often display reduced fitness, compared to their sensitive counterparts. The present study explored the fitness cost associated with phage-resistance as an opportunity to isolate an attenuated strain of *S. aureus*. The phage-resistant strain A172 was isolated from the phage-sensitive strain A170 in the presence of the M^{5a} phage. Acquisition of phage-resistance altered several properties of A172, causing reduced growth rate, under-expression of numerous genes and production of capsular polysaccharide. *In vivo*, A172 modulated the transcription of the TNF- α , IFN- γ and IL-1 β genes and, given intramuscularly, protected mice from a lethal dose of A170 (18/20). The heat-killed vaccine also afforded protection from heterologous methicillin-resistant *S. aureus* (MRSA) (8/10 mice) or vancomycin-intermediate *S. aureus* (VISA) (9/10 mice). The same vaccine was also effective when administered as an aerosol. Anti-A172 mouse antibodies, in the dose of 10 μ l/mouse, protected the animals (10/10, in two independent experiments) from a lethal dose of A170. Consisting predominantly of the sugars glucose and galactose, the capsular polysaccharide of A172, given in the dose of 25 μ g/mouse, also protected the mice (20/20) from a lethal dose of A170. The above results demonstrate that selection for phage-resistance can facilitate bacterial vaccine preparation.

Citation: Capparelli R, Nocerino N, Lanzetta R, Silipo A, Amoresano A, et al. (2010) Bacteriophage-Resistant *Staphylococcus aureus* Mutant Confers Broad Immunity against Staphylococcal Infection in Mice. PLoS ONE 5(7): e11720. doi:10.1371/journal.pone.0011720

Editor: Frank R. DeLeo, National Institute of Allergy and Infectious Diseases, National Institutes of Health, United States of America

Received: April 9, 2010; **Accepted:** June 29, 2010; **Published:** July 22, 2010

Copyright: © 2010 Capparelli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The funder (Regione Campania) had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: iannelli@unina.it

Introduction

Staphylococcus aureus can cause minor infections as well as life-threatening diseases. Endocarditis, osteomyelitis, pneumonia, surgical wound and intravascular device infections caused by *S. aureus* represent a major public health concern [1–3]. In the United States, about 60% of nosocomial *S. aureus* bloodstream infections and 40–60% of surgical wound infections are caused by methicillin-resistant strains of *S. aureus* (MRSA) [3]. One of these MRSA strains has been reported to cause an alarming number of necrotizing fasciitis cases [3]. Strains of *S. aureus* with reduced susceptibility to vancomycin are also emerging [2,4]. In the wake of the rising antimicrobial resistance [5], new strategies for the control of *S. aureus* infections are needed. This study describes the development of a vaccine active against *S. aureus*.

In the presence of an antibiotic or a bacteriophage (a bacteria-attacking virus), resistance is clearly beneficial to the bacteria. As might be expected, in such conditions, resistant bacteria are swift to emerge [6]. However, *in vivo* and *in vitro* experiments

demonstrate that, in the absence of the antibiotic or phage, resistant bacteria often display reduced fitness compared to their sensitive counterparts [6–8]. This fitness cost is particularly high in the case of phage-resistant bacterial strains. To infect bacteria, phages often select an essential component of the bacterial cell wall as a receptor [8]. To gain resistance, bacteria must dispose of this component or alter its conformation [8]. However, this strategy is costly to bacteria, which often become less virulent or avirulent. The present study explored the cost of phage resistance to isolate an attenuated strain of *S. aureus*. In a mouse model of infection, this phage-resistant strain prevented infections from diverse and clinically relevant *S. aureus* strains.

Materials and Methods

Bacteria

S. aureus strains A170, A177, A179 and A181 were isolated from clinical samples collected from patients with infected surgical wounds (one of the patients had diabetes; one wound from a car

incident; the remaining two were patients who underwent abdominal surgery). Patients were hospitalized at the Medical School of the University of Naples Federico II. Clinical samples were streaked on trypticase soy agar (TSA) (Oxoid, Milan, Italy) and single colonies expanded in trypticase soy broth (TSB) (Oxoid). The strains listed above and their phage-resistant mutants (A172, A178, A180, A182) were all identified as *S. aureus* [9]. For in vivo and in vitro experiments, bacteria were grown in TSB at 37°C, harvested while in exponential growth phase (OD₆₀₀: 1.5 to 1.8), centrifuged ($8 \times 10^3 \times g$ for 10 min) and washed with saline (0.15 M NaCl). Serial 10-fold dilutions in saline were then plated on TSA.

Mice

Experiments were carried out on female Balb/c mice (aged 8 to 10 weeks) at the animal facility of the University of Naples. Mice were infected via the intramuscular or aerosol routes. In the case of the intramuscular infection, mice received 5×10^6 – 10^8 bacteria (as detailed in each experiment) suspended in sterile saline (100 µl/mouse). In the case of aerosol infection, mice were placed in a nebulizing chamber and exposed to an aerosol solution (10^7 CFU/ml) for 10 min. When inspected on the day of infection, the lungs displayed about $1.6 \times 10^5 \pm 5.5 \times 10^3$ CFU/g/mouse (average of 3 mice). Organs were dissected and weighed. Samples were homogenized in 1 ml saline and serially diluted in saline. Colony forming units (CFU) were evaluated by the plate count assay and expressed as CFU/g. Animal experiments were approved by the Animal Care Committee of the University of Naples (permit number 86/609/EEC).

Isolation of the phage-resistant strains

Phage M^{Sa} was isolated from the *S. aureus* strain A170 following induction with mitomycin [10]. Phages M^{Sa1}, M^{Sa2} and M^{Sa3} were isolated from *S. aureus* strains A177, A179 and A181, respectively; phage release was again induced with mitomycin [10]. Phage-resistant *S. aureus* strains A172, A178, A180 and A182 were isolated by plating dilutions of overnight susceptible cultures on TSA containing increasing concentrations of the phage used for selection. Ten single colonies growing at the highest phage concentration were selected and subcultured twice on TSA agar in the absence of phage. To ensure stability of resistance, two colonies from each phage-resistant strain were further subcultured (20 times or more) in the absence of phage. Induction of the phage-resistant strains (including A172) with mitomycin excluded the presence of prophages in these strains.

Titration of anti-A172 antibodies

Mice were immunized with the A172 strain (10^8 CFU/mouse) and two weeks later sacrificed and the blood pooled. The protein A gene is under-expressed in the A172 strain. Yet, to avoid interference with the protein A possibly present on the bacterial surface, the wells of a 96-well plate (Falcon, Milan) were coated with the *S. aureus* protein A negative strain Wood 46 [11], quenched with 3% bovine serum albumin (50 µl/well; 2 h), washed with PBS, incubated with anti-A172 serum diluted (10^{-2} – 10^{-4}) in PBS (50 µl/well; 2 h), washed with PBS and incubated, in succession, with peroxidase-labelled rat anti mouse IgG (50 µl/well; 2h) and peroxidase substrate (100 µl/well; Bio-Rad, Milan).

Carbohydrate analysis

Teichoic acids from the A170 (A170^{TA}) and A172 (A172^{TA}) isolates were prepared as described [12] and analyzed gas chromatography-combined mass spectrometry. Monosaccharides

were identified as acetylated O-methyl glycosides derivatives, fatty acids or O-methyl ester derivatives. After methanolysis with methanolic HCl (2M HCl/CH₃OH; 85°C, 24 h) and acetylation with acetic anhydride in pyridine (85°C; 20 min), samples were analyzed by gas chromatography-combined mass spectrometry (GC-MS) and compared with standards. GC-MS analyses were carried out on a Hewlett-Packard 5890 instrument using SPB-5 capillary column (0.25×3 m; Supelco, PA). Ring size and attachment points were determined by methylation analysis as described [13]. The sample (1–2 mg) was methylated with CH₃I/NaOH in DMSO, hydrolysed with 2M trifluoroacetic acid (100°C, for 4 h), reduced with NaBD₄, acetylated with acetic anhydride in pyridine and analyzed by GC-MS. The temperature programme was: 150°C for 5 min and then 5°C/min to 300°C over 10 min.

Capsular polysaccharide purification

The A172 strain (100 ml) was grown in TSB. When in the exponential growth phase, the culture was centrifuged ($8 \times 10^6 \times g$ for 10 min) and the supernatant precipitated with ethanol (four volumes) in the cold (–20°C overnight). Supernatant and precipitate were separated by centrifugation ($8 \times 10^6 \times g$ for 10 min). The precipitate, insoluble in H₂O and several organic solvents (methanol, ethanol, chloroform, or dimethyl sulfoxide), was discarded. The supernatant was dialyzed (cut-off of the dialysis tube: 3500 Daltons) against water (3 days), lyophilised, weighed and tested for the capacity to protect mice. Mice vaccinated with the supernatant (25 µg/mouse) were fully protected against a lethal

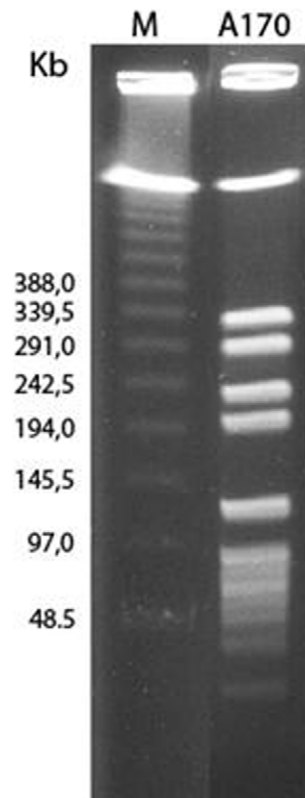
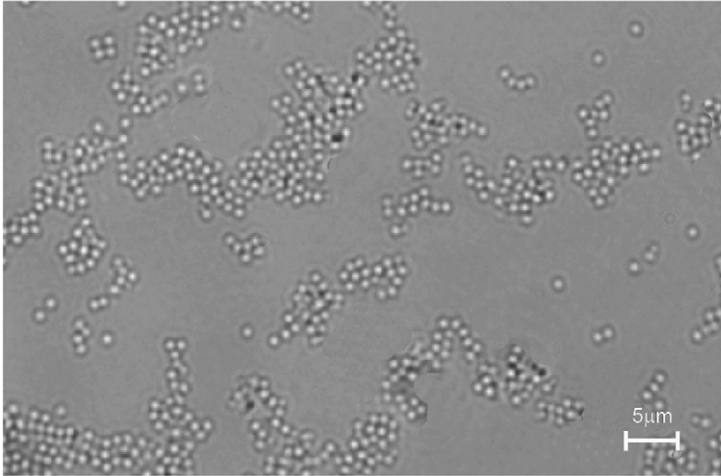


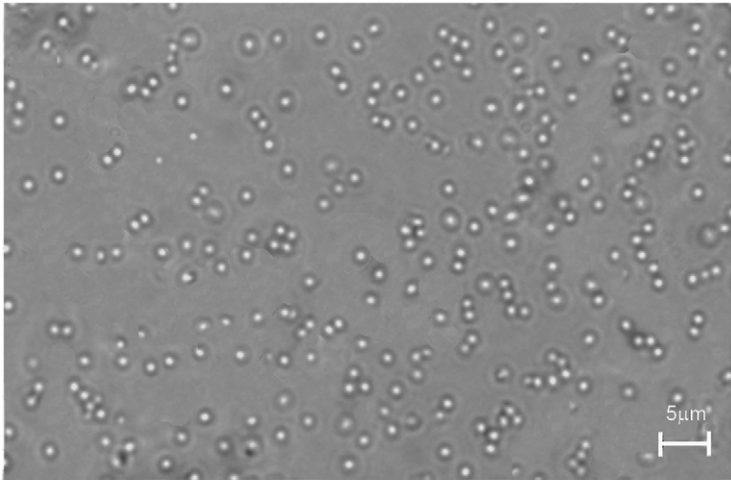
Figure 1. Pulsed field electrophoresis pattern of *Sma* I digests from the *Staphylococcus aureus* A170 strain. The A172 strain (not shown) displayed an identical pattern. M: DNA Size Standard, Lambda Ladder (Concatemers of λ cl857 Sam7) (Bio-Rad Laboratories, Hercules, CA).

doi:10.1371/journal.pone.0011720.g001

A



B



C

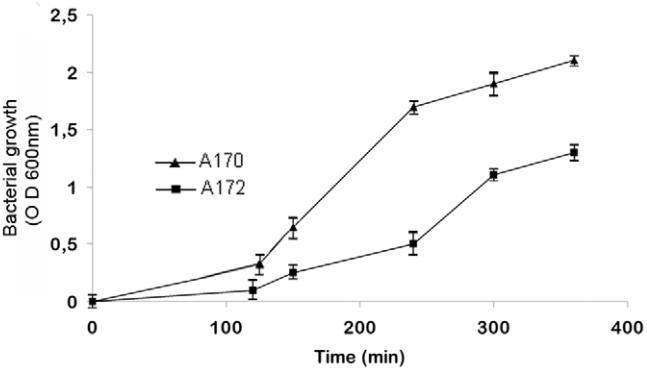


Figure 2. Differences between the phage M^{Sa}-sensitive strain A170 and the phage M^{Sa}-resistant strain A172. The strain A172 grows in larger clumps (A) compared to the strain A170 (B). Bacteria were grown in liquid culture and collected for microscopic examination at the early stationary phase. (C) The A172 strains displays also a slower growth rate, compared to A170. doi:10.1371/journal.pone.0011720.g002

dose of A170 bacteria. This biologically active fraction (3.4 g) was hydrolysed with 0.1 M HCl at 100°C for 48 h. The mixture was neutralized with 2 M NaOH and reduced with sodium borohydride in distilled water (3 g/0.5 ml). The alditols were acetylated with pyridine (100 µl) and acetic anhydride (100 µl) at 85°C for 30 min. The monoses were identified as acetate alditols by gas liquid chromatography (GLC) [14] using the Agilent GC-MS 7890A instrument (Waghaeusel-Wiesental, Germany) equipped with a capillary column (30 m×0.25 mm, 0.25 µm film thickness) of SPB-5 (Sigma Aldrich, Milan, Italy) and applying a temperature gradient of 150°C (3 min) to 320°C at 3°C/min.

Detection of A172 revertants

Individual wells of 96-well plates were filled with 200 µl TSB medium and inoculated with one colony of A172 bacteria isolated from mice vaccinated with the same strain and then grown in vitro for 15 generations under stress conditions (10 µg/ml novobiocin) [15]. Bacteria were grown for 3 h at 37°C. Individual cultures (10⁹ CFU/200 µl) were quantitatively spread over a phage lawn (10¹² PFU/5 ml soft agar). Bacteria were thus surrounded by large numbers of virus particles. Plates were inspected for the presence of plaques of lysis during the next 4 days. The detection limit of the test was established by carrying out the test on the same A172 cultures spiked with 12, 6, 3 or 1 A170 M^{Sa}-sensitive bacteria. The dilutions with 12, 6, 3 or 1 A170 bacteria were assayed by plate count in triplicate.

The presence of revertants was also investigated by an independent method, based on the increase of phage titre of a liquid culture upon addition of phage-sensitive bacteria [16]. A172 bacteria grown in the presence of novobiocin (10⁹ CFU/200 µl) were incubated for 1 h with the M^{Sa} phage (10⁵ PFU). The cultures were then transferred quantitatively on a lawn of the A170 M^{Sa}-sensitive bacteria (10⁵ CFU/4 ml soft agar) to determine whether the original phage titre increased. The detection limit of the test was established by testing in parallel the phage titre increase that occurred in the same A172 cultures (10⁹ CFU/200 µl) spiked with 12, 6, 3 or 1 A170 M^{Sa}-sensitive bacteria. The dilutions with 12, 6, 3 or 1 A170 bacteria were assayed as described above.

Phage lysis inhibition

A170 or A172 bacteria (10⁵ CFU/100 µl) were grown (4 h) in 2 ml TSB, in the presence or absence of: 5–20 mM N-acetylglucosamine (GlcNAc), 5–20 mM glucose (Sigma, Milan), 0.1–4 nmol/tube teichoic acid from A170 (A170^{TA}) or A172 (A172^{TA}). The M^{Sa} phage (10⁹ PFU/tube) was then added and bacterial growth (OD₆₀₀) measured 30 min later.

Phage inactivation by N-acetyl-glucosaminidase

A170 or A172 bacteria (10⁵ CFU/100 µl) were grown (4 h) in 2 ml TSB, in the presence or absence of N-acetyl-glucosaminidase from *Canavalia ensiformis* (Sigma; 4 U/tube). The M^{Sa} phage was then added (10⁹ PFU) and bacterial growth (OD₆₀₀) measured 30 min later.

Pulsed-field electrophoresis of *S. aureus* strains

The procedure adopted was that described in [17]. Briefly, inserts of intact DNA were digested in 200 µl of appropriate buffer

supplemented with 40 U of *Sma* I (Promega, Milan). Pulsed field gel electrophoresis (PFGE) of the restriction digests was performed by using the CHEF system (Bio-Rad Laboratories, Hercules, CA, USA) with 1% (wt/vol) agarose gels and 0.5×TBE as running buffer, at 10°C. Restriction fragments were resolved in a single run, at constant voltage of 6 V cm² and an orientation angle of 120° between electric fields, by a single phase procedure for 24 h with a pulse ramping between 1 and 50s.

Other methods

The transcription level of bacterial virulence genes and mouse cytokine genes was analyzed by real-time reverse transcription PCR (RT-PCR) as described [18]. Fluorescein isothiocyanate labelled A170 (A170^{FITC}) and A172 (A172^{FITC}) bacteria were prepared as described [19]. Troponin was measured using the Mini Vidas automated immunoassay analyzer and the Vidas troponin kit (bioMereux, Florence, Italy). The sequence type of the *S. aureus* strains were determined by sequencing the hypervariable region of *S. aureus* protein A gene (*spa*) as described [20]. The *spa* types were assigned using the Ridom *spa*-server (<http://spaserver.ridom.de>). Multilocus sequence typing (MLST) of *S. aureus* strains was carried out as described [21]. The bacterial genes *eta*, *etb*, *tst*, *lukS-PV-lukF-PV*, *lukE-lukD* and *lukM*, (coding for the exfoliative toxin A, exfoliative toxin B, toxic shock syndrome toxin 1, Pantone-Valentine leukocidin components S and F, the leukotoxin LukELukD, and the leukotoxin LukM, respectively) were detected by PCR [10]. The staphylococcal enterotoxin (*se*) genes *se-a*, *se-b*, *se-c*, *se-d*, and *se-e* were detected as described [22]. The restriction endonucleases analysis-enterotoxin gene cluster (REA-*egc*) type was carried out as described [23]. Survival rates of mice were analyzed using Fisher's exact test. Bacterial counts and gene expression levels were analyzed using the

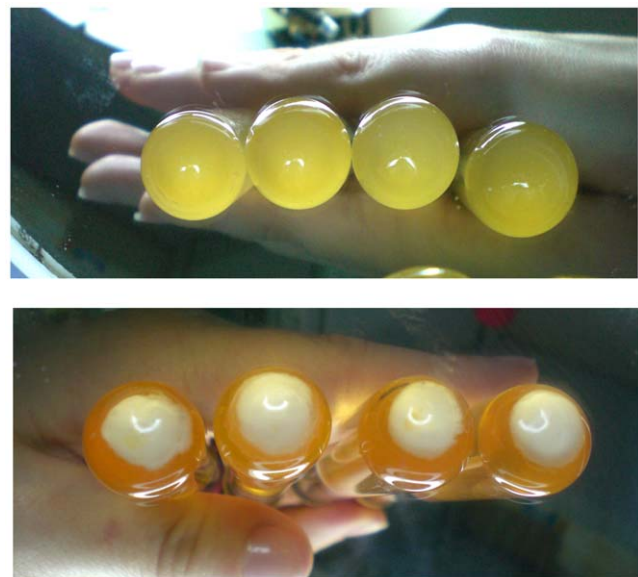


Figure 3. In *S. aureus*, phage-resistance comes with production of capsular polysaccharide. (A) Strains A170, A177, A179, A181 (sensitive to the phages M^{Sa}, M^{Sa1}, M^{Sa2}, M^{Sa3}, respectively). (B) Strains A172, A178, A180, A182 (resistant to the phages M^{Sa}, M^{Sa1}, M^{Sa2}, M^{Sa3}, respectively). doi:10.1371/journal.pone.0011720.g003

paired t test (P values are two-tailed values). The molar concentration of the teichoic acids from the *S. aureus* strains A170 and A172 were calculated attributing to A170 and A172 the molecular weight of 6350 daltons (the average molecular weight of the teichoic acid from *S. aureus* strain Copenhagen, which ranges from 5400 to 7300 daltons) [24]

Results

Isolation and characterization of A172

The strains A170 and A172 were positive for the exfoliative toxin B, the Pantone-Valentine leukocidin, while negative for the exfoliative toxin A, toxic shock syndrome toxin 1, the leukotoxin LukELukD, the leukotoxin LukM and the staphylococcal enterotoxins-a, -b, -c, -d and -e. The strains A170 and A172 displayed also the same *egc* type (*egc*-4), the multilocus sequence type (ST 45) and *spa*-type (t6668) (*spa*-type repeats succession: 08-16-02-16-34-13-17-34-34-34-34-34, equivalent to Kreiswirth's IDs: XKAK-BEMBBBBB). According to the Ridom *spa*-server (<http://spaserver.ridom.de>), t6668 represents a novel *spa* type. PFGE analysis displayed a close genetic relation between the parental phage M^{Sa}-sensitive strain A170 and the derivative phage-resistant strain A172, as proved by the identical PFGE pattern of the *Sma* I digests (Figure 1). A172 was isolated by growing the parental strain A170 in the presence of increasing concentrations of the M^{Sa} phage [10]. In addition to M^{Sa}, the strain A172 is also resistant to phages M^{Sa1}, M^{Sa2} and M^{Sa3}. Prolonged liquid subculture for several months in the absence of phage or prolonged storage at -80°C did not alter the resistance of A172 to M^{Sa}.

The strain was also highly stable in vivo. A172 bacteria were isolated from a vaccinated mouse, grown for 15 generations under stress conditions (10 $\mu\text{g}/\text{ml}$ novobiocin) [15] and used to establish 400 independent bacterial cultures, which yielded a total number of about 4×10^{11} bacteria (10^9 , the number of cells/culture $\times 400$, the number of cultures). Plated with phage in excess (10^{12} PFU/5 ml soft agar), the 400 independent bacterial cultures did not yield any plaque of lysis. Control cultures (10^9 CFU A172/200 μl spiked with 3 A170 bacteria), assayed in triplicate in three independent experiments, displayed all at least one plaque and, on

average, 2.2 ± 0.7 plaques/culture (number of replicas: 9). The absence of A172 revertants was confirmed in a second independent experiment. When the M^{Sa} phage (10^5 PFU) was mixed with the A172 bacteria grown under stress conditions (10^9 CFU/200 μl), no measurable increase in the phage titre occurred. Control cultures (10^9 CFU A172/200 μl spiked with 3 A170 bacteria), compared to the non-spiked cultures, displayed an increase in phage titre from $1.4 \times 10^5 \pm 24 \times 10^3$ to $1.2 \times 10^6 \pm 0.5 \times 10^3$ (n: 9; P: 0.0057). Conservatively, it is estimated that A172 revertants occur at a frequency $< 3 \times 10^{-9}$.

In liquid culture, compared to the parental strain, A172 formed colonies with the tendency to clump together starting at the early stationary phase (Figure 2A–B). Bacterial suspensions of A170 and A172 with the same density (OD₆₀₀: 0.6) were assayed by plate count. In three independent experiments, A170 cultures contained about 3 fold as many CFU as the A172 cultures ($3.4 \times 10^7 \pm 6.7 \times 10^6$ vs $1.2 \times 10^7 \pm 2.5 \times 10^6$; P 0.003). A172 displayed reduced growth rate during the exponential phase (Figure 2C) and reduced doubling time (A172: 45 min \pm 2.8 min; A170: 30 \pm 2.5 min; P 0.002). Also, unlike A170, A172 produced capsular polysaccharide (Figure 3). Phage resistance altered the transcription of several virulence factors. Among 14 ORFs of A172 analyzed, 13 were significantly down-regulated compared to A170 (Figure 4). Down-regulated ORFs comprise 9 genes, which control virulence factors [25].

The receptor of the M^{Sa} phage is the N-acetylglucosamine

Phages often use teichoic acids for adsorption on the cell wall of Gram positive bacteria. They can use as receptor the glucose side chains of *Bacillus subtilis* teichoic acids [26,27] or the N-acetylglucosamine (GlcNAc) side chains of *S. aureus* [28]. To identify the receptor site of the M^{Sa} phage, the A170 and A172 strains were grown in the presence or absence of 5–20 mM GlcNAc or 5–20 mM glucose. GlcNAc inhibited the lysis of A170 by phage M^{Sa}, while glucose did not. GlcNAc inhibition was dose-dependent (Figure 5A–B). The experiment was repeated using the teichoic acid from (A170^{TA}) or from (A172^{TA}) as inhibitor. Since

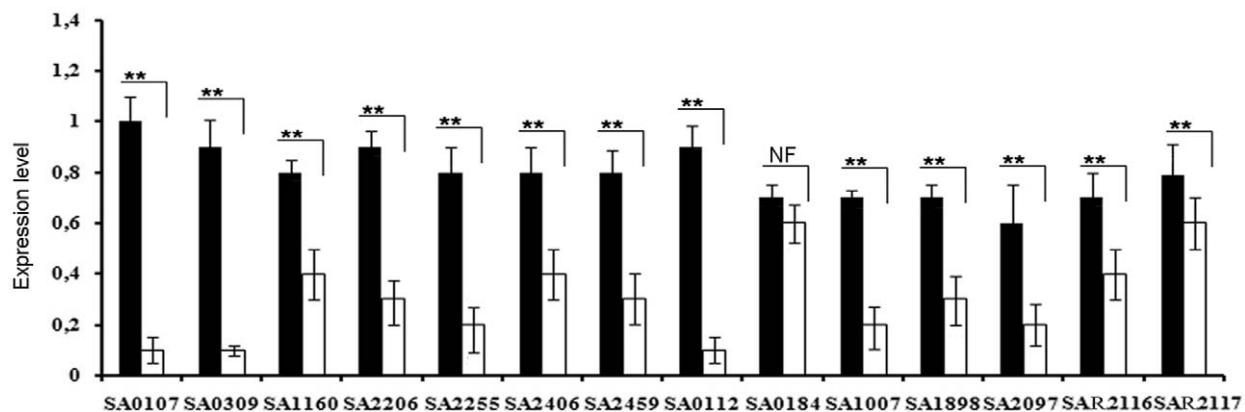


Figure 4. Expression levels of *S. aureus* A170 and A172 virulence factors. Acquisition of phage-resistance by A172 is accompanied by extensive alterations in virulence gene expression levels. Out of the 14 ORFs examined, 13 are significantly under-expressed, compared to the phage-sensitive strain A170. Bacteria were collected for transcriptional analysis during the exponential growth phase. SA0107 (*spa*, IgG binding protein A precursor); SA0112 (hypothetical protein, similar to cysteine synthase); SA0184 (hypothetical protein); SA0309 (*geh*; glycerol ester hydrolase); SA1007 (α -haemolysin); SA1160 (*nuc*; thermonuclease); SA1898 (hypothetical protein, similar to SceD precursor); SA2097 (hypothetical protein, similar to secretory antigen precursor SsaA); SA2206 (*sbi*; IgG-binding protein SBI); SA2255 (oligopeptide transporter substrate binding protein); SA2406 (*gbsA*; glycine betaine aldehyde dehydrogenase *gbsA*); SA2459 (*ica*; N-glycosyltransferase); SAR2116 (*groEL*; chaperonin GROEL); SAR2117 (*groES*; co-chaperonin GRES). ORFs numbers correspond to the *S. aureus* N315 genome sequence. Gene designations (when known) and proteins function are shown in parentheses.

doi:10.1371/journal.pone.0011720.g004

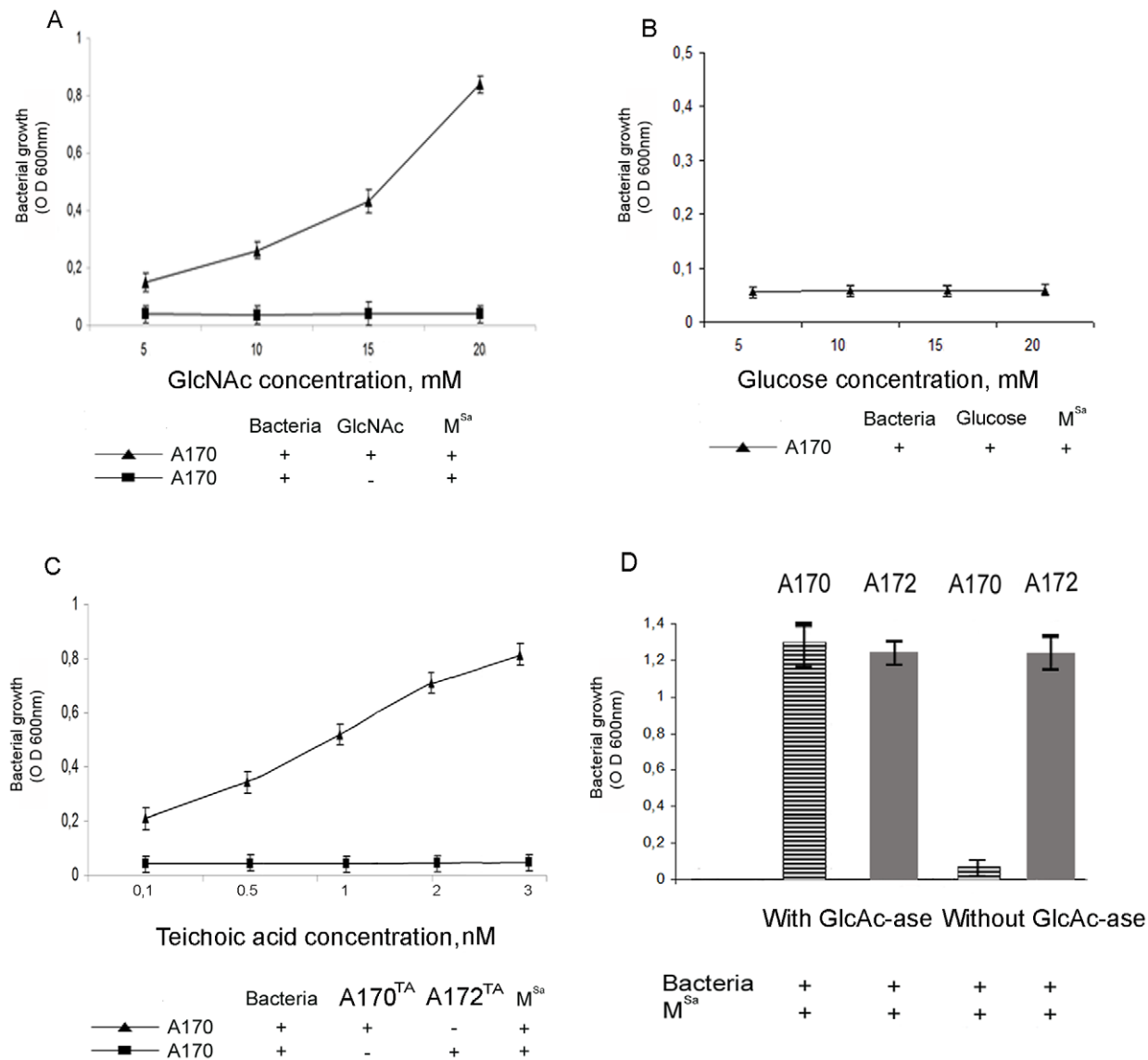


Figure 5. Phage M^{SA} is inhibited by N-acetyl-glucosamine (GlcNAc), the teichoic acid from A170 (A170^{TA}) or by treatment with N-acetyl-glucosaminidase (GlcNA-ase) from *Canavalia ensiformis*; phage M^{SA} is not inhibited by the teichoic acid from A172 (A172^{TA}) or glucose. (A) GlcNAc inhibits the lysis of the A170 strain by the phage M^{Sa}, while glucose (B) does not. (C) A170^{TA} inhibits the lysis of the A170 strain by the phage M^{Sa}, while A172^{TA} does not. (D) The phage-sensitive strain A170 grows in the presence of the M^{Sa} phage, if pre-treated with GlcAc-ase (4 U/tube; 2 h at 37°C).

doi:10.1371/journal.pone.0011720.g005

the quantity of teichoic acid that is isolated from the A170 or A172 strains varies significantly (2.4 mg from 10¹⁰ CFU A170; 0.9 mg from 10¹⁰ CFU A172), the experiment was conducted using an ample concentration range of the A170^{TA} and A172^{TA} reagents. Under these conditions, phage inactivation occurred with A170^{TA}, but not with A172^{TA} (Figure 5C). A170 bacteria were then treated with N-acetylglucosaminidase (4 U/tube; 2 h at 37°C). Enzymatic hydrolysis of GlcNAc from the A170 bacteria destroyed phage lysis capacity (Figure 5D). Finally, analysis of teichoic acids from the A170 (wild) and A172 (mutant) strains of *S. aureus* displayed the absence of ribitol (Figure 6A) and terminal GlcNAc (t-GlcNAc) residues (Figure 6B) from A172^{TA}. Collectively, the experiments described above demonstrate that A172 conforms with the tendency of several bacterial species - *B. subtilis* [26,27], *S. aureus* [28], *Salmonella enterica* [18,29] - to acquire phage-resistance by altering a cell wall polysaccharide component.

The A172 strain prevents *S. aureus* infection

For a more succinct presentation of the results, the distinctive features of the experiments described in this and the following sections are detailed in the Tables. One group of mice was immunized with A172 live (A172-L) and two weeks later infected with a lethal dose of A170. At the time of infection, in the serum of immunized animals the titre of the antibodies against A172-L ranged from 1:5000 to 1:12000. Control mice were also infected with a lethal dose of A170. The A172-L vaccine protected 90% of the mice (surviving mice: 18/20) from A170, while the control mice (10/10) died within 4 days (P: 0.008; Table 1, experiment 1). When examined at the end of the experiment (14 days after challenge), the kidneys of surviving A172-L treated animals displayed significantly fewer CFU compared to the kidneys of the control mice at the time of death (Table 1, experiment 1).

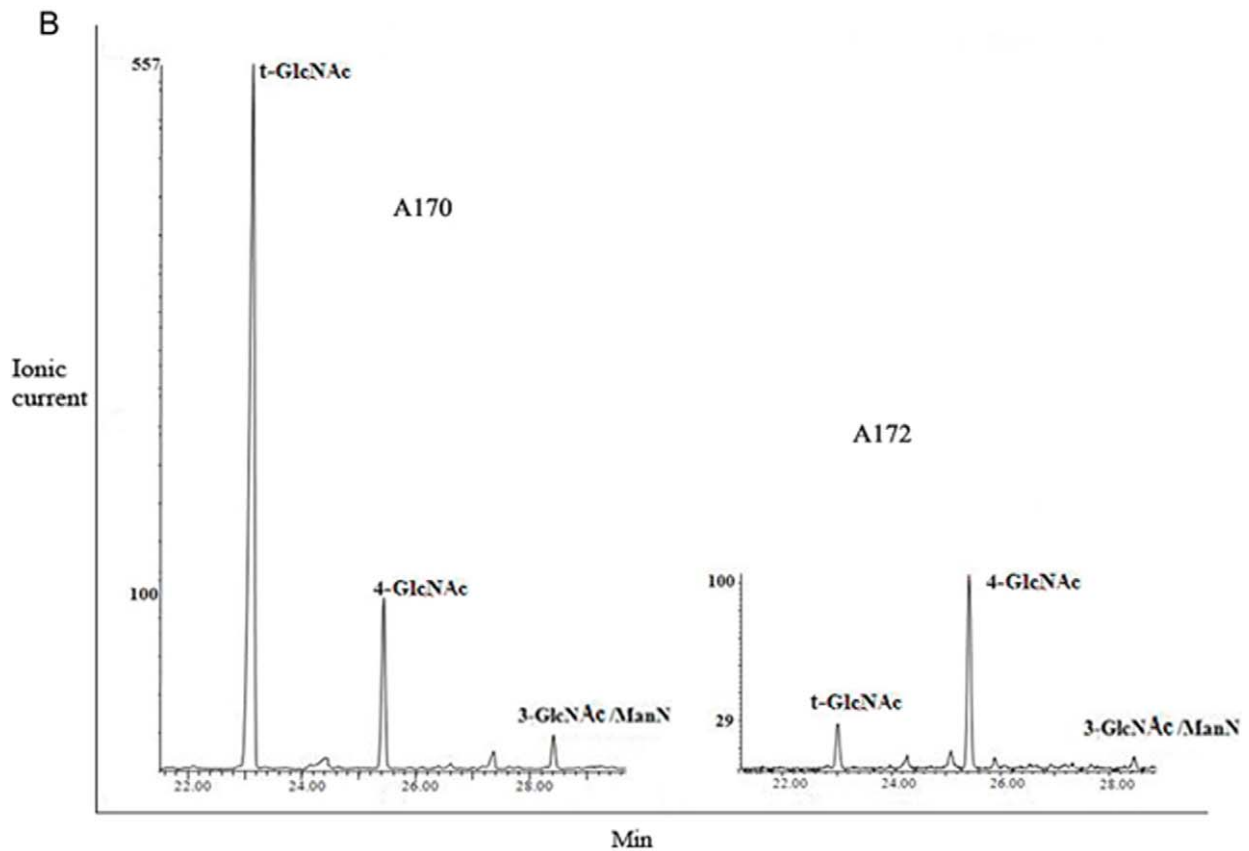
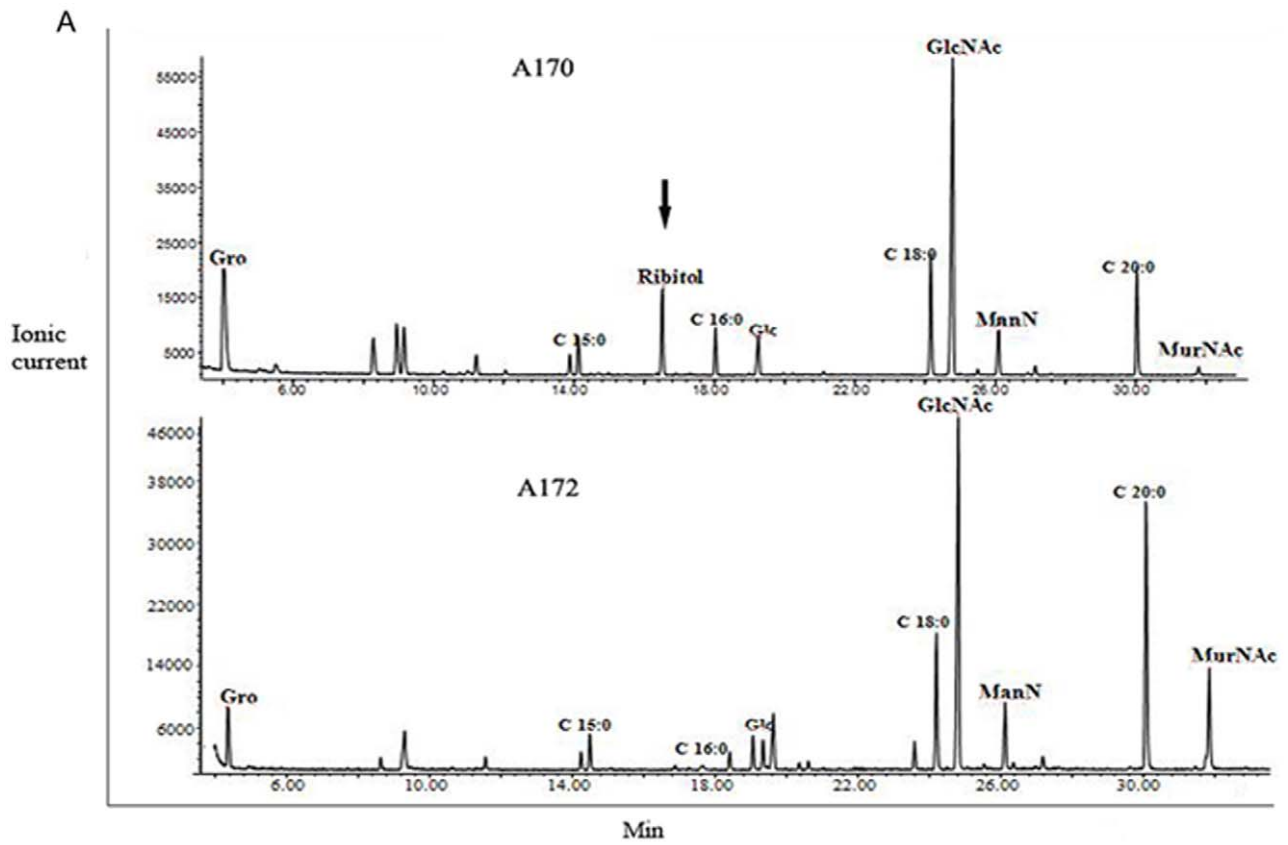


Figure 6. Gas chromatography-combined mass spectrometry spectra of the teichoic acids from the A170 and A172 strains. Acetylated O-methyl glycosides from the phage-sensitive A170 and phage-resistant A172 strains of *S. aureus*. Loss of ribitol (A) and t-GlcNAc linked to ribitol (B) in the teichoic acids is part of the strategy adopted by A172 to gain phage-resistance. The peak of 4-GlcNAc was chosen as reference. Gro: glycerol; GlcNAc: N-acetylglucosamine; ManN: N-acetylmannosamine; MurNAc: N-acetyl muramic acid; Glc: glucose; C15:0: pentadecanoic acid; C16:0: esadecanoic acid; C18:0: octadecanoic acid; C20:0: eicosanoic acid; 4-GlcNAc: 4-linked N-acetylglucosamine; 3GlcAcNMaN: 3-linked N-acetylmannosamine; t-GlcNAc: terminal GlcNAc.
doi:10.1371/journal.pone.0011720.g006

To exclude that the above results represented the response elicited by a sublethal infection with live *S. aureus*, mice were infected with a sublethal dose of A170 and two weeks later challenged with a lethal dose of A170. There was no significant difference in surviving between immunized and control mice (P: 1; Table 1, experiment 2). Thus, the protective response described above is to be ascribed specifically to A172.

Being a live vaccine, A172 poses the risk of reverting to the virulent phenotype. This risk is remote (see the section “Isolation and morphological characterization of A172”), albeit considerable when the use of a large number of vaccine doses is prospected. The study therefore explored whether A172 remained protective once it was heat-killed (15 min at 100°C). Mice were vaccinated with heat-killed vaccine (A172-HK) and 2 weeks later infected with a lethal dose of A170. The vaccine was protective also in this form (P: 0.004; Table 1, experiment 3).

When the interval between vaccination and challenge was extended to 20 weeks, control mice all died within 4 days (10/10); the vaccinated mice (37/40) were alive 14 days after challenge (P: 0.004; Table 1, experiment 4). Thus, the protection provided by the vaccine lasts at least 20 weeks.

One more experiment was designed to ascertain whether also A170 - in the heat-killed form (A170-HK; 15 min at 100°C)- was protective. No significant difference in survival between immunized and control mice was observed (P: 1; Table 2, experiment 5).

Additional properties of the A172 vaccine

The efficacy of the A172-HK vaccine was tested against two methicillin-resistant *S. aureus* (MRSA) (A174 and A175) and one

vancomycin-intermediate *S. aureus* (VISA) (A176) isolates derived from patients with staphylococcal infections. Three groups of mice were immunized with A172-HK and 20 weeks later infected with a lethal dose of the A174, A175 or A176 isolates. The difference in survival between vaccinated and control mice was statistically significant in each case (P: 0.01–0.02; Table 2).

Patients with cystic fibrosis are highly susceptible to *S. aureus* lung infection [30]. *S. aureus* is also one of the most common causes of pneumonia in paediatric and adult populations [31]. The study therefore explored the efficacy of A172-HK against lung infections. Mice were immunized by aerosol with A172-HK and 20 weeks later infected with A170 by the same route. The vaccine was also effective when administered by aerosol (P: 0.004; Table 3).

Collectively, the experiments described in this and the preceding section demonstrate that A172-HK protects mice against a lethal dose of *S. aureus*; protection correlates with increased survival and reduced colonization lasting at least 20 weeks and extending into drug-resistant isolates of *S. aureus*; the vaccine is effective when administered, in single dose and without adjuvant, by the intramuscular or aerosol routes.

Anti-A172 antibodies protect against *S. aureus* A170 infection *in vivo*

Two groups of mice (control mice) were injected intramuscularly with normal mouse serum (10 µl or 20 µl diluted 1:10 with sterile saline, respectively). Two more groups of mice were treated intramuscularly with serum from mice immunized with A172-HK (10 µl or 20 µl diluted 1:10 with sterile saline,

Table 1. Active protection of A172 in mice challenged with *S. aureus* strain A170.

Exp	Vaccination			Interval vaccination challenge (weeks)	Challenge			Survival	P value ^b	CFU/g ^a (mean ± SD)
	Dose CFU/mouse	Vaccine	Route		Dose CFU/mouse	Pathogen	Route			
1	10 ⁸	A172-L ^c	Im ^d	2	10 ⁸	A170	Im	18/20	0.008	266 ± 11 ^e
					-	-	-	0/10		5.6 × 10 ⁸ ± 3.5 × 10 ⁸
2	5 × 10 ⁶	A170	-	2	10 ⁸	A170	Im	0/10	1	5.7 × 10 ⁸ ± 3.2 × 10 ⁸
					-	-	-	0/10		5.2 × 10 ⁸ ± 1.5 × 10 ⁸
3	10 ⁸	A172-HK ^f	-	2	10 ⁸	A170	Im	19/20	0.004	260 ± 25 ^e
					-	-	-	0/10		3.5 × 10 ⁸ ± 1.8 × 10 ⁸
4	10 ⁸	A172-HK	-	20	10 ⁸	A170	Im	37/40	0.004	200 ± 11 ^e
					-	-	-	0/10		5.7 × 10 ⁸ ± 3.3 × 10 ⁸
5	10 ⁸	A170-HK	-	2	10 ⁸	A170	Im	0/10	1	5.2 × 10 ⁸ ± 1.8 × 10 ⁸
					-	-	-	0/10		4.7 × 10 ⁸ ± 2.3 × 10 ⁸

^aColony forming units per g of kidneys (experiments 1–4) or lungs (experiment 5); values calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cA172 Live.

^dIntramuscular.

^eP < 0.001.

^fHeat killed A172.

doi:10.1371/journal.pone.0011720.t001

Table 2. Protection of heat killed vaccine A172 against drug-resistant *S. aureus* strains.

Vaccination			Interval vaccination challenge (weeks)	Challenge					
Dose (CFU/mouse)	Vaccine	Route		Dose (CFU/mouse)	Pathogen	Route	Survival	P value ^b	CFU/g ^a (mean ± SD)
10 ⁸	A172-HK ^c	Im ^d	20	10 ⁸	A174	Im	8/10	0.02	250±22 ^e
				-	-	-	0/10		6.7×10 ⁸ ±1.8×10 ⁸
10 ⁸	A172-HK	-	20	10 ⁸	A175	Im	9/10	0.01	330±37 ^e
				-	-	-	0/10		5.8×10 ⁸ ±1×10 ⁸
10 ⁸	A172-HK	-	20	10 ⁸	A176	Im	9/10	0.01	380±26 ^e
				-	-	-	0/10		5.5×10 ⁸ ±3.6×10 ⁸

^aColony forming units per g of kidneys; values calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cHeat killed A172.

^dIntramuscular.

^eP<0.001.

doi:10.1371/journal.pone.0011720.t002

respectively). The titre of the antibodies against A172-HK was 1:8000. The next day, the mice were all challenged with a lethal dose of *S. aureus* A170. Survival was monitored for seven days. In two independent experiments, the antibodies against A172-HK - given at the dose of 10 µl/mouse - provided protection to all the mice (10/10) (P 0.01; Table 4, experiments 1 and 2). Instead, the same serum - given at the dose of 20 µl/mouse - killed 40%–50% of the mice (P 0.05–0.06; Table 4, experiments 1 and 2). Thus, anti-A172 antibodies can be protective or detrimental, depending upon the dose being given. This phenomenon has been described since the development of serum therapy [32]. In mouse models of *Streptococcus pneumoniae* [32] and *Cryptococcus neoformans* [33], antibody-mediated protection was also observed within a narrow dose range. Since the therapeutic use of antibodies in humans is sometimes associated with myocardial damage [34,35], mice passively immunized with antibodies against A172-HK were tested for evidence of heart damage by measuring the level of the cardiac troponin T in the serum. This marker is highly specific for myocardial damage [34]. Mice treated with 20 µl antibodies, compared to those treated with 10 µl, displayed a statistically significant higher level of serum troponin T (0.09±0.01 ng/ml vs 0.02±0.01 ng/ml; P: 0.001). It was not investigated whether heart damage was the primary cause of death or a secondary complication of antibody administration. Mice treated with 10 µl or 20 µl antibodies displayed both significantly fewer

CFU compared to the mice treated with normal mouse serum (Table 4, experiments 1 and 2).

In an independent experiment, mice were immunized with 10 µl of anti-A172 mouse serum absorbed with rat anti total mouse immunoglobulin fraction (5–10 µl) and diluted to 100 µl with sterile saline. The next day, the mice were challenged with a lethal dose of *S. aureus* A170. Absorption of anti-A172 mouse serum completely eliminated the protective effect of the serum (Table 4, experiments 3). Collectively, the above results demonstrate that the passive protection shown by the antibodies against A172-HK is dose dependent and specific (residing in the antibody fraction of the serum).

The A172 strain controls inflammation

Ten mice were immunized intramuscularly with A172-HK (10⁸ CFU/mouse). A172 modulated transcription of the genes coding for the pro-inflammatory cytokines TNF-α, IFN-γ and IL-1β, while inducing that of the genes coding for the anti-inflammatory cytokines Il-4 and Il-6, when these genes were analyzed by RT-PCR, 24 and 48 h after immunization (Figure 7). The same pattern of cytokine production was observed in mice immunized with A172-HK and 20 weeks later infected with A170 (data not shown). Given the abundant release of pro-inflammatory cytokines during *S. aureus* infection [36–37], this property of A172 is biologically relevant.

Table 3. Protection of heat killed vaccine A172 administered by aerosol.

Vaccination			Interval vaccination challenge (weeks)	Challenge					
Dose (CFU/mouse)	Vaccine	Route		Dose (CFU/mouse)	Pathogen	Route	Survival	P value ^b	CFU/g ^a (mean ± SD)
10 ⁸	A172-HK ^c	Aer ^d	20	10 ⁸	A170	Aer	19/20	0.004	216±3 ^e
				-	-	-	0/10		7.1×10 ⁸ ±3.1×10 ⁸

^aColony forming units per g of kidneys; values calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cHeat killed A172.

^dAerosol.

^eP<0.001.

doi:10.1371/journal.pone.0011720.t003

Table 4. Mouse anti A172-HK antibodies protect mice challenged with *S. aureus* A170 strain.

Exp	Passive vaccination			Challenge			Survival	P value ^b	CFU/g ^a (mean ± SD)
	Treatment	Dose (μ l/mouse)	Route	Dose (CFU for mouse)	Strain	Route			
1	Normal mouse serum	10	Im ^c	10 ⁸	A170	Im	0/10	0.01	5.8×10 ⁸ ±1×10 ^{8d}
	Anti-A172-HK serum	-	-	-	-	-	10/10		90±7
	Normal mouse serum	20	Im	10 ⁸	A170	Im	0/10	0.05	4.7×10 ⁸ ±2×10 ^{7d}
	Anti-A172-HK serum	-	-	-	-	-	6/10		5.5×10 ⁵ ±3.6×10 ^{4d}
2	Normal mouse serum	10	Im	10 ⁸	A170	Im	0/10	0.01	3.5×10 ⁸ ±2×10 ^{8d}
	Anti-A172-HK serum	-	-	-	-	-	10/10		73±6
	Normal mouse serum	20	Im	10 ⁸	A170	Im	0/10	0.06	6.3×10 ⁸ ±1×10 ^{8d}
	Anti-A172-HK serum	-	-	-	-	-	5/10		2.6×10 ⁵ ±10 ⁵
3	Anti-A172-HK serum absorbed with rat anti-mouse IgG	10	Im	10 ⁸	A170	Im	0/10		5.4×10 ⁸ ±4.2×10 ⁷

^aColony forming units per g of kidneys; values calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cIntramuscular.

^dP<0.0001.

doi:10.1371/journal.pone.0011720.t004

Selection for phage resistance as a general approach for vaccine production

Three groups of mice were immunized (10⁸ CFU/mouse, intramuscularly) with the M^{Sa1}-resistant strain A178 (derived from the M^{Sa1}-susceptible isolate A177), M^{Sa2}-resistant strain A180 (derived from the M^{Sa2}-susceptible isolate A179) or M^{Sa3}-resistant strain A182 (derived from the M^{Sa3}-susceptible isolate A181), respectively. Two weeks later, immunized and control mice were challenged with A170 (10⁸ CFU, intramuscularly). The phage-resistant strains all provided protection in terms of survival and level of host colonization (Table 5; P: 0.01–0.02) and – like A172 – all lost t-GlcNAc (data not shown) while all gained the capacity to secrete capsular polysaccharide (Figure 3). The experiment demonstrates the general applicability of selection for phage resistance as a means to curb bacterial virulence. The experiment demonstrates also how different *S. aureus* isolates use the same defence strategy against different phages.

The capsular polysaccharide from A172 provides protection

Capsular polysaccharide (CP) production is a characteristic common to the four protective strains (A172, A178, A180, A182)

(Figure 3). In an attempt to identify the protective component of A172, the CP from this strain (CP-A172) was purified and analyzed by gas liquid chromatography (GLC). CP-A172 contains glucose and mannose in the molar ratio of 3:1 and traces (molar ratio: 0.3) of galactose and GlcN (Figure 8). Immunization of mice with CP-A172 (25 μ g/mouse) or treatment with the serum from mice immunized with CP-A172 (10 μ l diluted 1:10 with sterile saline/mouse) protected mice against a lethal dose of A170 (P: 0.004; Table 6, experiments 1–2).

Discussion

Several virulence factors of *S. aureus* are being investigated for their potential use as vaccines [37–38]. However, in view of the numerous virulence factors produced by this pathogen [39], it is unlikely that a single component vaccine might protect against a multitude of virulence factors. Also, virulence genes often display high levels of variability, as result of the selection pressure to evade the immune system of the host [40]. The single antigen may therefore have limited distribution among clinical isolates or may not be expressed in vivo. The objective of the present study was to

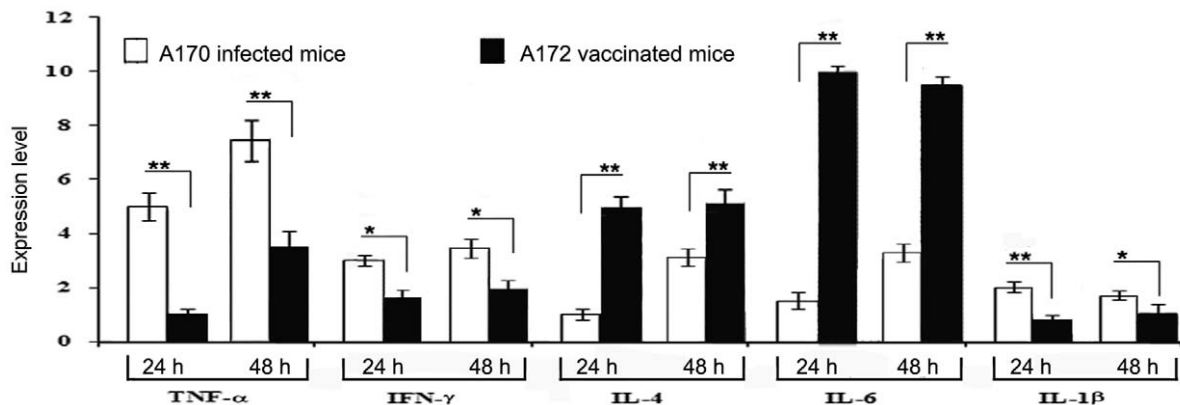


Figure 7. Anti-inflammatory activity of the A172 vaccine. A172 modulates transcription of the genes coding for pro-inflammatory cytokines (TNF- α , INF- γ and IL-1 β) and induces transcription of the genes coding for anti-inflammatory cytokines (IL-4 and IL-6). doi:10.1371/journal.pone.0011720.g007

Table 5. Selection for phage-resistance as a general approach for *S. aureus* vaccine production.

Vaccination			Interval vaccination-challenge (weeks)	Challenge			Survival	P value ^b	CFU/g ^a (mean ± SD)
Dose (CFU/mouse)	Strain	Route		Dose (CFU/mouse)	Strain	Route			
10 ⁸	A178	Im	20	10 ⁸	A177	Im	9/10	0.01	230 ± 40 ^c
			-	-	-	-	0/10		4.6 × 10 ⁸ ± 1.5 × 10 ⁸
10 ⁸	A180	-	20	10 ⁸	A179	Im	10/10	0.01	160 ± 34 ^c
			-	-	-	-	0/10		5.3 × 10 ⁸ ± 2 × 10 ⁸
10 ⁸	A182	-	20	10 ⁸	A181	Im	8/10	0.02	180 ± 25 ^c
			-	-	-	-	0/10		3.5 × 10 ⁸ ± 1.2 × 10 ⁸

^aColony forming units per g of kidneys; value calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cP < 0.001.

doi:10.1371/journal.pone.0011720.t005

develop a live attenuated *S. aureus* strain containing multiple antigens and therefore likely to elicit cross-protection against a spectrum of pathogenic strains. The attenuated strain (A172) proved effective as vaccine even when heat-killed. Thus, heat-killed A172 (A172-HK) possesses the advantages of a multiple component vaccine, without the problems of safety (reversion to phage sensitivity and the accompanying virulent phenotype), distribution and manufacturing of the live vaccines [41].

Phage-resistance often (though not always) occurs at a fitness cost [6,7]. This trade-off explains why phage-resistant and phage-sensitive bacteria coexist [42]. Bacterial selection for phage-resistance was therefore exploited as an approach to isolating an attenuated strain of *S. aureus*. The approach had already been used successfully to produce a vaccine against *Salmonella enterica* serovar Paratyphi B [18]. The phage-sensitive strain A170, following growth in vitro in the presence of the M^{Sa} phage, yielded the phage-resistant strain A172. Given their common origin, the two strains share several properties, but display also crucial differences. They share the spa-type (t6668), the MLST type (ST 45), the egc type (egc-4) and part of the genome (Figure 1). At the same time, A170 and A172 differ in their teichoic acid structure and the capacity to secrete CP. The teichoic acids of A172 lack the

terminal GlcNAc residues (Figure 6B) that phages often use for adsorption on the cell wall of *S. aureus* [28]. More importantly, GlcNAc and the teichoic acids from A170 (A170^{TA}) inhibit lysis by the M^{Sa} phage, while glucose or the teichoic acids from A172 (A172^{TA}) do not (Figure 5C). These results suggest that A172 gained phage-resistance by losing the phage adsorption site (terminal GlcNAc). However, A172 also displays CP production, a strategy that some bacteria put in action to mask their phage adsorption site [43]. In conclusion, while teichoic acid alteration is the likely mechanism adopted by A172 to gain phage resistance, the contribution of CP production cannot be excluded.

In a mouse model of infection, A172-HK elicited high levels of protection against lethal doses of the A170 strain (P 0.004; Table 1, experiment 3). A single dose of A172-HK, without use of adjuvant, was fully protective after vaccine administration by the intramuscular or aerosol routes and protection lasted at least 20 weeks (Tables 1–3). Significantly, A172-HK was effective against methicillin-resistant (MRSA) or vancomycin-intermediate clinical isolates of *S. aureus* (VISA) (Table 2). Sometime cross-protection has been attributed to the persistence of the vaccine strain within the host [44]. This possibility can be excluded in the present study since the vaccine persists in the immunized animals for only 8

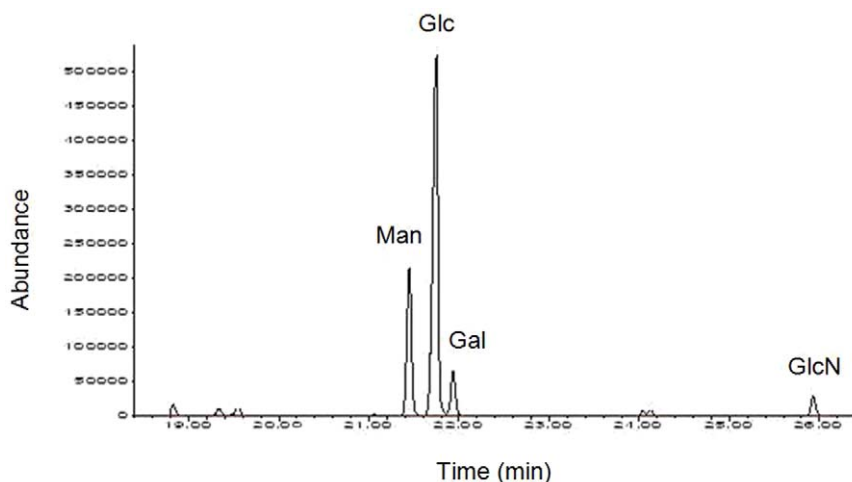


Figure 8. Gas liquid chromatography profile of monosaccharides obtained by acid hydrolysis of the capsular polysaccharide from *S. aureus* A172. Abundance expresses the relative ratio between monosaccharides.

doi:10.1371/journal.pone.0011720.g008

Table 6. Active and passive protection afforded by CPA172.

Exp	Treatment			Interval treatment challenge	Challenge			Survival	P value ^b	CFU/g ^a (mean ± SD)
	Treatment	Dose μg/mouse	Route		Dose CFU	Pathogen	Route			
1	CPA172	25	Im ^c	20 weeks	10 ⁸	A170	Im	20/20	0.004	138±27 ^d
					-	-	-	0/10		4×10 ⁸ ±10 ⁷
2	Anti-CPA172	10	Im	24 h	10 ⁸	A170	Im	20/20	0.004	205±42
	Normal mouse serum	10	-		-	-	-	0/10		3.8×10 ⁸ ±10 ⁷

^aColony forming units per g of kidneys; values calculated on 3 mice.

^bTwo tailed Fisher's exact test.

^cIntramuscular.

^dP<0.001.

doi:10.1371/journal.pone.0011720.t006

days, while mice were challenged with virulent *S. aureus* 20 weeks after vaccine administration (Table 1).

Passive immunization of mice with sera from A172-HK vaccinated mice provided protection against the challenge with a lethal dose of *S. aureus* A170. In two independent experiments, 10/10 of the animals passively immunized with an individual dose of 10 μl immune serum (diluted 1:10 with sterile saline) survived, while 10/10 untreated control mice died (P: 0.01; Table 4, experiment 1). When the experiment was repeated using a higher dose (20 μl immune serum diluted 1:10 with sterile saline/mouse), 40%–50% of the animals died (Table 5, experiment 2). These results indicate that high antibody doses can be detrimental. This possibility must be considered when interpreting passive protection experiments.

A172-HK was protective in vivo, while A170-HK was not (Table 1). This suggested that A172, along with the resistance to the M^{Sa} phage, also acquired the capacity to express a new molecule, that acts as a protective antigen. Also, the phage-resistant strains (A172, A178, A180, A182) all display the characteristic of secreting CP (Figure 3). CPs are heat stable and can confer protection [45–46]. Taken together, these facts provided sufficient ground to claim that the protective efficacy of A172 could reside in its CP (CP-A172). The results confirmed that CP-A172 provides mice with both active and passive resistance against a lethal dose of A170 (Table 6). The main components of CP-A172 are mannose and glucose (Figure 8). This feature distinguishes CP-A172 from the poly-N-succinyl β-1-6 glucosamine vaccine described by McKenney [45] and the capsular polysaccharide vaccine, that has acetylated fucosamine and mannosamine as major components [46].

When compared with A170 infected mice, A172-HK vaccinated animals displayed an elevated transcription level of IL-6 (Figure 7). IL-6 is an anti-inflammatory cytokine controlling the expression of pro-inflammatory cytokines (IFN-γ, IL-β and

TNF-α) [36–38]. Staphylococcal infections cause a profound release of proinflammatory cytokines, which enhance endothelial cell permeability and extravasation of bacteria from blood to the tissues [38]. In particular, elevated levels of IFN-γ cause necrosis in the kidneys of *S. aureus*-infected mice [37]. Thus, the A172 vaccine, inducing a balanced pro- and anti-inflammatory response, tempers the detrimental effects of an excessive inflammation.

It is not uncommon for clinical isolates to be resistant to staphylococcal phages. This finding deserves a final comment. Resistant bacteria, despite of their reduced fitness, sometimes persist because they attenuate the impact of phages on sensitive bacteria and thus contribute to maintain diversity within the bacterial population [47].

In conclusion, given the multiple virulence factors expressed by *S. aureus*, a whole cell bacterial vaccine would have significant potential. However, the application of these vaccines (against *S. aureus* and bacterial pathogens in general) is limited by the difficulty of developing reliable procedures of bacterial attenuation [41]. Selection for phage resistance promises to serve this purpose well. Of course, it must be recalled that the trade-off associated with phage resistance can be paid by the pathogen in a form other than lowered virulence. With these caveats, the evidence so far available – [18] and this article - suggests that selection for phage resistance as a means to curb bacterial virulence is not particular to the phage M^{Sa} or the bacterial strain A172, but more general. Phage resistance, long considered a problem in the fight against bacteria [48], is shown here to be an opportunity.

Author Contributions

Conceived and designed the experiments: RC RL AA MI MP DI. Performed the experiments: NN AS CG KB GB AE AC CM SR FR LR. Analyzed the data: RC RL AA KB GB AE AC MI MP CM SR FR LR DI. Wrote the paper: DI.

References

- Boyce JM (2005) Meticillin-resistant *Staphylococcus aureus*. Lancet Infect Dis 5: 653–654.
- Marris E (2005) Bugs gain vital ground in their battle against drugs. Nat Med 11: 461.
- Miller LG, Perdreaux-Remington F, Rieg G, Mahdi GS, Perroth JA, et al. (2005) Necrotising fasciitis caused by community-associated methicillin resistant *Staphylococcus aureus* in Los Angeles. N Engl J Med 352: 1445–1453.
- Sakoulas G, Eliopoulos GM, Fowler Jr. VG, Moellering Jr. RC, Novick R, et al. (2005) Reduced susceptibility of *Staphylococcus aureus* to vancomycin and platelet microbial protein correlates with defective autolysis and loss of accessory gene regulator (*agr*) function. Antimicrob Agents Chemother 49: 2687–2692.
- Lowy FD (1988) *Staphylococcus aureus* infections. N Engl J Med 339: 520–532.
- Zahid MS, Udden SM, Faruque AS, Calderwood SB, Mekalanos JJ, et al. (2008) Effect of phage on the infectivity of *Vibrio cholerae* and emergence of genetic variants. Infect Immun 76: 5266–5273.
- Jensen MA, Faruq SM, Mekalanos JJ, Levin BR (2006) Modeling the role of bacteriophage in the control of cholera outbreaks. Proc Natl Acad Sci USA 103: 4652–4657.
- Scott AE, Timms AR, Connerton PL, Carrillo CL, Radzum KA, et al. (2007) Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. PLOS Pathogens 3: 1142–1151.
- Mason W, J. Blevins J, Beenken K, Wibowo N, Ojima N, et al. (2001) Multiplex PCR protocol for the diagnosis of staphylococcal infections. J Clin Microbiol 39: 3332–3338.

10. Capparelli R, Parlato M, Borriello G, Salvatore P, Iannelli D (2007) Experimental phage therapy against *Staphylococcus aureus* in mice. *Antimicrob Agents Chemother* 51: 2765–2773.
11. Lachica V, Genigeorgis C, Hoepflich P (1979) Occurrence of protein A in *Staphylococcus aureus* and closely related *Staphylococcus* species. *J Clin Microbiol* 10: 752–753.
12. Peschel A, Otto M, Jack RW, Kalbacher H, Jung G, et al. (1999) Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J Biol Chem* 274: 8405–8410.
13. Ciucanu I, Kerek F (1984) A simple method for the permethylation of carbohydrates. *Carbohydr Res* 131: 209–217.
14. Cimmino A, Marchi G, Surico G, Hanuszkiewicz A, Evidente A, Holst O (2008) The structure of the O-specific polysaccharide of the lipopolysaccharide from *Pantoea agglomerata* strain FL1. *Carbohydr Res* 343: 392–396.
15. Korman R (1975) Cell wall composition of novobiocin-resistant pleiotropic mutant staphylococci. *J Bact* 124: 724–730.
16. Luria SE, Delbruck M (1943) Mutations of bacteria from virus sensitive to virus resistance. *Genetics* 28: 491–511.
17. Blaiotta G, Moschetti G, Simeoli E, Andolfi R, Villani F, et al. (2001) Monitoring lactic acid bacteria strains during “caciocotta” cheese production by restriction endonucleases analysis and pulsed-field gel electrophoresis. *J Dairy Res* 68: 139–144.
18. Capparelli R, Nocerino N, Iannaccone M, Ercolini D, Parlato M, et al. (2009) Phage therapy of *Salmonella enterica*: a fresh appraisal of phage therapy. *J Infect Dis* 201: 52: 52–61.
19. Weidenmaier C, Kokai-Kun J, Kristian SA, Chanturiya T, H. Kalbacher H, et al. (2004) Role of teichoic acids in *Staphylococcus aureus* nasal colonization, a major risk factor in nosocomial infections. *Nat Med* 10: 243–245.
20. Mellmann A, Friedrich AW, Rosenkotter N, Rothganger J, Karch H, et al. (2006) Automated DNA sequence-based early warning system for the detection of methicillin-resistant *Staphylococcus aureus* outbreaks. *PLoS Med* 3: e33.
21. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 38: 1008–15.
22. Blaiotta G, Ercolini D, Pennacchia C, Fusco V, Casaburi A, et al. (2004) PCR detection of staphylococcal enterotoxin genes in *Staphylococcus* spp. strains isolated from meat and dairy products. Evidence for new variants of *seG* and *seI* in *S. aureus* AB-8802. *J Appl Microbiol* 97: 719–730.
23. Blaiotta G, Fusco V, von Eiff C, Villani F, Becker K (2006) Biotyping of enterotoxigenic *Staphylococcus aureus* by enterotoxin gene cluster (*egc*) polymorphism and *spa* typing analyses. *Appl Environ Microbiol* 72: 6117–6123.
24. Sanderson A, Strominger J, Nathenson S (1962) Chemical structure of Teichoic acid from *Staphylococcus aureus*, strain Copenhagen. *J Biol Chem* 237: 3603–3613.
25. McCallum N, Karazum H, Getzmann R, Bischoff M, Majcherczyk P, et al. (2006) In vivo survival of teicoplanin-resistant *Staphylococcus aureus* and fitness cost of teicoplanin resistance. *Antimicrob Agents Chemother* 50: 2352–2360.
26. Glaser L, Ionesco H, Schaeffer P (1966) Teichoic acids as components of a specific phage receptor in *Bacillus subtilis*. *Biochem Biophys Acta* 124: 415–417.
27. Young FE (1967) Requirement of glucosylated teichoic acid for adsorption of phage in *Bacillus subtilis* 168. *Proc Natl Acad Sci USA* 58: 2377–2384.
28. Coyette JJ, Ghuyssen M (1968) Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. IX. Teichoic acid and phage adsorption. *Biochemistry* 7: 2385–2389.
29. Baggesen D, Wegener H, Madsen M (1997) Correlation of conversion of *Salmonella enterica* serovar *enteritidis* phage type 1, 4 or 6 to phage 7 with loss of lipopolysaccharide. *J Clin Microbiol* 35: 330–333.
30. Goerke C, Gressinger M, Endler K, Breikopt C, Wardechi CK, et al. (2007) High phenotypic diversity in infecting but not in colonizing *Staphylococcus aureus* populations. *Environ Microbiol* 9: 3134–3142.
31. Wardenburg JB, Schneewind O (2008) Vaccine protection against *Staphylococcus aureus* pneumonia. *J Exp Med* 205: 287–294.
32. Goodner K, Horsfall F (1935) The protective action of Type 1 antipneumococcus serum in mice. I. Quantitative aspects of the protection test. *J Exp Med* 62: 359–374.
33. Taborda C, Rivera J, Zaragoza O, Casadevall A (2003) More is not necessarily better: prozone-like effects in passive immunization IgG. *J Immunol* 170: 3621–3630.
34. Sparano JA, Wolf A, Brown D (2000) Troponins for predicting cardiotoxicity from cancer therapy. *Lancet* 356: 1947–1948.
35. Yeh ET (2006) Cardiotoxicity induced by chemotherapy and antibody therapy. *Ann Rev Med* 57: 485–498.
36. Dinarello CA (1996) Cytokines as mediators in pathogenesis of septic shock. In *Pathology of septic shock. Current Topics in Microbiology and Immunology. Vol 216* ERIetschel, HWagner, eds. Berlin: Springer-Verlag. pp 133–165.
37. Hu DL, Omoe H, Sasaki S, Sashinami H, Sakuraba H, et al. (2003) Vaccination with non-toxic mutant toxic shock syndrome toxin I protects against *Staphylococcus aureus* infection. *J Infect Dis* 188: 743–752.
38. Nilsson IM, Patti JM, Bremell T, Tarkowski A (1998) Vaccination with a recombinant fragment of collagen adhesion provides protection against *Staphylococcus aureus*-mediated septic death. *J Clin Invest* 101: 2640–2649.
39. Kuklin N, Clark D, Secore S, Cook J, Cope L (2006) A novel *Staphylococcus aureus* vaccine: iron surface determinant B induces rapid antibody response in rhesus macaques and specific increased survival in a murine *S. aureus* sepsis model. *Infect Immun* 74: 2215–2223.
40. Telford JL (2008) Bacterial genome variability and its impact on vaccine design. *Cell Host & Microbe* 3: 408–416.
41. Winzler EA (2008) Malaria research in the post-genomic era. *Nature* 455: 751–756.
42. Bohannan BJ, Lenski RE (2000) Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol Lett* 3: 362–377.
43. Labrie S, Samson J, Moineau S (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiology* 8: 317–327.
44. Hormaeche CE, Joisey H, Desilva L, Izhar M, Stocker MB (1991) Immunity conferred by Aro-*Salmonella* live vaccines. *Microb Pathog* 10: 149–158.
45. McKenney D, Kimberley L, Pouliot L, Wang Y, Murthy V, et al. (1999) Broadly protective vaccine for *Staphylococcus aureus* based on an in vivo-expressed antigen. *Science* 284: 1523–1527.
46. O’Riordan K, Lee J (2004) *Staphylococcus aureus* capsular polysaccharides. *Clin Microbiol Rev* 17: 218–234.
47. Middelboc M, Hamstrong A, Blackburn N, Sinn R, Fischer U, et al. (2001) Effect of bacteriophage on the population dynamics of four strains of pelagic marine bacteria. *Microb Ecol* 42: 395–406.
48. Lederberg J (1996) Smaller fleas...ad infinitum: therapeutic bacteriophage redux. *Proc Natl Acad Sci USA* 93: 3167–3168.

A Proteomic Approach to Investigate Myocarditis

Andrea Carpentieri*, Chiara Giangrande*,
Piero Pucci and Angela Amoresano
*Department of Organic Chemistry and Biochemistry,
Facoltà di Scienze Biotechnologiche,
Federico II University of Naples, Naples,
Italy*

1. Introduction

Myocarditis is an inflammatory disease of the cardiac muscle which might be related to viral (mainly parvovirus B19 and many others), protozoan (*Borrelia burgdorferi*, *Trypanosoma cruzi*, *Toxoplasma gondii*), bacterial (*Brucella*, *Corynebacterium diphtheriae*, *Gonococcus*, *Haemophilus influenzae*, *Actinomyces*, *Tropheryma whipplei*, *Vibrio cholerae*, *Borrelia burgdorferi*, *Leptospira*, *Rickettsia*), fungal (*Aspergillus*) and other non viral pathogens (Rezkalla SH et al., 2010, Blauwet LA et al., 2010, Cihakova D et al., 2010) infections; It has been reported, however, that this kind of inflammation might be caused by an hypersensitivity response to drugs (Kühl U et al., 2009). The final effect in each case is represented by myocardial infiltration of immunocompetent cells following any kind of cardiac injury.

Myocarditis presents with many symptoms, from chest pain that spontaneously resolves without treatment to cardiogenic shock and sudden death (Kühl U et al., 2010, Taylor CL et al 2010). The major long-term consequence is dilated cardiomyopathy with chronic heart failure (Lv H et al. 2011, Stensaeth KH et al 2011).

Nowadays, diagnostic tools available for this disease are mainly related to general investigations (such as electrocardiography) and analysis of the most abundant serum proteins, whose alteration is related to cardiac pathology, even if it's not specifically connected to myocarditis itself. Here we propose preliminary speculations on the serum proteins profiling (both in the expression level and in the characterization of post translational modifications) and on free peptides identification in myocarditis affected patients (compared to healthy individual), that could be helpful in finding specific markers for this pathology.

As many other inflammatory events this disease involves in fact different kinds of biological macromolecules, here we focus in particular on proteins. More than 50% of total protein content in a cell is post translationally modified. The pattern of Post Translational Modifications (PTMs) on proteins constitute a molecular code that dictates protein conformation, cellular location, macromolecular interactions and activities, depending on cell type, tissue and environmental conditions (Diernfellner AC et al., 2011, Savidge TC,

* Equally contributed to the work.

2011, Hao P et al.,2011). It is very well known that biological function of many proteins is strictly related to the presence of the appropriate set of PTMs. Most importantly, PTMs deregulation might be involved in the development of diseases , in fact, as a consequence of many pathologies, PTMs set might be altered (Dell A et al., 2001, Kim YJ. Et al 1997, Dube DH. Et al., 2005, Granovsky M. Et al 2000).

Proteomics investigations offer all useful tools to deeply investigate this kind of alterations. Two dimensional electrophoresis coupled with software mediated image analysis, affinity chromatographies and especially Mass Spectrometry (MS) present high levels of sensitivity, accuracy and reproducibility and these are all fundamental requirements that this kind of study needs.

As previously described by our group (Carpentieri A, Giangrande C, et al., 2010), serum glycoproteome characterization might be crucial in clinical investigation. In fact we showed that the *N*-glycan profiling of serum glycoproteins extracted from myocarditis affected donors compared to the ones of healthy people shows many peculiarities. We demonstrated that many of the extracted oligosaccharides are in fact incomplete or truncated structures whilst others show a high level of fucosylation, all these results fully matched previously published data about the glycosylation in proteins during chronic inflammation events (Dell A et al., 2001, Kim YJ. Et al 1997, Dube DH. Et al., 2005, Granovsky M. Et al 2000, Carpentieri A, Giangrande C, et al., 2010).

Thanks to the high sensitivity of the most modern analytical techniques, several research groups focus the research at level of peptides rather than at protein level (Taylor-Papadimitriou J. et al., 1994 Amado F et al., 2010, Menschaert G et al. 2010). In fact, these endogenous peptides are referred to as the peptidome. Initially, peptidomic analyses were conducted as a method to study neuropeptides and peptide hormones; these are signaling molecules that function in a variety of physiological processes (Ludwig M. 2011, Colgrave ML et al., 2011). Recent studies have found large numbers of cellular peptides with half-lives of several seconds, raising the possibility that they may be involved in biological functions (Gorman PM et al 2003).

Here we propose preliminary data to show molecular basis of myocarditis using a proteomic approach. The analyses were focused on the serum proteins profiling, namely the study of the expression level and the characterization of glycoproteins involved in the pathology. A classical two-dimensional gel electrophoresis procedure to obtain protein maps and a “gel-free” comparison of the glycoproteomes in healthy and myocarditis human sera by using advanced mass spectrometry were reported. Free peptides identification in myocarditis affected patients (compared to healthy individual), was achieved in order to provide possible specific markers for this pathology.

2. Materials and methods

2.1 Materials

Human serum samples from 8 healthy donors (all Caucasian 4 males and 4 females aged between 60 and 80 years, all other clinical informations were covered by laws on privacy) and from 5 myocarditis affected patients (all Caucasian 3 males and 2 females aged between 60 and 85 years, all other clinical informations were covered by laws on privacy), respectively, have been obtained from the “Servizio Analisi” Policlinico, Napoli. Aliquots of serum samples from different donors were pooled in order to obtain an average overview of glycoforms distribution.

Guanidine, dithiothreitol (DTT), trypsin, α -cyano-4-hydroxycinnamic acid were purchased from Sigma. Iodoacetamide (IAM), tris(hydroxymethyl)aminomethane, calcium chloride and ammonium bicarbonate (AMBIC) were purchased from Fluka as well as the MALDI matrix 2,5-, α -cyano-4-hydroxycinnamic. Methanol, trifluoroacetic acid (TFA) and acetonitrile (ACN) are HPLC grade type from Carlo Erba, whereas the other solvents are from Baker. Gel filtration columns PD-10 are from Pharmacia, the HPLC ones from Phenomenex, whereas the pre-packed columns Sep-pak C-18 are from Waters.

Comassie Brilliant Blue was from Bio-Rad. PNGase F were purchased from Boheringer. Ion exchange resins Dowex H+ (50W-X8 50-100 mesh) was provided by BDH. Concanavalin A sepharose resin was purchased from Amersham Biosciences.

2.2 Protein concentration determination

Sera protein concentration was determined by Bradford assay method, using bovine serum albumin (BSA) as standard. Known amounts of BSA were diluted in 800 μ L of H₂O and then mixed to 200 μ L of Comassie Brilliant Blue. 5 different BSA concentrations were determined by measuring absorbance at 595 nm and used to obtain a linear calibration curve. Three different sera dilutions were measured at 595 nm. Absorbance data were interpolated on the calibration curve, allowing the determination of protein concentration in the different samples.

2.3 Serum depletion

The depletion of the most abundant proteins from each serum sample was performed using Multiple Affinity Removal Spin Cartridges from Agilent. The procedure was performed at room temperature according to manufacturer's instructions and then immediately frozen to -20°C.

2.4 Free peptides analysis

300 μ L acetonitrile was added to 100 μ L of serum, incubated at room temperature for 30 minutes and then centrifuged at 13000 rpm. Supernatants were collected and concentrated by vacuum centrifugation (SAVANT) and resuspended in 20 μ L formic acid 0.1%. Samples were desalted by C18 Zip Tip (Millipore) and diluted 100 times prior mass spectrometry analyses.

LC/MS-MS HPLC-Chip/Q-TOF 6520

Peptides were analyzed by a HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The capillary column works at a flow of 4 μ L/min, concentrating and washing the sample in a 40 nL enrichment column. The sample was then fractionated on a C18 reverse-phase capillary column (75 μ m~43 mm in the Agilent Technologies chip) at flow rate of 400 nL/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min.

Data were acquired through MassHunter software (Agilent Technologies). Proteins identification was achieved by using Mascot software (Matrix science), with a tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and choosing methionine oxidation and glutamine conversion in pyro-glutamic acid as variable modifications.

MALDI-TOF/TOF

Peptides were also analyzed by MALDI-TOF/TOF using a 4800 Plus MALDI-TOF/TOF (Applied Biosystems). Samples were mixed on MALDI plate to the matrix consisting of a

solution of 10 mg/mL α -cyano-4-hydroxycinnamic, whose preparation consisted in the resuspension of α -cyano-4-hydroxycinnamic in water and acetonitrile 10:1 (v/v). The instrument was calibrated using a mixture of standard peptides (Applied Biosystems). Spectra were register even in reflector positive. MS/MS spectra were performed with CID using air as collision gas. Spectra were manually interpreted.

2.5 2D-gel electrophoresis

IEF (first dimension) was carried out on non-linear wide-range immobilized pH gradients (pH 4-7; 7 cm long IPG strips; GE Healthcare, Uppsala, Sweden) and achieved using the EttanTM IPGphorTM system (GE Healthcare, Uppsala, Sweden). Analytical-run IPG-strips were rehydrated with 125 μ g of total proteins in 125 μ l of rehydration buffer (urea 8 M, CHAPS 2%, 0,5% (v/v) IPG Buffer, bromophenol blue 0,002%) for 12 h at 20°C. The strips were then focused according to the following electrical conditions at 20°C: 500 V for 30 min, from 1000 V for 30 min, 5000 V until a total of 15000 Vt was reached. After focusing IPG strips were equilibrated for 15 min in 6 M urea, 30% (vol/vol) glycerol, 2% (wt/vol) SDS, 0,05 M Tris-HCl, pH 6,8, 2% (wt/vol) DTT, and subsequently for 15 min in the same urea/SDS/Tris buffer solution but substituting the 2% (wt/vol) DTT with 2,5% (wt/vol) iodoacetamide. The second dimension was carried out on 12% polyacrylamide gels at 25 mA/gel constant current until the dye front reached the bottom of the gel. MS gel was stained with colloidal comassie.

2.6 Image analysis

Gels images were acquired with an Epson expression 1680 PRO scanner. Computer-aided 2-D image analysis was carried out using the ImageMasterTM 2D Platinum software (GE Healthcare, Uppsala, Sweden). Differentially expressed spots were selected for MS analysis.

In order to find differentially expressed proteins, comassie stained gel image of serum proteins from healthy individual was matched with the one of myocarditis affected patient. The apparent isoelectric points and molecular masses of the proteins were calculated with ImageMaster 2D Platinum 6.0 using identified proteins with known parameters as references.

Relative spot volumes (%V) ($V = \text{integration of OD over the spot area}$; $\%V = V \text{ single spot} / V \text{ total spot}$) were used for quantitative analysis in order to decrease experimental errors. The normalized intensity of spots on three replicate 2-D gels was averaged and standard deviation was calculated for each condition.

2.7 Protein identification by mass spectrometry

In situ digestion

Protein spots were excised from the gel and destained by repetitive washes with 0.1 M NH_4HCO_3 pH 7.5 and acetonitrile. Enzymatic digestion was carried out with trypsin (12.5 ng/ μ l) in 10 mM ammonium bicarbonate buffer pH 7.8. Gel pieces were incubated at 4 °C for 2 h. Trypsin solution was then removed and a new aliquot of the same solution was added; samples were incubated for 18 h at 37 °C. A minimum reaction volume was used as to obtain the complete rehydration of the gel. Peptides were then extracted by washing the

gel particles with 10mM ammonium bicarbonate and 1% formic acid in 50% acetonitrile at room temperature.

Mass spectrometry and protein identification

LC-MS/MS analyses were performed as previously described for free peptides.

Mass spectrometric obtained data were used for protein identification using the software MASCOT that compare peptide masses obtained by MS and MS/MS data of each tryptic digestion with the theoretical peptide masses from all the proteins accessible in the databases (Peptide Mass Fingerprinting, PMF). Database searches were performed in NCBI databank (National Center for Biotechnology Information), restricting the analysis to the pertinent taxonomies. The parameters used for the identification were: tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and cysteine carbamidomethylation as fixed modification. Variable modifications were methionine oxidation, glutamine conversion in pyro-glutamic acid, and asparagine deamidation.

2.8 Boronate affinity chromatography

Glycoproteins were purified using PBA-bound agarose (Sigma-Aldrich, Munich, Germany). 500 μ l of sample previously diluted (1 : 1) with equilibration buffer (50 mM taurine/NaOH, pH 8.7, containing 3–10 mM MgCl₂) was incubated with 200 μ l of pre-washed immobilized ligand resin for 1 h on ice and with gentle shaking. After transfer of the resin into 1,5 ml eppendorf tubes, the non-binding fraction was collected by low speed centrifugation (10 s, 500 \times g). The resin was then thoroughly rinsed with equilibration buffer (six washes of 150 μ l each) and 1 N NaCl (three washes of 150 μ l each). For final elution of the bound fraction, a total of six washes (150 μ l each) with taurine buffer containing 50 mM sorbitol were used. Three successive fractions (150 μ l each) were pooled before further analysis.

2.9 Deglycosylation

Glycopeptides were lyophilized, resuspended in 10 mM AMBIC and incubated with PNGase F (5 U), for 12–16 h at 37°C. Deglycosylation was carried out also on unbound peptides, in order to release the glycans not recognized by Boronate affinity chromatography.

3. Results and discussion

3.1 Free peptides analysis

The analyses to investigate the “peptidomic” both in healthy and in pathological samples were accomplished by using a gel-free approach. The peptide component of the eluted fraction was analyzed by tandem mass spectrometry. The stringency of scoring parameters of the MASCOT algorithm minimized the number of false positive identifications. Most MS/MS spectra giving positive hits were derived from doubly and triply charged precursor ions that resulted predominantly in y-ion series.

Triplicate LC-MS/MS analysis of supernatants after ACN precipitation of serum proteins, showed the occurrence of many free peptides in both analyzed sera. As reported in Table 1, the total number of detected and identified peptides was 41. Among these, 9 peptides were unique in the pathologic sample. It should be noted that some peptides were identified in both samples.

m/z	RT	Sequence	Peptide	Protein	H/M ratio
567.95	17.09	QAGAAGSRMNFPRGVLS	(650-666)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
513.78	17.70	YYLQGAKIPKPEASFSPR	(627-644)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
823.17	22.11	MNFRPGVLSSRQLGLPGPPDVPDH AAYHPF	(658-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
1005.98	22.78	QLGLPGPPDVPDHAAYHPF	(669-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	0.68±0.18
757.71	19.96	SRQLGLPGPPDVPDHAAYHPF	(667-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	0.27±0.09
681.85	22.07	PGVLSSRQLGLPGPPDVPDHAAYHP F	(662-687)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	0.18±0.03
576.90	20.69	PGVLSSRQLGLPGPPDVPDHAAYHP FR	(662-688)	Q14624 Inter-alpha-trypsin inhibitor heavy chain H4	H: not detected
379.71	0.55	LAEGGGVR	(28-35)	P02671 Fibrinogen alpha chain	25±3.11
453.24	3.65	FLAEGGGVR	(27-35)	P02671 Fibrinogen alpha chain	2.48±0.57
510.76	11.29	DFLAEGGGVR	(26-35)	P02671 Fibrinogen alpha chain	1.33±0.76
539.27	11.70	GDFLAEGGGVR	(25-35)	P02671 Fibrinogen alpha chain	3.15±1.02
597.77	15.59	SGEGDFLAEGGGV	(22-34)	P02671 Fibrinogen alpha chain	0.85±0.23
603.79	12.57	EGDFLAEGGGVR	(24-35)	P02671 Fibrinogen alpha chain	2.42±0.63
632.30	13.50	GEGDFLAEGGGVR	(23-35)	P02671 Fibrinogen alpha chain	2.72±0.71
655.28	16.03	DSGEGDFLAEGGGV	(21-34)	P02671 Fibrinogen alpha chain	1.23±0.36
675.81	13.67	SGEGDFLAEGGGVR	(22-35)	P02671 Fibrinogen alpha chain	2.32±0.58
690.80	16.13	ADSGEGDFLAEGGGV	(20-34)	P02671 Fibrinogen alpha chain	3.25±0.82
733.33	14.04	DSGEGDFLAEGGGVR	(21-35)	P02671 Fibrinogen alpha chain	1.59±0.12
768.85	14.08	ADSGEGDFLAEGGGVR	(20-35)	P02671 Fibrinogen alpha chain	6.16±1.01
851.71	13.58	SSSYSKQFTSSTSYNRGDSTFES	(576-598)	P02671 Fibrinogen alpha chain	0.65±0.17

m/z	RT	Sequence	Peptide	Protein	H/M ratio
693.06	13.06	SSSYSKQFTSSTSYNRGDSTFESKS	(576-600)	P02671 Fibrinogen alpha chain	M: not detected
733.83	14.48	SSSYSKQFTSSTSYNRGDSTFESKSY	(576-601)	P02671 Fibrinogen alpha chain	M: not detected
619.75	17.89	QGVNDNEEGFF	(31-41)	P02675 Fibrinogen beta chain	3.25±0.89
663.26	16.54	QGVNDNEEGFFS	(31-42)	P02675 Fibrinogen beta chain	1.97±0.41
696.28	17.64	QGVNDNEEGFFSA	(31-43)	P02675 Fibrinogen beta chain	2.31±0.53
404.55	18.90	RIHWESASLL	(1310-1319)	P01024 Complement C3	H: not detected
402.22	14.29	THRIHWESASLLR	(1308-1320)	P01024 Complement C3	H: not detected
445.25	16.98	SKITHRIHWESASLL	(1305-1319)	P01024 Complement C3	H: not detected
415.20	7.60	HWESASL	(1312-1318)	P01024 Complement C3	H: not detected
471.74	16.21	HWESASLL	(1312-1319)	P01024 Complement C3	H: not detected
851.07	22.22	TLEIPGNSDPNMIPDGFNSYVR	(957-979)	P0C0L4 Complement C4-A	H: not detected
1054.53	25.86	DDPDAPLQPVTPLQLFEGR	(1429-1447)	P0C0L4 Complement C4-A	H: not detected
489.96	20.26	RHPDYSVLLLR	(169-180)	P02768 Serum Albumin	0.43±0.17
417.91	16.67	KFQNALLVRY	(426-435)	P02768 Serum Albumin	H: not detected
547.31	16.25	KVPQVSTPTLVEVSR	(438-452)	P02768 Serum Albumin	H: not detected
868.11	21.57	AVPPNNSNAAEDDLPTVELQGVVPR	(14-38)	P00488 Coagulation factor XIII A	2.40±0.84
920.14	21.09	RAVPPNNSNAAEDDLPTVELQGVVPR	(13-38)	P00488 Coagulation factor XIII A	23.60±3.61
837.93	21.39	TAFGGRRVAVPPNNSNAAEDDLPTVELQGVVPR	(7-38)	P00488 Coagulation factor XIII A	91.32±7.39
781.37	17.84	TATSEYQTFNPR	(315-327)	P00734 Prothrombin	H: not detected
868.46	25.47	TGIFTDQVLSVLKGE	(86-101)	P02655 Apolipoprotein C-II	H: not detected
572.95	12.91	DALSSVQESQVAQQAR	(45-60)	P02656 Apolipoprotein C-III	H: not detected
803.76	21.14	AATVGLAGQPLQERAQAWGERL	(210-232)	P02649 Apolipoprotein E	0.02

Table 1. List of free peptides identified by LC-MS/MS. Averaged area of chromatographic peaks from healthy(H) and myocarditis (M) ratio, indicates that some of them were differently represented. Peptide sequences were validated by MALDI-TOF/TOF analyses too.

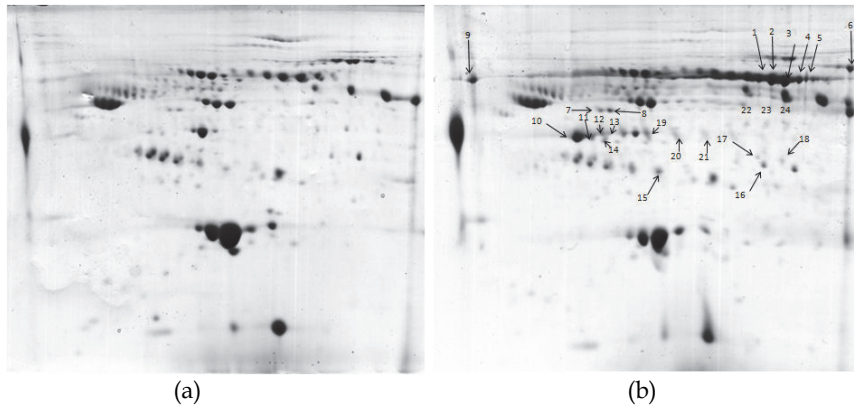


Fig. 2. Two-dimensional electrophoresis gels of healthy (A) and myocarditis affected (B) sera proteins. Arrows and numbers in (B) correspond to numbers of identified spots in Table 2.

It should be underlined that all the spots we further analysed were taken from the pathologic serum and results are summarized in Table 2.

Spot	Accession number	Protein
1	P02790	Hemopexin
2	P02790	Hemopexin
3	P02790	Hemopexin
4	P02790	Hemopexin
5	P02790	Hemopexin
6	P02787	Serotransferrin
7	P01008	Antithrombin III
8	P01008	Antithrombin III
9	P01011	Alpha-1 antichymotrypsin
10	P01024	Complement C3
11	P04196	Histidine rich glycoprotein
12	P01024	Complement C3
13	P00738	Haptoglobin
14	P04196	Histidine rich glycoprotein
15	Q14624	Inter-alpha-trypsin inhibitor heavy chain H4
16	P08603	Complement factor H
17	P08603	Complement factor H
18	P03952	Plasma kallikrein
19	P06727	Apolipoprotein A-IV
20	P00738	Haptoglobin
21	P00738	Haptoglobin
22	P36955	Pigment epithelium-derived factor
23	P36955	Pigment epithelium-derived factor
24	P36955	Pigment epithelium-derived factor

Table 2. List of proteins identified by LC-MS/MS of spots excised from two-dimensional gels of sera taken from myocarditis affected patients.

Proteins excised from the gel were reduced alkylated and, *in situ*, digested with trypsin. The resulting peptide mixtures were directly analysed by LC-MS/MS according to the peptide mass fingerprinting procedure. MS and MS-MS obtained data were used to search for a non-redundant sequence using the in-house MASCOT software, taking advantage of the specificity of trypsin and of the taxonomic category of the samples. The number of measured masses that matched within the given mass accuracy of 20 ppm was recorded and the proteins that had the highest number of peptide matches were examined.

Thanks to this approach, we could identify many proteins differently expressed in the two sera. Among identified proteins, hemopexin (Dooley H et al., 2010), complement C3 (Adamsson Eryd S et al., 2011, Onat A et al., 2011), plasma kallikrein (Kolte D et al., 2011) are undoubtedly related to inflammatory events and most interestingly they resulted clearly over expressed in the pathologic sample.

A preliminary speculation on identified proteins might involve the function of hemopexin. As shown by literature data (Dooley H et al., 2010, : Mauk MR et al., 2011, Larsen R et al., 2010), hemopexin is a serum protein with the very well known function of scavenging the heme released or lost by the turnover of heme proteins such as haemoglobin or by haemolysis caused by parasitic infection, and thus protects the body from the oxidative damage that free heme can cause (Larsen R et al., 2010). Myocarditis itself it's not related to haemolysis phenomena, but some viral infections may cause it, therefore, finding a very high level of hemopexin in a myocarditis affected patient might be a putative marker of the inflammation itself (quite common are in fact viral myocarditis).

Moreover we found some connections between differently expressed proteins in pathologic serum and identified free peptides; in particular we could detect some free peptides some peptides from proteins Complement C3, Inter-alpha-trypsin inhibitor heavy chain H4, Antithrombin III which results over expressed in pathologic serum, (see Tab 1).

3.3 Boronate affinity chromatography

The third part of this work focus on the investigation of one of the most important post-translational modification, the glycosylation. The importance of investigation of post-translational modifications (PTM) is notably increased in the proteomic era, as they play a critical role in cellular functioning and they vary in response to environmental stimuli, signalling modulators or development of diseases (Laurell E et al., 2011). PTMs can affect biological functions thus playing a critical role in cellular functioning. Moreover, they can vary in response to environmental stimuli, thus finely tuning cellular mechanisms and their deregulation might be involved in the development of diseases. A huge number of different types of PTMs have been identified but only a few are reversible and important for regulation of biological processes (Wu C et al. 2011). The pattern of PTMs on proteins constitute a molecular code that dictates protein conformation, cellular location, macromolecular interactions and activities, depending on cell type, tissue and environmental conditions. Understanding this code is the major challenge of proteomics in post-genomic era. Existing methodologies for PTMs identification essentially rely on specific enrichment procedures able to selectively increase the amount of modified peptides. These procedures have to be integrated with sophisticated mass spectrometric experiments to address the identifications of PTMs. The development of a variety of new technologies for exploring the structures of the sugar chains has opened up a new frontier in the glycomics field. Moreover recent progress in mass spectrometry led to new challenges in glycomics, including the development of rapid glycan enrichment. Recently our group introduced an

easy to handle strategy to give preliminary insights for the comparison of glycoproteomes in healthy and pathological human sera, by using a single Con A affinity chromatography step coupled with mass spectrometry techniques. The strategy led both to the identification of 69 different glycosylation sites within 49 different proteins and to the definition of the glycosylation patterns. Moreover, glycoform distribution in myocarditis and hepatic carcinoma has been reported. The analysis of glycan profiling, once extracted from serum glycopeptides, is essential for comparative studies on different sera samples thus providing a useful tool for the development of screening procedures (Carpentieri A, Giangrande C, et al., 2010).

In this paper, a different simple and rapid procedure to obtain an overview of the glycosylation sites profiling in the two samples was accomplished. This was achieved by enriching for the N-linked glycopeptides resulting from trypsin digestion of sera samples in order to enhance the identification of N-glycosylation sites using LC-MS/MS. The analyses have been carried out by using healthy sera as control.

To reduce the complexity of the whole sample, Boronate affinity purification was rapidly performed in batch after tryptic digestion. Thanks to the vicinal diols binding capacity no discrimination on the basis of the glycan type was performed thus, in a proof of principle, all glycopeptides could be selected. The recovered glycopeptides were then deglycosylated by PNGase F treatment and the peptide mixtures directly analysed by LC-MS/MS.

The analyses were performed on intact serum samples without any pre-purification step or removal of most abundant proteins. The peptide component of the eluted fraction was analysed by tandem mass spectrometry. Similar analyses were carried out on the unbound Boronate fractions, mainly containing non-glycosylated peptides.

The data were then pooled and summarised in Table 3. The results presented here demonstrated that Boronate affinity chromatography on serum tryptic digests is a useful tool to enhance the detection by LC-MS/MS of glycopeptide. Another advantage of the strategy relies on the fact that it was performed on glycopeptides instead of glycoproteins, therefore there were no SDS-PAGE step, no isolation of the individual glycoproteins and no *in situ* digestion. This allowed the detection of less abundant glycopeptides together with the most represented ones, such as those deriving from albumins or immunoglobulins.

As shown in Table 3, all the selected peptides still contained the conserved N-glycosylation motif (Asn-X-Ser/Thr), thus indicating that N-glycosylation peptides were isolated with high selectivity. This analysis led both to the localization of the modification sites and identification of glycoproteins.

The presence of a putative N-glycosylation site was confirmed by the fact that peptides mass was increase of 1 Da, due to the conversion of Asn into Asp after PNGase F incubation. However, some non-specific peptides, namely non-glycosylated peptides, were detected in the eluted Boronate fraction, and identified as belonging to most abundant proteins like albumin.

Spontaneous deamidation seems rather unlikely for generating the results presented, although it cannot be excluded completely. Most MS/MS spectra giving positive hits were derived from doubly charged precursor ions that resulted predominantly in y-ion series.

As a whole, using Boronate affinity approach we could confirm the previously identified glycosylation sites based on ConcanavalinA enrichment and 5 more glycosylation sites were identified thus refining previous data (Carpentieri A, Giangrande C, et al., 2010) on myocarditis glycoproteome.

	Protein	Sequence	Peptide
P01011	Alpha-1-antichymotrypsin	YTGN*ASALFILPDQDKH,M	268-283
P02763	Alpha-1-acid glycoprotein	SVQEIQATFFYFTPN*KTEDTIFLR ^{H,M} QDQCIYN*TTYLNVQR ^{H,M}	58-81 87-101
P01009	Alpha-1-antitrypsin	YLGN*ATAIFFLPDEGKH,M QLAHQSN*STNIFFSPVSIATAFAMLSL GTKH ADTHDEILEGLNFN*LTEIPEAQIHEGF QELLRH,M	268-283 64-93 94-125
P01023	Alpha-2-macroglobulin	SLGNVN*FTVSAEALQELCGTEVPS VPEHGRKH	864-886
P43652	Afamin	DIENFN*STQKH,M YAEDKFN*ETTEKH	28-37 396-407
P01008	Antithrombin III	LGACN*DTLQQLMEVFKH,M SLTFN*ETYQDISELVYGAKH,M	124-139 183-201
P04114	Apolipoprotein B-100	FN*SSYLOGTINQITGRH,M FVEGSHN*STVSLTTKH AEEEMLEN*VSLVCPKM YDFN*SSMLYSTAKM	1522-1536 3405-3419 27-41 3462-3474
P05090	Apolipoprotein D	ADGTVNQIEGEATPVN*LTEPAKM	83-104
P02749	Apolipoprotein H	VYKPSAGN*NSLYRH LGN*WSAMPCKM	155-167 251-261
O75882	Attractin	IDSTGN*VTNELRH,M GPVKMPSQAPTGNFYQPPLN*SSMC LEDNRH	411-422 1023-1052
P00450	Ceruloplasmin	EHEGAIYPDN*TTDFQRH,M EN*LTAPGSDSAVFFEQGTTRH,M	129-144 396-414
P08603	Complement factor H	ISEEN*ETTCYMGKH MDGASN*VTCINSRH,M IPCSQPPQIEHGTIN*SSRM SPDVIN*GSPISQKH	907-919 1024-1036 868-885 212-224
P10909	Clusterin	LAN*LTQGEDQYYLRH,M	372-385
Q8IWW2	Contactin-4	LN*GTDVDVTGMDFRM	64-76
P05156	Complement factor 1	FLNN*GTCTAEGKH,M	100-111
P01024	Complement C3	TVLTPATNHMGN*VFTTIPANRH	74-94
P0C0L4	Complement C4-A	GLN*VTLSTGRH,M	1326-1336
P02748	Complement component C9	AVN*ITSENLIDDVVSLIRH,M	413-430
Q14517	Protocadherin-fat 1	QVYN*LTVRAKDKM FSMDYKTGALTVQN*TTQLRSRM	994-1005 1930-1950
P02765	Alpha-2-HS-glycoprotein	VCQDCPLLAPLN*DTRH,M AALAAFNAQNN*GSNFQLEEISRH	145-159 166-187
Q03591	Complement factor H related protein 1	LQNNENN*ISCOVERH	120-132

	Protein	Sequence	Peptide
Q13439	Golgin subfamily A member 4	HN*STLKQLMREFNTQLAQKH	1990-2008
P02790	Hemopexin	SWPAVGN*CSSALRH,M ALPQPQN*VTSLLGCTHH,M	181-193 447-462
P00738	Haptoglobin	VVLHPN*YSQVDIGLIKH,M MVSHHN*LTTGATLINEQWLLTTAKH ,M NLFLN*HSEN*ATAKH,M	236-251 179-202 203-215
P04196	Histidine rich glycoprotein	VEN*TVYYLVLDVQESDCSVLSRH VIDFN*CTTSSVSSALANTKH,M	61-83 121-139
P05155	Plasma protease C1 inhibitor	VGQLQLSHN*LSLVILVPQNLKH,M	344-364
P01857	Ig alpha-1 chain C region	LSLHRPALEDLLGSEAN*LTCTLTGL RH	127-153
P01859	Ig gamma-2 chain C region	TKPREEQFN*STFRH TPLTAN*ITKH	168-180 200-208
P01860	Ig gamma-1 chain C region	EEQYN*STYRH,M	136-144
P01591	Immunoglobulin J chain	EN*ISDPTSPLRH,M	48-58
P40189	Interleukin-6 receptor beta	QQYFKQN*CSQHESPDISHFERM	812-833
P56199	Integrin alpha-1	SYFSSLN*L TIRM	1096-1106
P29622	Kallistatin	DFYVDEN*TTVRH	232-242
P01042	Kininogen-1	LNAENN*ATFYFKH,M	389-400
P11279	Lysosome-associated membrane glycoprotein 1	DPAFKAAN*GSLRM	314-325
P01871	Ig mu chain C region	YKN*NSDISSTRH GLTFQQN*ASSMCVPDQDPAIRM	44-54 204-223
Q9HC10	Otoferlin	NEMLEIQVFN*YSKVFSNKH,M	59-76
Q5VU65	Nuclear pore membrane glycoprotein 210-like	EVVVN*ASSRH	1551-1559
Q9Y5E7	Protocadherin beta 2	ETRSEYN*ITITVTDFGTPRM	414-432
P36955	Pigment epithelium derived factor	VTQN*LTLIEESLTSEFIHDIDRH	282-303
P27169	Serum paraoxanase/arylesterase1	HAN*WTLTPLKH	250-259
P49908	Selenoprotein P	EGYSN*ISYIVVNHQGISSRH	79-97
Q9Y275	Tumor necrosis factor ligand superfamily member 13B	CIQNMPETLPN*NSCYSAGIAKM	232-252

	Protein	Sequence	Peptide
P40225	Thrombopoietin	IHELLN*GTRGLFPGPSRRM	250-267
P02787	Serotransferrin	QQQHILFGSN*VTDCSGNFCLFRH,M CGLVPVLAENYN*KSDNCEDTPEAGY FAVAVVKM	622-642 421-452
P07996	Thrombospondin 1	VVN*STTGPGEHLRH,M	1065-1077
P04004	Vitronectin	NN*ATVHEQVGGPSLTSDLQAQSKM	85-107
P25311	Zinc-alpha-2-glycoprotein	DIVEYIN*DSN*GSHVLQGRH,M	100-117
P02766	Transthyretin	ALGISPFHEHAEVVFTAN*DSGPR ^H	101-123

Table 3. LC/MSMS analysis for the identification of glycosylation sites, H indicates peptides deriving from healthy serum and M indicates the ones from myocarditis serum.

4. Conclusions

In biomedical applications, a comparative approach is usually employed to identify proteins that are up and down regulated in a disease specific manner for use as diagnostic markers or therapeutic targets. This report represents an overview of the investigation at molecular level of myocarditis by using a proteomic approach.

Serum proteins (including the N-glycosylation sites profiling) and glycoproteins and free peptides occurring in human sera from healthy donors were compared to the ones from myocarditis patients. This procedure, allowed the identification of several N-glycosylation sites by a single-step proteomic approach, contemporarily probing an entire complex sample by LC-MS/MS. Thanks to the depletion of the serum most abundant proteins, we could detect some of the very weakly represented free peptides, whose presence is connected to the pathology itself. The high resolution, the sensitivity and the reproducibility of the used techniques led to the identification of some up regulated proteins in the serum from a myocarditis affected patient, all these proteins are connected to inflammatory events and one in particular (hemopexin) opens the way to new speculations in serum proteins as a specific marker for pathologic state.

Finally, this proteomic approach represents a new opportunity for therapeutics and early diagnostics, for the screening of proteic biomarkers in pathological status. Finding a biomarker molecule that precisely indicates certain kind of pathology, is something quite difficult to achieve since it requires a huge background in many different fields of clinical investigation. Here we contribute with putative diagnostic species that could really be helpful for an early diagnosis myocarditis event.

5. Acknowledgments

This work was supported by grants from the INBB, Ministero dell'Universita` e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale 2002, 2003, 2005, 2006; FIRB 2001). Support from the National Center of Excellence in Molecular Medicine (MIUR - Rome) and from the Regional Center of Competence (CRdC ATIBB, Regione Campania - Naples) is gratefully acknowledged. The authors declare that this paper is novel and does not overlap with any already published articles.

6. References

- Adamsson Eryd S, Smith JG, Melander O, Hedblad B, Engström G. Inflammation-sensitive proteins and risk of atrial fibrillation: a population-based cohort study. *Eur J Epidemiol.* 2011 Mar 19;18(4):539-50.
- Amado F, Lobo MJ, Domingues P, Duarte JA, Vitorino R. Salivary peptidomics. *Expert Rev Proteomics.* 2010 Oct;7(5):709-21.
- Blauwet LA, Cooper LT. Myocarditis. *Prog Cardiovasc Dis.* 2010 Jan-Feb;52(4):274-88.
- Carpentieri A, Giangrande C, Pucci P, Amoresano A. Glycoproteome study in myocardial lesions serum by integrated mass spectrometry approach: preliminary insights. *Eur J Mass Spectrom (Chichester, Eng).* 2010;16(1):123-49.
- Cihakova D, Rose NR. Pathogenesis of myocarditis and dilated cardiomyopathy. *Adv Immunol.* 8;99:95-114.
- Colgrave ML, Xi L, Lehnert SA, Flatscher-Bader T, Wadensten H, Nilsson A, Andren PE, Wijffels G. Neuropeptide profiling of the bovine hypothalamus: Thermal stabilization is an effective tool in inhibiting post-mortem degradation. *Proteomics.* 2011 Apr;11(7):1264-76.
- Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science.* 291,2351-2356 (2001).
- Diernfellner AC, Schafmeier T. Phosphorylations: making the *Neurospora crassa* circadian clock tick. *FEBS Lett.* 2011 Mar 28 [Epub ahead of print].
- Dooley H, Buckingham EB, Criscitiello MF, Flajnik MF. Emergence of the acute-phase protein hemopexin in jawed vertebrates. *Mol Immunol.* 2010 Nov-Dec;48(1-3):147-52.
- Dube DH, Bertozzi CR. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov.* 4, 477-88 (2005).
- Gorman PM, Yip CM, Fraser PE, Chakrabarty A. Alternate aggregation pathways of the Alzheimer beta-amyloid peptide: Abeta association kinetics at endosomal pH. *J Mol Biol.* 2003 Jan 24;325(4):743-57.
- Granovsky M, Fata J, Pawling J, Muller WJ, Khokha R, Dennis JW. Suppression of tumor growth and metastasis in *Mgat5*-deficient mice. *Nat Med.* 6, 306-12 (2000).
- Hao P, Guo T, Sze SK. Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography. *PLoS One.* 2011 Feb 23;6(2)
- Herrero M, Ibañez E, Cifuentes A. Capillary electrophoresis-electrospray-mass spectrometry in peptide analysis and peptidomics. *Electrophoresis.* 2008 May;29(10):2148-60
- Kim YJ, Varki A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj J.* 14, 569-76 (1997).
- Kolte D, Bryant J, Holsworth D, Wang J, Akbari P, Gibson G, Shariat-Madar Z. Biochemical characterization of a novel high-affinity and specific plasma kallikrein inhibitor. *Br J Pharmacol.* 2011 Apr;162(7):1639-49.
- Kühl U, Schultheiss HP. Myocarditis in children. *Heart Fail Clin.* 2010 Oct;6(4):483-96.
- Kühl U, Schultheiss HP. Viral myocarditis: diagnosis, aetiology and management. *Drugs.* 2009 Jul 9;69(10):1287-302.
- Larsen R, Gozzelino R, Jeney V, Tokaji L, Bozza FA, Japiassú AM, Bonaparte D, Cavalcante MM, Chora A, Ferreira A, Marguti I, Cardoso S, Sepúlveda N, Smith A, Soares MP.

- A central role for free heme in the pathogenesis of severe sepsis. *Sci Transl Med*. 2010 Sep 29;2(51):51.
- Laurell E, Beck K, Krupina K, Theerthagiri G, Bodenmiller B, Horvath P, Aebersold R, Antonin W, Kutay U. Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell*. 2011 Feb. [Epub ahead of print]
- Ludwig M. Are neuropeptides brain hormones? *J Neuroendocrinol*. 2011 Apr;23(4):381-2.
- Lv H, Havari E, Pinto S, Gottumukkala RV, Cornivelli L, Raddassi K, Matsui T, Rosenzweig A, Bronson RT, Smith R, Fletcher AL, Turley SJ, Wucherpfennig K, Kyewski B, Lipes MA. Impaired thymic tolerance to α -myosin directs autoimmunity to the heart in mice and humans. *J Clin Invest*. 2011 Mar 23. [Epub ahead of print]
- Mauk MR, Smith A, Grant Mauk A. An alternative view of the proposed alternative activities of hemopexin. *Protein Sci*. 2011 Mar 14. [Epub ahead of print]
- Menschaert G, Vandekerckhove TT, Baggerman G, Schoofs L, Luyten W, VanCriekeing W. Peptidomics coming of age: a review of contributions from abioinformatics angle. *J Proteome Res*. 2010 May 7;9(5):2051-61.
- Onat A, Can G, Rezvani R, Cianflone K. Complement C3 and cleavage products in cardiometabolic risk. *Clin Chim Acta*. 2011 Mar 23. [Epub ahead of print]
- Rezkalla SH, Kloner RA. Influenza-related viral myocarditis. *WMJ*. 2010 Aug;109(4):209-13.
- Stensaeth KH, Hoffmann P, Fossum E, Mangschau A, Sandvik L, Klow NE. Cardiac magnetic resonance visualizes acute and chronic myocardial injuries in myocarditis. *Int J Cardiovasc Imaging*. 2011 Feb 24.
- Taylor CL, Eckart RE. Chest pain, ST elevation, and positive cardiac enzymes in an austere environment: Differentiating smallpox vaccination-mediated myocarditis and acute coronary syndrome in operation Iraqi freedom. *J Emerg Med*. 30
- Taylor-Papadimitriou J, Epenetos AA. Exploiting altered glycosylation patterns in cancer: progress and challenges in diagnosis and therapy. *Trends Biotechnol*. 12, 227-33 (1994).
- ter Weeme M, Vonk AB, Kupreishvili K, van Ham M, Zeerleder S, Wouters D, Stooker W, Eijnsman L, Van Hinsbergh VW, Krijnen PA, Niessen HW. Activated complement is more extensively present in diseased aortic valves than naturally occurring complement inhibitors: a sign of ongoing inflammation. *Eur J Clin Invest*. 2010 Jan;40(1):4-10. Epub 2009 Oct 15.
- van den Broek I, Sparidans RW, Schellens JH, Beijnen JH. Sensitive liquid chromatography/tandem mass spectrometry assay for absolute quantification of ITH4-derived putative biomarker peptides in clinical serum samples. *Rapid Commun Mass Spectrom*. 2010 Jul 15;24(13):1842-50.
- Wu C, Parrott AM, Fu C, Liu T, Marino SM, Gladyshev VN, Jain MR, Baykal AT, Li Q, Oka S, Sadoshima J, Beuve A, Simmons WJ, Li H. Thioredoxin 1-Mediated Post-Translational Modifications: Reduction, Transnitrosylation, Denitrosylation and Related Proteomics Methodologies. *Antioxid Redox Signal*. 2011 Apr 1.

Oligosaccharides characterization by dansyl labelling and integrated mass spectrometry techniques

Chiara Giangrande¹, Maura Melchiorre¹, Gennaro Marino^{1,2} and Angela Amoresano^{1,2}

¹Department of Organic Chemistry and Biochemistry; University of Naples “Federico II”, Naples, Italy

²School of Biotechnology Sciences

To whom correspondence should be addressed: Dr. A. Amoresano, Department of Organic Chemistry and Biochemistry, Federico II University of Naples, Monte S. Angelo, via Cynthia 4, 80126 Naples Italy.

Phone: +39081674114, Fax: +39081674313

Email: angamor@unina.it

ABSTRACT. This work aims at the development and set up of a new methodological platform based on the integration of affinity chromatography, a novel derivatization procedure and multistage mass spectrometry (MSⁿ) for the glycan characterization. In order to improve ionization an aromatic moiety and a charged basic group are introduced by dansyl derivatization of carbohydrate by using dansylhydrazine. This procedure has been set up firstly on synthetic oligosaccharides, revealing that dansyl derivatization improves fragmentation properties both in MALDI-TOF/TOF and in ESI-MS/MS experiments. This derivatisation method has been successfully applied to the glycan population from a complex matrix made of oligosaccharides extracted by white and yolk egg glycoproteins. A sub-glycoproteome characterisation was achieved by analysing glycopeptide component integrating ConA affinity purification and LC-MS/MS protein identification.

INTRODUCTION. Post-translational modifications (PTMs) play a critical role in cellular functioning and they vary in response to environmental stimuli, signalling modulators or development of diseases. Thus PTMs can affect biological functions finely tuning cellular mechanisms. Among PTMs, glycosylation is a co- and post-translational modification through which oligosaccharide chains are covalently attached to proteins. Glycoproteins are ubiquitous proteins characterized by the presence of oligosaccharide moieties. They are synthesised by any living organisms, from eubacteria to eukaryotes(1) and as complex mixtures of glycosylated variants (glycoforms). The micro-heterogeneity is due to the incomplete nature of glycosylation and deglycosylation reactions (2) and depends on cell type and physiological status. Anomalous pattern of glycosylation may be caused by several pathophysiological conditions or even by developmental events (3). Nowadays mass spectrometry is answering to most of the challenges connected to the study of glycans: the problem of glycans heterogeneity has been overcome by mass spectrometry methodologies, which allow to achieve qualitative and semi-quantitative information on oligosaccharides structures present in the mixture, with high sensitivity (21 and references therein).

Analysis of glycoproteome can be performed by treating the entire protein extract with endoglycosidases so that oligosaccharides and peptides can be analysed separately. Experimental procedures are needed to explore glycan sequences and their function. Such procedures are difficult due to the fact that glycan synthesis is not a template-driven and thus no mechanism amplifying glycan structures is available. Moreover carbohydrates detection is not an easy task since no chromophores are available so far. In order to facilitate their detection, chromophores are generally introduced at the reducing end of oligosaccharides. Hydrazine derivatives are generally used to derivatize carbohydrates and their reaction leads up to hydrazone derivatives that, if needed, can be reduced to open ring structures (4). Derivatized sugars can be analysed by mass spectrometry techniques in order to produce structural information. Both ESI-MS and MALDI-MS are generally used for oligosaccharides analysis. Derivatization strongly affects ionization efficiency by introducing additional charges and giving sugars a more hydrophobic

character. Moreover, labels influence glycans fragmentation behaviour and help the assignment of fragmentation spectra by tagging the reducing end (5).

Our group has already reported a new approach involving dansyl derivatives to selectively label specific PTMs (6, 7, 8, 9, 10, 11, 12). The dansyl-derivatization introduces an hydrophobic and aromatic moiety and basic secondary nitrogen into the molecule thus enhancing ionization efficiency. Furthermore the dansyl moiety shows a distinctive fragmentation pattern with product ions at m/z 170 and 234 in MS2 and the diagnostic m/z 234 \rightarrow 170 fragmentation in the MS3 mode (12). Recently we demonstrated that peptide dansylation resulted to be a rapid and convenient method to improve the performances in MALDI-MS/MS analysis at proteomic scale (13).

Here, we suggest a new derivatisation protocol of glycans reducing ends by dansylhydrazide. Once set up, the procedure was successfully applied to the characterisation of oligosaccharides from a complex matrix such as egg yolk and white.

EXPERIMENTAL SECTION. Protein extraction from egg white and yolk. Eggs were purchased at the local market. White and yolk were separately diluted in 10 mL of MilliQ water and centrifuged at 13000 rpm for 30 min, in order to separate the soluble fraction from the other components. This fraction was then subjected to ammonium sulphate precipitation. 76 g of ammonium sulphate (Sigma Aldrich) were added to 1 L of solution. After 12-16 h at 4°C the solution was centrifuged for 30 min at 13000 rpm in order to induce proteins precipitation. The excess of salts was removed using Centricon (Millipore).

Protein concentration determination. Egg protein concentration was determined by Bradford assay method, using BSA as standard. Known amounts of BSA were diluted in 800 μ L of H₂O and then mixed to 200 μ L of Coomassie Brilliant Blue. 5 different BSA concentrations were determined by measuring absorbance at 595 nm. Determined absorbance data were interpolated on the calibration curve, allowing the determination of protein concentration in the different samples.

Reduction and alkylation. Samples were dissolved in 200 μ L of denaturation buffer (guanidine 6 M, TRIS 0.3 M, EDTA 10 mM, pH 8). Reduction was carried out by using a 10:1 DTT : cysteins molar ratio. DTT was purchased from Fluka. After incubation at 37°C for 2 h, iodoacetamide (Fluka) was added to perform carboxyamidomethylation using an excess of alkylating agent of 5:1 with respect to the moles of thiolic groups. The mixture was then incubated in the dark at room temperature for 30 min. The alkylation reaction was stopped by addition of formic acid, in order to achieve an acidic pH. The excess of salts and reagents was then removed by gel filtration on PD-10 columns (GE Healthcare). Elution was performed using ammonium bicarbonate (Fluka) 10 mM buffer. The collected fractions were analyzed by a spectrophotometer Beckman DU 7500, measuring the absorbance at 220 and 280 nm. Protein containing fractions were collected and concentrated in a centrifuge Speed-Vac.

Trypsin digestion. Digestion was carried out in AMBIC 10 mM buffer using trypsin (Sigma Aldrich) at a 50:1 trypsin : protein mass ratio (w/w). The sample was incubated at 37°C for 12-16 h.

Concanavalin A affinity chromatography. The peptide mixtures were incubated for 2 h with 500 μ L of ConA Sepharose resin (Amersham Biosciences), which had been activated and washed in a 20mM Tris-HCl buffer; 0,5M NaCl; 1mM CaCl₂ at 4°C and pH 7.4. After 2 h the ConA containing Eppendorf was centrifuged at 3000 rpm at 4°C. The unbound peptides were removed in the supernatant. The resin was then washed six times with the same buffer described before. The elution phase was based on methyl- α -D-mannopyranoside (Sigma Aldrich) 0,25 M competition with glycopeptides and was performed for 30 min at 4°C. This procedure was repeated twice in order to be sure of having eluted all the glycopeptides bound to ConA.

RP-HPLC. Both the peptides and the glycopeptides needed a desalting step in order to remove salts and methyl- α -D-mannopyranoside.

Both the unbound and eluted peptides were desalted by HPLC on a C-18 resin. HPLC was performed on an Agilent chromatograph, at a flow rate of 0,2 ml/min and a linear gradient from 5% B (95%

acetonitrile, 5% H₂O e 0,07% trifluoroacetic acid) to 95% B in 1 min, allowing the fast elution of the unfractionated peptides.

Deglycosylation. Egg glycopeptides were treated with PNGase F in 50 mM ammonium bicarbonate. Reactions were performed for 12-16 h at 37°C. Deglycosylation was carried out also on unbound peptides, in order to release the glycans that had not been recognized by ConA.

Oligosaccharide purification. For glycan purification, the sample was loaded on a C-18 silica based bonded phase (Sep-pak, Waters). The column was activated with 10 mL methanol and equilibrated with 10 mL isopropanol, 10 mL H₂O and 10 mL acetic acid 5%. The oligosaccharide component was eluted in 3 mL of 5% acetic acid as eluent whereas peptides were collected in 3 mL of 40% isopropanol and 5% acetic acid. Alternatively an RP-HPLC step was performed by using a C-18 column and a gradient profile as follows: 7 min at 5% B, 5 to 95% B in 2 min and 10 min at 95% B.

LC-MS/MS. Peptides were analyzed by a HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The capillary column works at a flow of 4 µL/min, concentrating and washing the sample in a 40 nL enrichment column. The sample was then fractionated on a C18 reverse-phase capillary column (75 µm~43 mm in the Agilent Technologies chip) at flow rate of 300 nl/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min. Data were acquired through MassHunter software (Agilent Technologies). Proteins identification was achieved by using Mascot software (Matrixscience), with a tolerance of 10 ppm on peptide mass, 0.6 Da on MS/MS, and choosing cysteine carbamidomethylation as fixed modification. Variable modifications were methionine oxidation, glutamine conversion in pyro-glutamic acid, and asparagine deamidation.

Oligosaccharides derivatization. Glycans were subjected to chemical derivatization with dansyl-hydrazine. Lyophilized sugars were dissolved in 3-6 µL of water, to which a solution containing 200

mM dansyl-hydrazine in 90% acetonitrile (Romil) and 10% TFA (Carlo Erba) was added. The reaction was carried out in the dark at 37°C for 3 h.

MALDI-TOF/TOF analysis of glycans. Oligosaccharides were analysed on a 4800 Plus MALDI-TOF/TOF (Applied Biosystems). Samples were mixed on MALDI plate to the matrix consisting of a solution of 20 mg/mL DHB (Fluka), whose preparation consisted in the resuspension of DHB in water and acetonitrile 10:1 (v/v). The instrument was calibrated using a mixture of standard peptides (Applied Biosystems). Spectra were register even in reflector positive or negative mode. MS/MS spectra were performed with CID using air as collision gas. Spectra were manually interpreted and checked against GlycoWorkbench software (14).

Q-TRAP analysis of glycans. Supplementary MS experiments were performed on a 4000 Q-Trap system (Applied Biosystems). A micro-ionspray source was used at 2 kV with liquid coupling, with a declustering potential of 50 V, using an uncoated silica tip (o.d. 150 µm, i.d. 20 µm, tip diameter 10 µm) from New Objectives. In product ion scanning the scan speed was set to 4000 Da/s, the resolution of Q1 was set to “low” and the best collision energy was calculated as 35.

RESULTS AND DISCUSSION. In this study we provided a new methodology for the full characterization of glycome in a complex sample probed by a proteomics approach. As a valid step for enrichment, selecting specific glycoforms and simplifying the initial peptidic mixture, concanavalin A affinity chromatography was applied to complex matrix from egg fractions.

Concanavalin A affinity chromatography on egg proteins. White and yolk fractions were separated and proteins were precipitated in ammonium sulphate as described. SDS-PAGE was carried out on 50 µg of proteins deriving from both the white and the yolk to verify the correctness of separation procedure. Aliquots of the two samples were reduced, carbamidomethylated and trypsin digested. Concanavalin A affinity chromatography was performed on both the digested fractions (15).

Characterization and derivatization of oligosaccharides. Oligosaccharides characterization is arousing a particular interest even being a challenging analytical problem. Glycan molecules present extremely various structures as concerning the type of monosaccharides they are built of, the branching, the anomericity of the glycosydic linkage and the global structure that the carbohydrate moieties confer to protein. Nowadays mass spectrometry methodologies resolve most of the challenges connected to the study of glycans such as the glycoforms heterogeneity achieving qualitative and quantitative information on oligosaccharides structures present in the analyte, with enormous sensitivity (21). MALDI-TOF mass spectrometry technique was able to detect any difference between the analytes in terms of speed and sensitivity in the analyses.

Oligosaccharides characterization by MALDI-MS was achievable on the basis of the m/z value together with the knowledge of oligosaccharides biosynthesis. Hypothetical structures were then confirmed through tandem mass spectrometry. MS/MS spectra were interpreted using the software Glycoworkbench (14), which simulates glycans fragmentation and compares the theoretical fragments with the experimental ones.

MALDI-TOF spectra of oligosaccharides, released from protein backbone as described, were recorded in reflector mode. Peaks were manually attributed and then confirmed by MALDI-MS/MS.

The bound or the unbound ConA fractions both from yolk and white were treated with PNGase F. Oligosaccharides were purified from peptides by RP-HPLC and analysed by MALDI-MS in reflectron positive and negative mode. In positive ion mode oligosaccharides usually ionize as MNa^+ but sometimes we noticed the presence of sodium adducts, especially when the structure contains sialic acid residues. It is quite useful to register negative spectra in order to detect oligosaccharides bearing sialic acid residues due to ionize as MH^- of sugars. Fig. 1 shows the mass spectra in positive ion mode of the ConA bound and unbound fraction from egg white. The attribution of each peak is reported in Table 1. The glycans detected belonged to the complex type as confirmed by MS/MS data. However, it should be noted that some peak could be attributed to more than one structure. The MS/MS analysis let us to

confirm the presence of isobaric structures or to discriminate between the different structure only for the most abundant species. As an example, the peak occurring at m/z 1542,5 was identified as a biantennary glycan carrying a GlcNAc bisecting rather than as a triantennary complex structure. The heterogeneity in the peak attribution still remained for the less abundant ones. The main differences between the fractions consisted in the higher branching of unbound oligosaccharides.

Fig. 2 displays the positive ion mode mass spectra of ConA bound and unbound fractions of yolk. In both fraction the presence of sialylated structures, which were not detected in the white sample can be observed. These structures were identified in positive ion mode due to the presence of a negative charge for each sialic acid, which was neutralized by a sodium ion. For this reason we performed the analyses in negative ion mode in order to confirm the presence of sialic acid residues. A high heterogeneity was found in the MALDI spectra of the glycan moieties essentially belonging to the complex type. Moreover, differences between egg white and yolk could be appreciated and were essentially ascribed to the total absence of sialic acid in the oligosaccharides from white egg fraction.

In order to improve MS/MS glycan fragmentation a novel derivatisation method was proposed. Despite the fact that the MALDI ionization mechanism is not yet completely understood, it is widely acknowledged that the ionization efficiency depends on the chemical physical properties of the analyte. In particular, the presence of basic and/or hydrophobic groups (16) enhances the yields of the ionization process. Dansyl chloride (DNS-Cl) is a fluorescent reagent which has extensively used in biochemistry during the eighties (17) and used by our group to selectively label specific post-translational modification in a selective proteomics approach (11, 12, 18).

In a recent paper we demonstrated that dansylation labeling improves the MALDI ionization and fragmentation of peptides (13); moreover this derivatisation was applied to fragmentation of glycosylated peptides whose analysis is often problematic (19). On the basis of the experience recently made with dansyl chloride and other dansyl derivatives, we used a commercially available dansylated reagent, namely dansyl-hydrazine, to selectively label free oligosaccharides.

Our experimental procedure was set up on a synthetic oligosaccharide, maltoheptaose, by using dansyl-hydrazine as label. The reaction was carried out as described in the experimental section. The unlabeled sample was 2-fold more concentrated in comparison with dansylated samples, because of the dilution in the reagent solution. To monitor the effects of glycan dansylation, we diluted the unlabeled sample with a volume of ACN and we used it as control for the further experiments. The analyses were carried out by MALDI-TOF mass spectrometry where the labelled oligosaccharides showed an increment of 247.3 Da compared to the unmodified signal. Tandem mass spectrometry was performed on both the labelled and unlabelled oligosaccharide. Signals were assigned according to the nomenclature proposed by Domon and Costello (20).

Fig.3 shows the MS/MS spectra of maltoheptaose registered on a MALDI-TOF/TOF mass spectrometer where the DNS-sugar undergoes a loss of 234 Da, corresponding to the dansyl unit. We noticed the presence of Y and B ions, as well as cross-ring cleavages. As shown in the figure, the intensity strongly increased for the labelled specie. The difference in ionization efficiency could be appreciated in the difference in quality of the MS/MS data between dansylated and native oligosaccharides because the fragmentation profiles are modified. The fragmentation pattern of dansylated oligosaccharide exhibited enhanced B-ion and C-ion series in MS/MS experiments (Fig. 3B) compared to the native ones (Fig. 3A). This improvement in fragmentation could be attributed to the presence of the dansyl fluorescent moiety at the reducing end. This behavior was in agreement with previous data where dansyl moiety was added at the N-terminus of different peptides (13).

The same analysis was performed by using an hybrid Q-Trap instruments and the results are reported in Fig.4. As shown, the ESI/MS/MS spectrum of the DNS- oligosaccharide reveals a better fragmentation, when compared to the unmodified one. Even in this case both Y and B ions were present, but cross-ring cleavages were less frequent than for MALDI-MS.

Once set up, the dansyl labelling procedure was applied to the characterisation of glycans in a real sample, namely the oligosaccharides released from egg glycoproteins. The N-glycans enzymatically

released from glycoproteins were derivatised and subjected to MS/MS analysis in order to assign the correct structure. As an example, Fig 5 showed the MALDI-MS/MS spectrum of the precursor ion at m/z 1419.39, compared to the dansyl-labelled form.

The oligosaccharide was identified as a high mannose structure on the basis of the interpretation of the fragmentation spectrum. Similar analyses were carried out on the other glycans thus confirming previous results. As a whole the data obtained by MALDI-MS/MS analyses of glycans are summarised in Table 2. It should be worth noting that DNS-derivatives are characterized by higher ionization efficiency than the unlabelled ones, producing better spectra with a higher S/N ratio with respect to the unmodified ones (Fig. 5B). Thus, the derivatization protocol by dansyl-hydrazine was effective in the improvement of ionization efficiency, introducing a strong basic group. This labelling strategy not only increased peaks resolution in MS/MS spectra, but also introduced the possibility for further analysis of sugars by exploiting the potential of precursor ion scan analysis (manuscript in preparation) as already performed for the selective identification of other PTMs (12).

Protein identification. Identification of proteins and the direct localization of N-glycosylation sites was achieved by LC-MS/MS analysis. In ConA bound fractions proteins identification was obtained exclusively by taking in account the peptides containing the N-glycosylation consensus sequence. As described before (15), in order to identify not glycosylated peptides and glycopeptides bearing a different glycosylation pattern (not recognized by concanavalin A) an LC-MS/MS experiment was performed on the unbound fraction.

Proteins identification was carried out by using a bioinformatics tool, namely Mascot MS/MS ion search program through the NCBI database, specifying *Gallus gallus* as taxonomy. The NCBI identified proteins were then searched against Swiss-prot database, in order to univocally recognize each protein. For the identification, we considered asparagine deamidation as a possible variable modification that took place in the consensus sequence N-X-S/T after glycosidase treatment. The data obtained are summarized in Table 2. It should be noted that several glycopeptides were detected in the unbound

fraction from egg samples and identified by the diagnostic increment of 1 Da respect to the theoretical mass. Some proteins, as the vitellogenin-2, ovalbumin, ovotransferrin and clusterin, are indicated to be glycosylated on the basis of similarity with the counterpart from other species. For several egg proteins, the glycosylation sites are indicated in the SwissProt database as putative N-glycosylation sites. The MS analyses performed in this work, led to the direct identification of N-glycosylation sites for several egg proteins. To the best of our knowledge, these results represent the first data on the fine localisation of N-glycosylation sites thus complementing the data already available on the characterisation of egg proteome.

CONCLUSIONS. As a whole, in this work an optimized dansylation protocol for glycoproteomic studies by MALDI-MS/MS analysis is reported. The presence of the aromatic dimethylaminonaphthalene group at the reducing end increased the glycan fragmentation leading to a better coverage of the oligosaccharide sequence. This tendency could be rationalized considering that the UV-adsorbing naphthalene group in the reducing end could increase the energy transfer from the laser in PSD-like fragmentation, favoring the glycan bond destabilization/protonation as observed for peptides. Glycan dansylation was demonstrated to be a rapid and cost-effectively valuable method useful to optimize the ionization efficiency of oligosaccharides. Well established techniques were conjugated with novel derivatisation procedure thus providing a new method for the glycoproteome investigation.

ACKNOWLEDGMENT. This work was supported by Italian MIUR grants, PRIN 2008, and FIRB Italian Human ProteomeNet Project 2007




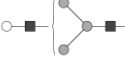





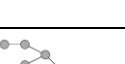
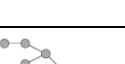

REFERENCES

- (1) Spiro RG. *Glycobiology*. 2002, 12, 43R-56R.
- (2) Qiu R, Regnier FE. *Anal Chem*. 2005, 77, 2802-9.
- (3) Dell A, Morris HR. *Science*. 2001, 291, 2351-6.

- (4) Lattová E, Perreault H. *Methods Mol Biol.* 2009, 534, 65-77.
- (5) Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wührer M. *Anal Bioanal Chem.* 2010, ;397, 3457-81.
- (6) Palmese A, De Rosa C, Marino G, Amoresano A. *Rapid Commun Mass Spectrom.* 2011, 15, 25, 223-31.
- (7) Chiappetta G, Corbo C, Palmese A, Galli F, Piroddi M, Marino G, Amoresano A. *Proteomics.* 2009 Mar;9(6):1524-37. Erratum in: *Proteomics.* 2009 , 9, 3220.
- (8) Amoresano A, Cirulli C, Monti G, Quemeneur E, Marino G. *Methods Mol Biol.* 2009, 527, 173-90, ix.
- (9) Amoresano A, Chiappetta G, Pucci P, Marino G. *Methods Mol Biol.* 2008, 477, 15-29.
- (10) Cirulli C, Chiappetta G, Marino G, Mauri P, Amoresano A. *Anal Bioanal Chem.* 2008, 392, 147-59.
- (11) Amoresano A, Chiappetta G, Pucci P, D'Ischia M, Marino G. *Anal Chem.* 2007, 79, 2109-17.
- (12) Amoresano A, Monti G, Cirulli C, Marino G. *Rapid Commun Mass Spectrom.* 2006, 20, 1400-4.
- (13) Chiappetta G, Ndiaye S, Demey E, Haddad I, Marino G, Amoresano A, Vinh J. *Rapid Commun Mass Spectrom.* 2010;24, 3021-32.
- (14) Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. *J Proteome Res.* 2008, 7, 1650-9.
- (15) Carpentieri A, Giangrande C, Pucci P, Amoresano A. *Eur J Mass Spectrom (Chichester, Eng).* 2010;16(1):123-49.
- (16) Mollé D, Jardin J, Piot M, Pasco M, Léonil J, Gagnaire V. *J Chromatogr A.* 2009, 1216, 2424-32.

- (17) Walker JM. *Methods Mol Biol.* 1994, 32, 329-34
- (18) Cirulli C, Marino G, Amoresano A. *Rapid Commun Mass Spectrom.* 2007
- (19) Napoli A, Aiello D, Di Donna L, Moschidis P, Sindona G. *J Proteome Res.* 2008, 7, 2723-32.
- (20) Domon B, Costello CE. *Glycoconjugate J.* 1988, 5, 397-409.
- (21) North SJ, Hitchen PG, Haslam SM, Dell A. *Curr. Opin. Struct. Biol.* 2009, 19, 498-506.

Table 1: list of the oligosaccharides structures detected by MALDI-MS, where BY=ConA bound fraction of yolk, UY=ConA Unbound fraction of yolk, BW=ConA bound fraction of white, UW=ConA unbound fraction of white.

m/z	Composition	Structure
1095.37 ^{BY}	HexNac ₂ Hex ₄	
1136.39 ^{BY}	HexNac ₃ Hex ₃	
1257.42 ^{BY,BW}	HexNac ₂ Hex ₅	
1298.42 ^{BY}	HexNac ₃ Hex ₄	
1339.44 BY,BW,UY,UW	HexNac ₄ Hex ₃	
1419.47 ^{BY,BW}	HexNac ₂ Hex ₆	
1460.50 ^{BY,BW,UY}	HexNac ₃ Hex ₅	
1501.53 BY,BW,UY,UW	HexNac ₄ Hex ₄	
1542.55 BY,BW,UY,UW	HexNac ₅ Hex ₃	
1581.52 ^{BY}	HexNac ₂ Hex ₇	
1622.55 ^{BW}	HexNac ₃ Hex ₆	
1663.58 ^{BY,BW}	HexNac ₄ Hex ₅	

1704.61 BY,BW,UY,UW	HexNac ₅ Hex ₄	
1745.63 BY,BW,UY,UW	HexNac ₆ Hex ₃	
1850.66 ^{BY,UY}	FucHexNac ₅ Hex ₄	
1866.66 BW,UY,UW	HexNac ₅ Hex ₅	
1905.63 ^{BY,UY}	HexNac ₂ Hex ₉	
1907.68 ^{BW,UW}	HexNac ₆ Hex ₄	
1948.71 ^{BW,UY,UW}	HexNac ₇ Hex ₃	
1954.67 ^{BY,UY}	HexNac ₄ Hex ₅ Sia	
2028.71 ^{BW,UY}	HexNac ₅ Hex ₆	
2067.69 ^{BY,UY}	HexNac ₂ Hex ₁₀	
2110.76 ^{UY,UW}	HexNac ₇ Hex ₄	
2151.79 ^{UY,UW}	HexNac ₈ Hex ₃	
2245.77 ^{BY,UY}	HexNac ₄ Hex ₅ Sia ₂	
2272.82 ^{UY,UW}	HexNac ₇ Hex ₅	





2303.81 ^{BY,UY}	FucHexNAc ₅ Hex ₅ Sia a	
2313.84 ^{UY,UW}	HexNAc ₈ Hex ₄	
2319.81 ^{BY,UY}	HexNAc ₅ Hex ₆ Sia	
2475.90 ^{UY,UW}	HexNAc ₈ Hex ₅	

Table 2: List of the proteins identified by LC-MS/MS analyses. N-glycosylation sites are marked with an asterisk after the asparagines residue. BY and BW indicate the bound fraction of egg yolk and white samples, respectively; UY and UW indicate the unbound fraction of egg yolk and white, respectively.

Protein	Sequence
Vitellogenin-1 (P87498)	MSFTCSFN ^K ^{BY,UY,UW} EVHINTSSAN [*] ITICPAADSSLLVTCNK ^{BY}
Ovomucoid (P01005)	FPN [*] ATDK ^{BY,BW,UY,UW} PMN [*] CSSYAN [*] TTSEDK ^{BY,BW,UY,UW} CNFCNAVVESN [*] GTLTSLSHFGK ^{BY,BW,UY,UW} LAAVSDVCSEYKPDCTAEDRPLCGSDN [*] K ^{BY,BW,UY,UW}
Vitellogenin-2 (P02845)	VQVFVTN [*] LTDSSK ^{BY} LLVDGAESPAN [*] ISLISAGASLWIHNENQGFALAAPGHGIDK ^{BY,UY}
Apovitellenin-1 (P02659)	NFLIN [*] ETAR ^{BY,UY,UW}
Serum albumin (P19121)	KQETTPINDN [*] VSQCCSQLYANR ^{BY}
Apolipoprotein B (Q197X2)	VASPLYN [*] VTWR ^{BY,BW} SLHN [*] STLGS�VPK ^{BY,BW} NQYGMTEVN [*] QTLK ^{BY,BW} INSPSNQILIPAMGN [*] ITYDFSEK ^{BY,BW} LNIPEADFSSQADLVNN [*] MTTEVEAGR ^{BY,BW} VASPFFTLSTQAEVHN [*] TTASANSPEFDTSLSAQATSK ^{BY,BW}
Complement C3 (Q90633)	FEIDHALSN [*] R ^{BY} AAN [*] LSDIVPNTSESEK ^{BY}
Ig mu chain C region* (P01875)	VNVSGTDWR ^{BY,UY} VTHN [*] GTSITK ^{UW}
Cathepsin D* (Q057441)	DTLQISN [*] ISIK ^{BY}
Arylsulfatase K* (Q5ZK90)	NDFLN [*] VSAPR ^{BY}
Complement factor B-like protease* (P81475)	TGN [*] WSTKPSCK ^{BY}
Ovotransferrin (P02789)	TAGWVIMPGLIHN [*] R ^{BY,BW} FGVN [*] GSEK ^{BW,UY,UW}
Ovalbumin-related protein Y (P10114)	SAN [*] LTISSVDNLMISDAVHGVMFMEVNEEGTEATGSTGAIGNIK ^{BY,BW,UW}
Ovoglycoprotein (Q8JIG5)	LN [*] ETCVVK ^{BY,UY,UW} HN [*] STLTHEDGQVVSMaelTHSDK ^{BY,BW}
Clusterin (P14018)	RYDDLlSAFQAEMLN [*] TSSLLDQLNR ^{UY,UW}
Ovalbumin (P01012)	YN [*] LTSVLMAMGITDVFSSSANLlSGISSAESLK ^{UW}

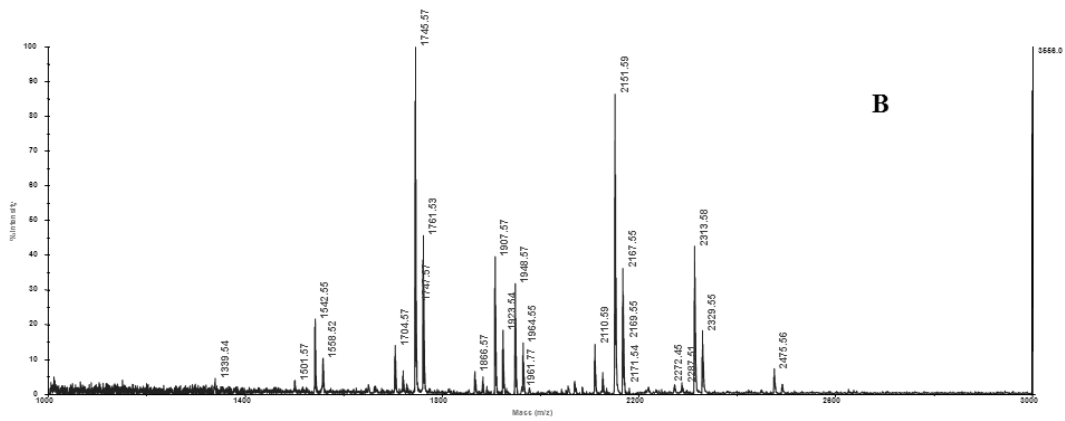
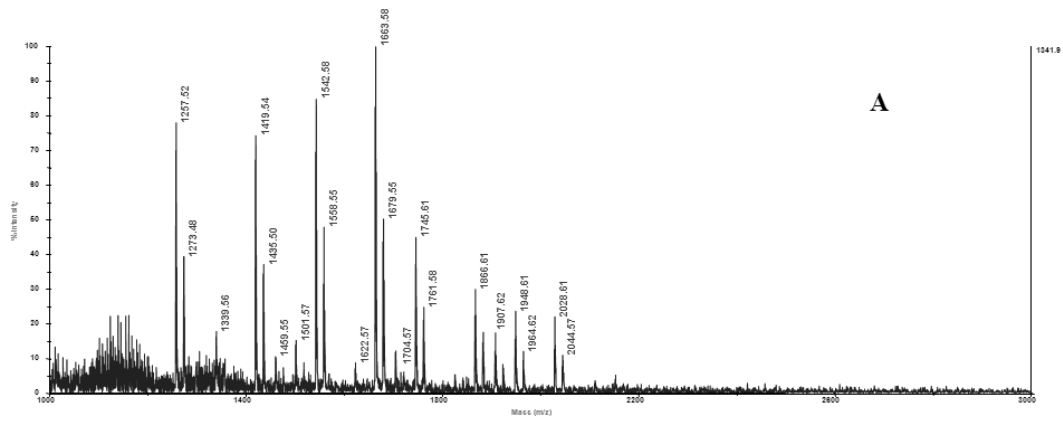


Fig. 1: positive ion mode MALDI-MS spectrum of ConA bound (panel A) and unbound fraction (panel B) oligosaccharides released from egg white glycoproteins.

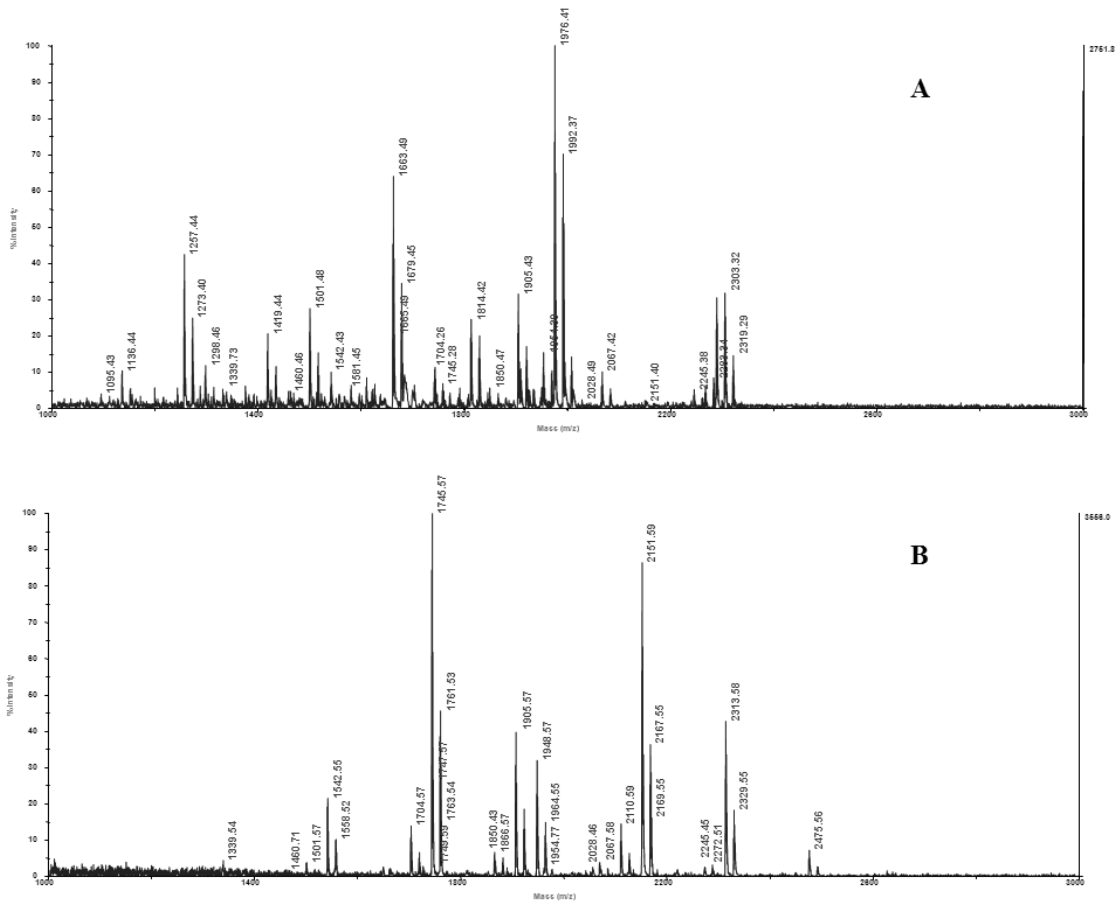


Fig. 2: positive ion mode MALDI-MS spectrum of ConA bound (panel A) and unbound fraction (panel B) oligosaccharides released from egg yolk glycoproteins.

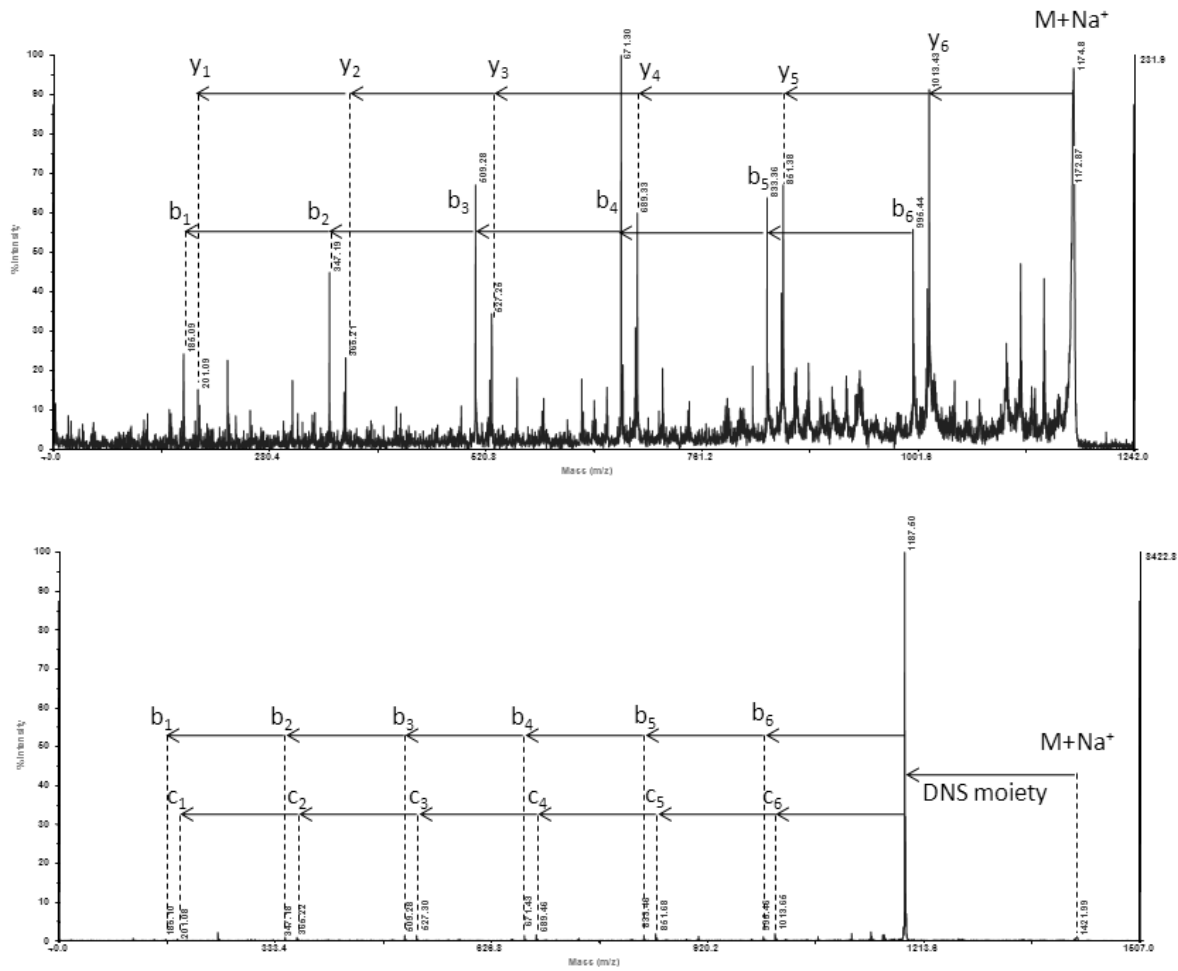


Fig. 3: positive ion mode MALDI-MS/MS spectrum of DNS-labelled maltoheptaose (B) compared to the unmodified one(A).

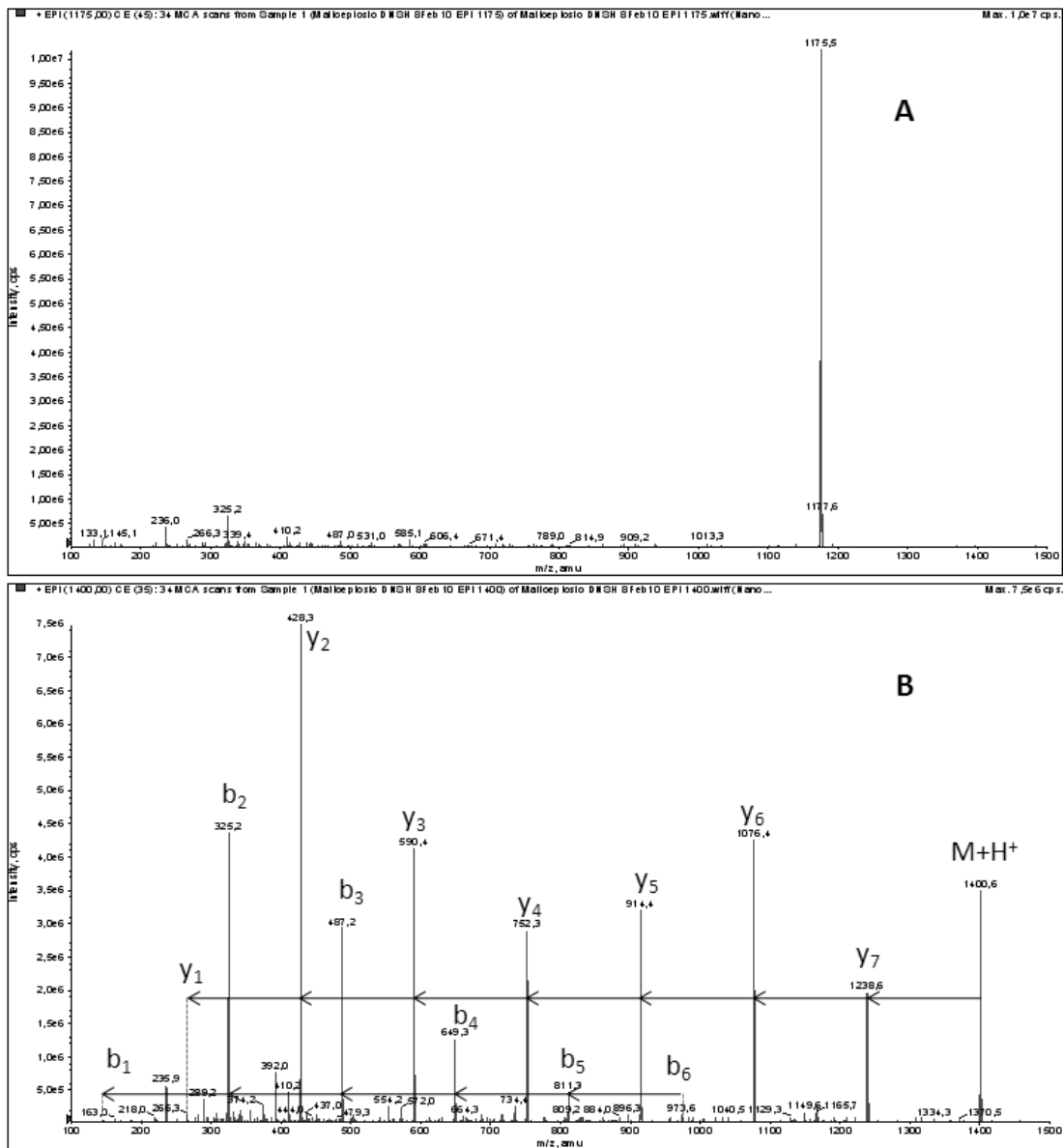


Fig. 4: ESI-MS/MS spectrum of DNS-labelled maltose (B) compared to the unmodified one(A) performed on a 4000 Q-TRAP.

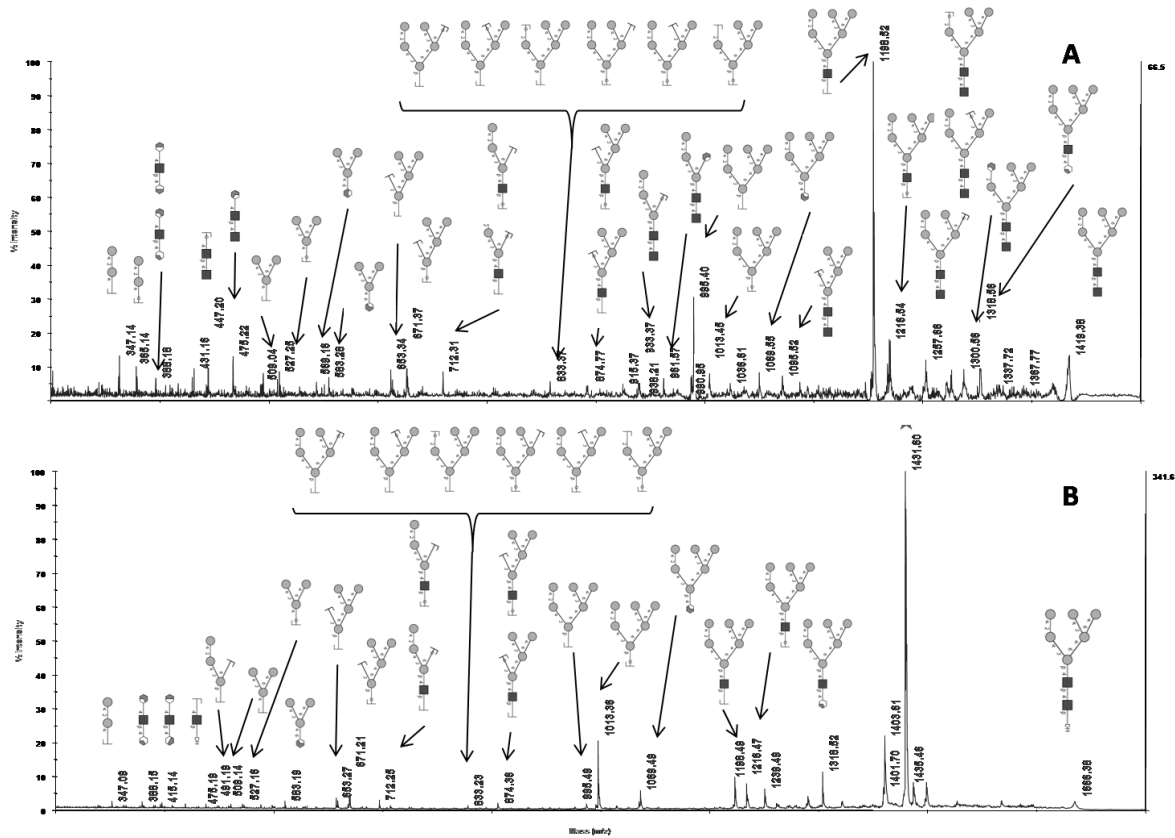


Fig. 5: positive ion mode MALDI-MS/MS spectrum of DNS-labelled oligosaccharide at m/z 1666.38 released from egg white ConA eluted fraction (B) compared to the unmodified one(A)

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Innate immunity probed by lipopolysaccharides affinity strategy and proteomics

Chiara Giangrande, Lucia Colarusso, Rosa Lanzetta, Antonio Molinaro, Piero Pucci and Angela Amoresano*

5 Received (in XXX, XXX) Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

Lipopolysaccharides (LPSs) are ubiquitous and vital components of the cell surface of Gram-negative bacteria that have been shown to play a relevant role in the induction of the immune system response. In animal and plant cells innate immune defenses toward microorganisms are triggered by the perception of pathogen associated molecular patterns (PAMPs). These are conserved and generally indispensable microbial structures such as LPSs that are fundamental in the Gram negative immunity recognition.

This paper reports the development of an integrated strategy based on lipopolysaccharide affinity methodology that represents a new starting point to elucidate the molecular mechanisms elicited by bacterial LPS and involved in the different steps of innate immunity response. Biotin-tagged LPS was immobilized on streptavidin column and used as a bait in an affinity capture procedure to identify protein partners from human serum specifically interacting with this effector. The complex proteins/lipopolysaccharide was isolated and the protein partners were fractionated by gel electrophoresis and identified by mass spectrometry.

This procedure proved to be very effective in specifically binding proteins functionally correlated with the biological role of LPS. Proteins specifically bound to LPS essentially gathered within two functional groups, regulation of the complement system (Factor H, C4b, C4BP and alpha 2 macroglobulin) and inhibition of LPS induced inflammation (HRG and Apolipoproteins).

The reported strategy might have important applications in the elucidation of biological mechanisms involved in the LPSs-mediated molecular recognition and anti-infection responses.

Introduction

In a biological context, the term immunity refers to the set of events and biological cascades that protect organisms against infectious diseases. This protection system is based on complex interconnections between cells and molecules whose synergic action determines immune response. The main physiological function of immune system is the protection against infectious agents or simply against exogenous elements. The ability to distinguish “self” from “non-self” is critical to prepare an effective immune response (1).

Immune system response can be classified into two categories: adaptive immunity and innate immunity. The adaptive immune system is based on specialized processes whose function is to eliminate external threats during the late phase of infection with the generation of immunological memory. Specificity is developed by clonal gene rearrangements form a range of antigen-specific receptors on lymphocytes (2). Only vertebrates are provided with this type of immunity, allowing the organism to recognize and remember the “non-self”. Innate immune system is necessary to activate adaptive immune system. It acts in a non-specific manner and does not confer any memory of the challenge. It is widespread amongst all animals and even in plants, and its main function is to provide immediate protection against pathogens. Innate immunity relies on specific cells as primary mediators, namely phagocytic cells and antigen presenting cells (APCs), such as granulocytes, macrophages, and dendritic cells (DCs) (3). Innate immunity mechanisms do not have specific pathogen targets, the principal targets are evolutionarily conserved molecular structures, termed “pathogen

associated molecular patterns” (PAMPs) whose expression is common to different pathogens. These structures are generally recognized by receptors on the membrane of the host cells that are responsible for the immune response. Such structures are called “pattern recognition receptors” (PRR). PRRs are constitutively expressed in the host on all cells of a given type (4). Lipopolysaccharide (LPS) is a glycolipid, generally referred to as endotoxin in Gram negative bacteria, belonging to PAMP family. It is a potent inducer of the innate immune system and the main cause of septic shock (5).

The recognition of this endotoxin is achieved by the interplay of a variety of proteins, either receptors or serum secreted proteins. The cascades of events following LPS infection has been investigated in numerous studies. However, most of them focused on the interaction of single specific proteins with LPS, such as the LPS-binding protein (LBP), whose mechanism of action was clarified in many works (6; 7). Further studies showed that other proteins possess LPS-binding domains (8; 9; 10). Intracellular pathways elicited by LPS infection was investigated by *in vitro* and *in vivo* analyses on different cell types, ranging from immune cells to epithelial cells (11; 12).

Toll-like receptor 4 (TLR-4) is the signal transducing receptor of lipopolysaccharides; lipid A is the real effector required to activate TLR4 signalling pathway in conjunction with a soluble co-receptor protein myeloid differentiation protein 2 (MD2), which directly and physically binds to LPS (13; 14). Entry of LPS or lipid A into the blood stream is responsible for the onset of septic shock (15). Aggregates of LPS bind to certain plasma proteins, such as albumin, LBP and soluble CD14 to facilitate the interaction with host cells, via TLR4. The complex of protein bound LPS, TLR4, and MD2, activates two intracellular

pathways (16): the MyD88-dependent pathway, responsible for early-phase NF- κ B and MAPK activation, which controls the induction of proinflammatory cytokines TNF- α , IL-1, IL-6, and the MyD88-independent, TRIF-dependent pathway that activates IRF3, which is necessary for the induction of IFN- β - and IFN-inducible genes (17). The activation of the TLR4-MD2 complex triggers the induction of inflammatory cytokines acting as endogenous mediator of infection, as well as the superoxide anion (O $_2^-$), hydroxyl radicals (OH), nitric oxide and antimicrobial peptides (18).

The low and balanced concentrations of these mediators and soluble immune response modulators leads to a resulting inflammation, one of the most important and ubiquitous aspect of the immune host defence against invading microorganisms. Beside these positive effects, an uncontrolled and massive immune response, due to the circulation of large amount of endotoxins, leads to the symptoms of the sepsis and of the septic shock. Therefore identification of all the molecular actors involved in the recognition cascade is of pivotal importance with particular emphasis on the differences in elicitation and reaction by different LPSs.

Numerous studies have been performed following pro-inflammatory expression patterns via microarray analyses (19) and the phage display method was used to capture new LPS-binding peptides (20).

Recently, proteomics approaches were addressed to the study of innate immunity. Comparative analysis of the human plasma proteome prior and after LPS injection was performed (21) and the human innate immunity interactome was investigated using 58 genes involved in transcriptional regulation of type I IFN tagged with the FLAG epitope following cells stimulation with LPS (22).

The aim of this study was the setting up of a new methodology for the development of lipopolysaccharides analyses to be used to capture LPS-interacting proteins in order to better elucidate the molecular mechanism of the innate immunity. Here, a strategy based on chemical manipulations, biochemical procedure and mass spectrometry techniques integrated in a functional proteomic workflow is proposed. The procedure was developed using an LPS derived from *Salmonella typhimurium* and the proteomic study was carried out by using human serum from a healthy donor as a model.

Materials and methods

LPS extraction

Dried cells of *Salmonella typhimurium* were extracted by the phenol/water method (23). The extracted phases were dialysed three times against distilled water and subjected to enzymatic digestions in order to remove nucleic acids and protein contaminants. Both water and phenol fractions were analysed through SDS-PAGE 13.5%.

LPS biotinylation

LPS biotinylation was achieved by a transesterification reaction with biotin-p-nitrophenylester (Sigma Aldrich). Lyophilized LPS was solubilised in 200 μ L pyridine (Romil), to which a certain amount of biotin-p-nitrophenylester was added. LPS: biotin-p-nitrophenylester ratio was calculated as 1:3 (weight:weight). The reaction was carried out in the dark for 2 hours at 80°C. The reaction mixture was dried under nitrogen and then the sample was dissolved in MilliQ water. The excess of reagents was removed by molecular exclusion chromatography on PD-10

columns (GE-Healthcare). LPS containing fractions were pooled and lyophilized. Unlabelled and biotinylated LPS were dissolved in Laemmli buffer, supplemented with 100 mM DTT and separated by SDS-PAGE in a 12.5% polyacrylamide gel. LPS was detected by silver nitrate staining for carbohydrates (Tsai Frash, 1982) or electrotransferred onto a PVDF membrane (Millipore). The membrane was blocked with 3% non fat dry milk, 1% BSA (Sigma Aldrich), in PBS containing 0.05% Tween 20 (Sigma Aldrich). Then it was incubated with 1:1000 streptavidin/HRP conjugate (Sigma Aldrich). Biotinylated sample was detected using SuperSignal West Femto Chemiluminescent Kit from Pierce.

LPS immobilization onto avidin beads

Biotin-LPS immobilization was carried out using 3 mg biotin-LPS/mL of settled avidin agarose resin (settled gel, Pierce). An aliquot of resin was extensively washed with binding buffer (0.1 M Na $_3$ PO $_4$; 0.3 M NaCl; pH 7.2) and incubated with biotin-LPS at 4°C overnight under shaking. Then the resin was washed three times with binding buffer.

Biotin-LPS and human serum interaction

The affinity experiment was preceded by a “precleaning step”. A protease inhibitors cocktail (Sigma Aldrich) was added to 5 mg human serum (Servizio Analisi, Policlinico), to a final concentration of 1 mM. This sample was diluted to a final volume of 600 μ L with binding buffer and incubated with 150 μ L of settled resin overnight at 4°C. The unbound fraction of precleaning step was therefore incubated with 150 μ L of biotin-LPS immobilized resin overnight at 4°C. Later on, the two aliquots of resin, that we term “control” and “test” were subjected to repeated washes with binding buffer. Elution was performed boiling the resin in Laemmli buffer and DTT.

SDS-PAGE was performed, loading on a 1.5 mm, 12.5 % gel all the samples deriving from the different steps of the fishing experiment. The gel was run at constant 25 mA for 1h. The gel was stained with colloidal Coomassie (Pierce).

In gel trypsin digestion

The analysis was performed on the Coomassie blue-stained protein bands excised from the gels. Gel particles were washed first with acetonitrile and then with 0.1 M ammonium bicarbonate. Protein samples were reduced incubating the bands with 10 mM dithiothreitol (DTT) for 45 min at 56 C. Cysteines were alkylated by incubation in 5 mM iodoacetamide for 15 min at room temperature in the dark. The bands were then washed with ammonium bicarbonate and acetonitrile. Enzymatic digestion was carried out with trypsin (12.5 ng/ μ L) in 50 mM ammonium bicarbonate buffer, pH 8.5. Gel particles were incubated at 4°C for 2h, in order to allow the enzyme to enter the gel. The buffer solution was then removed and a new aliquot of buffer solution was added for 18 h at 37 C. A minimum reaction volume, enough for complete gel rehydration was used. At the end of the incubation the peptides were extracted by washing the gel particles 0.1% formic acid in 50% acetonitrile at room temperature and then lyophilised.

LC-MS/MS analyses

Peptides mixtures were analyzed by LC-MS/MS, using a HPLC-

Chip LC system (Agilent 1200) connected to a Q-TOF 6520 (Agilent Technologies). Samples were diluted in 10 μ L 0.1% formic acid. After loading, the peptide mixtures were

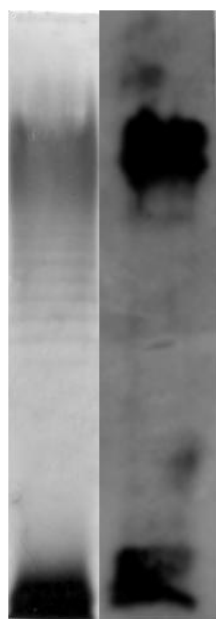


Fig. 1: SDS-PAGE of LPS after biotin-p-nitrophenylester derivatization detected by silver staining (on the left) and blotting with streptavidin/HRP incubation (on the right)

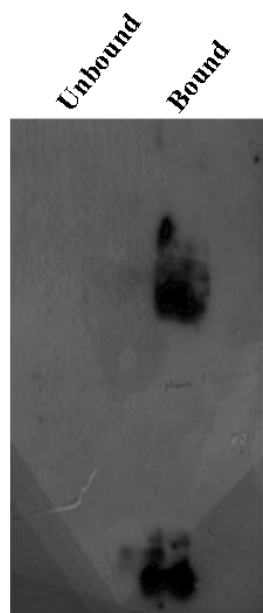


Fig. 2: immobilization procedure probed by blotting with streptavidin/HRP incubation. Two bands corresponding to LPS and Lipid A moiety were clearly detected in the bound fraction while no bands occurred in the unbound material

concentrated and washed at 4 μ L/min in 40 nL enrichment column with 0.2% formic acid in 2% acetonitrile. Fractionation was carried out on a C-18 reverse phase column (75 μ m x 43 mm) at a flow rate of 0.4 μ L/min with a linear gradient of eluent B (95% acetonitrile and 0.2% formic acid) in A (2% acetonitrile and 0.1% formic acid) from 7% to 80% in 51 min. Mass spectrometry analyses were performed using data dependent acquisition MS scans (mass range 300-2400 m/z), followed by MS/MS scans (mass range 100-2000 m/z) of the most intense ions of a chromatographic peak.

Raw data from LC-MS/MS were converted to mzData. The spectra were searched against the NCBI database (2006.10.17 version) using the licensed version of Mascot 2.1 (Matrix Science). The MASCOT search parameters were: taxonomy Homo sapiens; allowed number of missed cleavages 2; enzyme trypsin; variable post-translational modifications, methionine oxidation, pyro-glu N-term Q; peptide tolerance 10 ppm and MS/MS tolerance 0.6 Da; peptide charge, from +2 to +3.

Results

LPS were extracted by *Salmonella typhimurium* cells as described in Materials and Methods section and analysed by SDS-PAGE. The LPS fraction was found exclusively in the water phase as suggested by the presence of the typical ladder pattern in the gel analysis (data not shown).

LPS tagging

Immobilisation of the LPS represents the key step in the entire strategy to generate a labeled bait that could be linked to agarose beads. Biotinylation of LPS was chosen as the best suited derivatization process both for the simple modification reaction and to take advantage of the strong and specific interaction of biotin with avidin. Biotinylated LPS was analysed by SDS-PAGE following both silver staining and immunodetection with streptavidin/HRP conjugate. Figure 1 shows the corresponding gel. Staining procedures displayed the occurrence of two sample bands. The low molecular mass bands corresponds to the LipidA core component of LPS (R-type LPS) whereas the high molecular mass band represents the whole LPS molecule comprising the O-chain moiety (S-type LPS). Both components were responsive to western blot analysis demonstrating that they had been derivatised by biotin.

LPS immobilization

Biotin-LPS was conjugated to avidin agarose beads by incubation in 0.1 M Na_3PO_4 overnight. Both unbound and bound fractions were analyzed by SDS-PAGE followed by blotting and incubation with streptavidin/HRP conjugate. Fig. 2 shows the results of the immobilization procedure; two immunoresponsive bands corresponding to LPS and the Lipid A moiety were clearly detected in the bound fraction while no immunoresponsive bands occurred in the unbound material.

Identification of LPS-interacting proteins

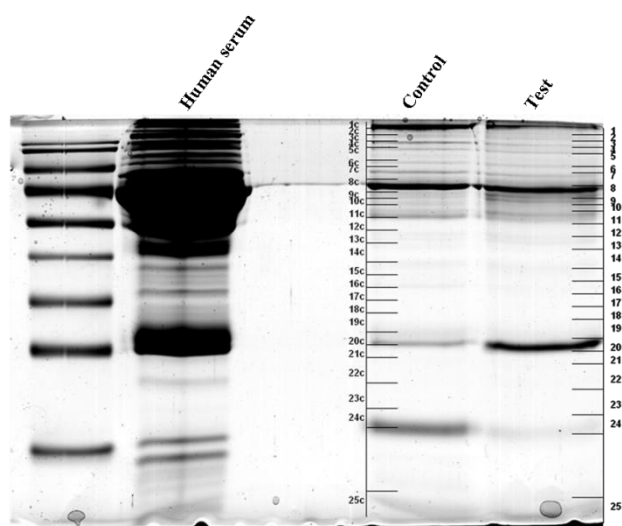


Fig. 3: 12,5% SDS-PAGE gel containing 15 µg of human serum together with control and sample bound fractions after elution in Laemmli buffer. The entire control and sample lanes of the gel were cut in 25 slices. Each slice was destained and trypsin digested, *in situ*.

The final step of the proposed strategy relies on the use of the immobilised LPS as bait in a functional proteomic experiment (24) aiming at capturing specific LPS interactors to investigate innate immunity molecular mechanisms in human serum. An aliquot of human serum proteins was incubated with underivatized avidin agarose beads overnight at 4°C as a pre-cleaning step to remove all the proteins that non specifically interact with agarose or avidin. The unbound fraction was then incubated with biotin-LPS avidin agarose beads and after

extensively washing in binding buffer both the bound fractions (control and sample) were eluted in Laemmli buffer and fractionated by SDS-PAGE gel. An aliquot of 15 µg of human serum was also loaded onto the gel (Fig.3). Due to the complexity of the gel patterns and the low resolution of 1D electrophoresis, several proteins can occur in the same gel band. Therefore, protein bands specifically present in the sample lane and absent in the control lane cannot be identified by simply comparing the two gel profiles. Thus the entire sample lane of the gel was cut in 25 slices. Each slice was destained and *in situ* trypsin digested. To check for non specific proteins, the same procedure was applied to the control lane. The resulting peptide mixtures were extracted from the gel and submitted to nanoLC/MS/MS analysis generating sequence information on individual peptides.

This information, together with the peptide mass values, was then used to search protein databases, leading to the identification of the protein components. The experiments were performed in two biological replicates and each protein mixture was run through LC/MS/MS three times. All proteins identified in both the sample and control lanes were discarded, while only those solely occurring in the sample lane and absent in the control lane were selected as putative LPS interactions. As further selection criteria, only proteins found in both biological replicates and identified with at least 2 significative peptides in MASCOT search were considered, providing a full list of putative LPS interactors. The results are summarised in Table 1.

Discussion

Lipopolysaccharides are major constituents of the outer membrane of Gram-negative bacteria and important molecules in the induction of the immune system response. The innate immune system constitutes the first line of defence against

Sample	MW	PROTEIN	ID	SCORE	PEPTIDES
3	~180 KDa	Alpha-2-macroglobulin	<u>P01023</u>	170	8
		factor H	<u>P08603</u>	44	2
4	~130KDa	ceruloplasmin	<u>P00450</u>	136	6
		phospholipase D	<u>P80108</u>	114	4
6	~100KDa	Coagulation factor II (Thrombin)	<u>P00734</u>	56	4
		S plasma protein	<u>P07225</u>	140	8
7	~80 KDa	histidine-rich glycoprotein precursor	<u>P04196</u>	54	2
		C4 binding protein C4bp	<u>P04003</u>	266	16
10	~60 KDa	Beta2-Glycoprotein I	<u>P02749</u>	91	5
		Proapolipoprotein A1	<u>P02647</u>	86	3
13	~50KDa	complement C4B	<u>P0C0L5</u>	92	3
18	~30KDa	Apolipoprotein D	<u>P05090</u>	56	2
19	~30KDa	Apolipoprotein A-IV precursor	<u>P06727</u>	74	4
24	~15KDa	Prolactin-inducible precursor	<u>P12273</u>	64	2

Table 1: list of putative LPS-interacting proteins after *in situ* digestion of the bands from control and sample bound fractions. As selection criteria only proteins solely present in the sample lane and completely absent in the control lane were considered putative interactors. Analyses were carried out in replicates and only proteins present in both replicates and identified with at least 2 peptides in MASCOT search were considered

microorganisms and plays a primordial role in the activation and regulation of adaptive immunity. In humans, components of the innate immune system include members of the complement cascade and soluble pattern recognition molecules (PRRs), functional ancestors of antibodies (25). In particular, the complement system plays a key role in the elimination of microorganisms after entrance in the human host.

The integrated strategy based on functional proteomic approach described in this paper represents a new starting point to elucidate the molecular mechanisms elicited by bacterial LPS and involved in the different steps of innate immunity response. Identification of the specific LPS-binding proteins in human serum was rarely addressed since most investigations focused on the LPS-mediated effects on cellular pathways and transcriptional responses. Only recently, a single example was reported using a solid-phase chemically immobilised LPS (26).

The strategy based on affinity capture procedure using biotin LPS as a bait was optimized and human serum from a healthy donor was used as model. A number of proteins were specifically retained by the LPS bait and subsequently identified by mass spectrometric methodologies. The identity of these proteins represented *per se* a validation of the developed affinity-capture procedure. These proteins, in fact, essentially gathered within two functional groups, regulation of the complement system (Factor H, C4b, C4BP and alpha 2 macroglobulin) and inhibition of LPS induced inflammation (HRG and Apolipoproteins). Both groups are functionally correlated with the biological role of LPS.

The complement system is an essential component of the innate immune system that participates in elimination of pathogens. Three different pathways synergistically contribute to the protection mechanism elicited by the complement proteins, the classical pathway, the alternative pathway and the lectin pathway. Despite the specific pathway they belong to, complement proteins provide host defence towards bacterial infections by binding to cell surface components of exogenous microorganisms. In contrast, pathogenic microorganisms have evolved several strategies to escape these defence mechanisms. The most common procedure consists in the recruitment of complement inhibitory proteins to the bacterial surface thus impairing the immune system.

In this respect, the identification of well characterized protein inhibitors of the complement system among the LPS interactors has a clear biological significance. C4BP, Factor H and alpha 2 macroglobulin are all complement inhibitors able to downregulate activation of the classical, alternative and lectin pathways respectively. *Streptococcus pyrogenes*, the etiologic agent of important human infections, was shown to bind Factor H, C4BP and other complement proteins as a crucial step in the pathogenesis of these infections (27). *Bordetella pertussis* was able to escape the classical pathway of complement by binding the classical pathway inhibitor C4BP. In addition, very recently it was shown that this pathogen can also evade the alternative complement pathway by recruitment of Factor H on the cell surface (28).

On the other hand, alpha 2 macroglobulin exerts a specific inhibitory activity towards the lectin pathway by binding the mannose-binding lectin (MBL) and the MBL associated serine proteases MASP-1 and MASP-2 (29). Adding of alpha 2

macroglobulin to human serum totally reversed killing of *Neisseria gonorrhoeae*, preventing MBL mediated activation of the complement system (30).

However, the most effective complement inhibitor is Factor H, a soluble protein regulator essential for controlling the alternative immune pathway (31). Recognition and binding Factor H constitutes the infection mechanism adopted by several pathogens as a common immune evasion strategy. In addition several recent reports pointed out that besides its role as an alternative pathway downregulator, Factor H has an additional complement regulatory role in inhibiting activation of the classical pathway (32; 33).

A totally different biological role can be ascribed to the second group of proteins identified by the LPS affinity capture strategy. HRG and Apolipoproteins, in fact, are known to exert protective effects against LPS induced systemic inflammation preventing inhibition of the complement immune system. HRG binds strongly to several complement protein inhibitors including Factor H and C4BP thus assisting in maintenance of normal immune function and enhancing complement activation (34). *S. pyrogenes* was shown to grow more efficiently in HRG-deficient plasma while the presence of overexpressed HRG greatly increased clots formation, bacterial entrapment and killing (35).

Apolipoproteins AI, A-IV and D are all components of the high density lipoproteins (HDL) that have long been reported to bind bacterial LPS neutralizing its toxicity and preventing initiation of innate immunity. More recently evidences for an Apo AI-LPS specific interaction have been obtained (36). Moreover, the adenovirus mediated overexpression of this protein led to protection of mice against LPS-mediated systemic inflammation (37).

Conclusions

In conclusion, the novel LPS-mediated affinity capture strategy developed in this paper proved to be very effective in specifically binding proteins involved either in inhibiting the activation of the complement system or in preventing bacterial infection by sequestering the LPS moiety. Both processes are of relevant biological significance and the results obtained makes sense with the bait used. Moreover, although a proper sensitivity test was not performed, the LPS affinity procedure led to the identification of low abundant serum proteins even without any depletion step. The sensitivity of the LC-MS/MS analyses allowed us to identify specific proteins even when the protein band showed only a very faint staining. A further advantage of this procedure concerns the possibility of determining both the proteins directly bound to LPS bait and other components involved in the functional complex but not linking the LPS moiety. It will be now possible to foresee applications of this strategy to investigate biological mechanisms exerted by pathogenic bacteria at the molecular level and to possibly define new mechanisms of infection.

Notes:

Department of Organic Chemistry and Biochemistry, University of Naples Federico II, Via Cinthia 6, 80126 Naples, Italy.

Corresponding Author: Angela Amoresano, Department of Organic Chemistry and Biochemistry, University of Naples Federico II, Naples,

References

- 1 BA. Beutler, *Blood*, 2009, **113**, 1399-407
- 2 R. Medzhitov, C. Janeway, Jr, *N. Engl. J. Med.*, 2000, **343**, 338–344
- 10 3 A. Iwasaki, R. Medzhitov, *Nat. Immunol.*, 2004, **5**, 987–995
- 4 S. Akira, S. Uematsu, O. Takeuchi, *Cell*, 2006, **124**, 783-801
- 5 SI. Miller, RK. Ernst, MW. Bader, *Nat Rev Microbiol.*, 2005, **3**, 36-46
- 6 GL. Su, RL. Simmons, SC. Wang, *Crit Rev Immunol.*, 1995, **15**, 201-14
- 15 7 MM. Wurfel, ST. Kunitake, H. Lichenstein, JP. Kane, SD. Wright, *J Exp Med.*, 1994, **180**, 1025-35
- 8 ZQ. Wang, WM. Xing, HH. Fan, KS. Wang, HK. Zhang, QW. Wang, J. Qi, HM. Yang, J. Yang, YN. Ren, SJ. Cui, X. Zhang, F. Liu, DH. Lin, WH. Wang, MK. Hoffmann, ZG. Han, *J Immunol.*, 2009, **183**, 6646-56
- 20 9 T. Nakamura, F. Tokunaga, T. Morita, S. Iwanaga, *J. Biochem.*, 1988, **103**, 370-374
- 10 NS. Tan, B. Ho, JL. Ding, *FASEB. J.*, 2000, **14**, 859-870
- 11 Z. Xu, CX. Huang, Y. Li, PZ. Wang, GL. Ren, CS. Chen, FJ. Shang, Y. Zhang, QQ. Liu, ZS. Jia, QH. Nie, YT. Sun, XF. Bai, *J Infect.*, 2007, **55**, e1-9
- 25 12 L. Guillot, S. Medjane, K. Le-Barillec, V. Balloy, C. Danel, M. Chignard, M. Si-Tahar, *Biol Chem.*, 2004, **279**, 2712–2718
- 13 A. Poltorak, P. Ricciardi-Castagnoli, S. Citterio, B. Beutler, *Proc Natl Acad Sci U S A.*, 2000, **97**, 2163-7
- 30 14 U. Ohto, K. Fukase, K. Miyake, Y. Satow, *Science*, 2007, **316**, 1632-4
- 15 J. Andrä, T. Gutschmann, M. Müller, AB. Schromm, *Adv Exp Med Biol.*, 2010, **667**, 39-51
- 35 16 JC. Marshall, *Clin Infect Dis.*, 2005, **41**, S470-80
- 17 TH. Mogensen, *Clin Microbiol Rev.*, 2009, **22**, 240-73
- 18 WJ. Liaw, C. Tzao, JY. Wu, SJ. Chen, JH. Wang, CC. Wu, *Shock*, 2003, **19**, 281-8
- 19 RK. Kollipara, NB. Perumal, *Immunome Res.*, 2010, **6**, 5
- 40 20 M. Matsumoto, Y. Horiuchi, A. Yamamoto, M. Ochiai, M. Niwa, T. Takagi, H. Omi, T. Kobayashi, MM. Suzuki, *J Microbiol Methods*, 2010, **82**, 54-8
- 21 WJ. Qian, JM. Jacobs, DG. Camp, ME. Monroe, RJ. Moore, MA. Gritsenko, SE. Calvano, SF. Lowry, W. Xiao, LL. Moldawer, RW. Davis, RG. Tompkins, RD. Smith, *Proteomics*, 2005, **5**, 572-84
- 45 22 S. Li, L. Wang, M. Berman, YY. Kong, ME. Dorf, *Immunity*, 2011
- 23 O. Westphal, K. Jann, *Methods carbohydr. Chem.*, 1965, **5**, 83-91
- 24 M. Monti, M. Cozzolino, F. Cozzolino, R. Tedesco, P. Pucci, *Ital J Biochem.*, 2007, **56**, 310-4
- 50 25 L. Deban, S. Jaillon, C. Garlanda, B. Bottazzi, A. Mantovani, *Cell Tissue Res.*, 2010, **343**, 237-49
- 26 YG. Kim, YH. Yang, BG. Kim, *J Chromatogr B Analyt Technol Biomed Life Sci.*, 2010, **878**, 3323-6
- 27 MA. Oliver, JM. Rojo, S. Rodríguez de Córdoba, S. Alberti, *Vaccine*, 2008, **26** Suppl 8: I75-8
- 55 28 H. Amdahl, H. Jarva, M. Haanperä, J. Mertsola, Q. He, TS. Jokiranta, S. Meri, *Mol Immunol.*, 2011, **48**, 697-705
- 29 G. Ambrus, P. Gál, M. Kojima, K. Szilágyi, J. Balczer, J. Antal, L. Gráf, A. Laich, BE. Moffatt, W. Schwaeble, RB. Sim, P. Závodszyk, *J Immunol.*, 2003, **170**, 1374-82
- 60 30 S. Gulati, K. Sastry, JC. Jensenius, PA. Rice, S. Ram, *J Immunol.*, 2002, **168**, 4078-86
- 31 VP. Ferreira, MK. Pangburn, C. Cortés, *Mol Immunol.*, 2010, **47**, 2187-97
- 65 32 LA. Tan, B. Yu, FC. Sim, U. Kishore, RB. Sim, *Protein Cell.*, 2010, **1**, 1033-49
- 33 LA. Tan, AC. Yang, U. Kishore, RB. Sim, *Protein Cell.*, 2011, **2**, 320-332
- 34 GA. Manderson, M. Martin, P. Onnerfjord, T. Saxne, A. Schmidtchen, TE. Mollnes, D. Heinegård, AM. Blom, *Mol Immunol.*, 2009, **46**, 3388-98
- 35 O. Shannon, V. Rydengård, A. Schmidtchen, M. Mörgelin, P. Alm, OE. Sørensen, L. Björck, *Blood*, 2010, **116**, 2365-72
- 36 MF. Henning, V. Herlax, L. Bakás, *Innate Immun.*, 2011, **17**, 327-37
- 75 37 Y. Li, JB. Dong, MP. Wu, *Eur J Pharmacol.*, 2008, **590**, 417-22.