Tesi Doctoral

Glycoprobes for capture and identification of lectins from biological sources

Sira Defaus Fornaguera

Universitat de Barcelona

Universitat Pompeu Fabra Barcelona

IMIM Parc de Salut MAR Barcelona
Institut Hospital del Mar d'Investigacions Mèdiques

Parc Recerca Biomèdica Barcelona

Glycoprobes for capture and identification of lectins from biological sources

Sira Defaus Fornaguera

Memòria presentada per

Sira Defaus Fornaguera

per optar al grau de Doctor per la Universitat de Barcelona

Revisada per:

Dr. David Andreu Martínez
Universitat Pompeu Fabra
Director de Tesi

Dr. Ricardo Gutiérrez Gallego
Universitat Pompeu Fabra
Director de Tesi

Dr. Jordi Robles Brau
Universitat de Barcelona
Tutor de Tesi

Programa de Doctorat de Química Orgànica

Departament de Química Orgànica

Facultat de Química

Universitat de Barcelona

2014

Al meu pare, el meu millor amic i principal puntal.

I a la meva estimada Lola,
no passa dia que no t'enyori, iaia.

Fer una tesi és un repte personal..., però que no es pot fer sol. Per aquest motiu, des d'aquestes línies vull expressar el meu més sincer agraïment a totes aquelles persones que m'han acompanyant durant aquests anys tan intensos, i que de ben segur han contribuït en la culminació de la present tesi doctoral. Sense elles, res d'això hauria estat possible.

Als directors d'aquesta tesi, David i Ricardo, per la confiança dipositada en mi i pel constant suport. Gràcies per contribuir amb el vostre esforç, ensenyament i dedicació a fer d'aquesta etapa, una etapa de creixement acadèmic, professional, però sobretot, personal i emocional. Agrair també a en Jordi, el meu tutor, els ànims i consells rebuts sobretot en els meus inicis en la investigació.

Als companys de laboratori, tant de la UPF com l'IMIM. Gràcies pels moments viscuts i per fer-me el dia a dia més agradable. Ha estat un plaer poder compartir aquesta tesi envoltada de gent fantàstica.

A l'equip de Proteòmica per la vostra amabilitat i ajuda en totes les qüestions requerides; i als laboratoris del Dr. M. Avilés i Prof. M. Przybylski per la càlida acollida en ambdues estades.

Als meus amics, per donar color a la meva vida. Sonia i Carmen, amigues com vosaltres, són difícils de trobar, fàcils d'estimar, i impossibles d'oblidar. Gràcies especialment a la Sonia, per compartir tota una vida amb mi i pel vincle tan especial que ens uneix.

A la meva família, pel vostre amor incondicional i per estar al meu costat en tot moment. Gràcies per tot el que m'heu donat i ensenyat. Gràcies a vosaltres he arribat fins aquí. A la Meri, la meva germana, i als "peques", Arnau i Ada, gràcies per fer-me feliç.

I al Miquel, per ser el millor company de viatge. Gràcies per fer-me sentir tan afortunada d'haver trobat algú amb qui construir la meva vida. Diu el proverbi que "les petjades de les persones que van caminar juntes mai s'esborren". Doncs, dóna'm la mà i fem camí plegats.

A totes i a tots, gràcies de tot cor.

CONTENTS

RESUM

Introducció i objectius

En biologia trobem quatre classes principals de macromolècules constituents essencials dels éssers vius: els carbohidrats, els àcids nucleics, les proteïnes i els lípids. Mentre que les proteïnes i els àcids nucleics són polímers bàsicament lineals amb un únic tipus d'enllaç entre els monòmers constituents, els carbohidrats es caracteritzen per un ampli ventall d'arranjaments estructurals. Aquesta diversitat i resultant complexitat dels carbohidrats sorgeix principalment de la varietat estructural a nivell monomèric, on les gairebé il·limitades combinacions multidireccionals dels diferents monosacàrids, els diferents tipus d'enllaç, l'anomericitat i les possibles ramificacions generen un gran nombre d'estructures de glicans (d'oligo- a polisacàrids) possibles. Aquestes es poden modificar encara més per sulfatació, acetilació, metilació, fosforilació, *etc.*, i/o unir covalentment a altres molècules com ara pèptids o lípids per formar els corresponents glicoconjugats amb particulars funcions i localitzacions en els organismes vius (Figura i).

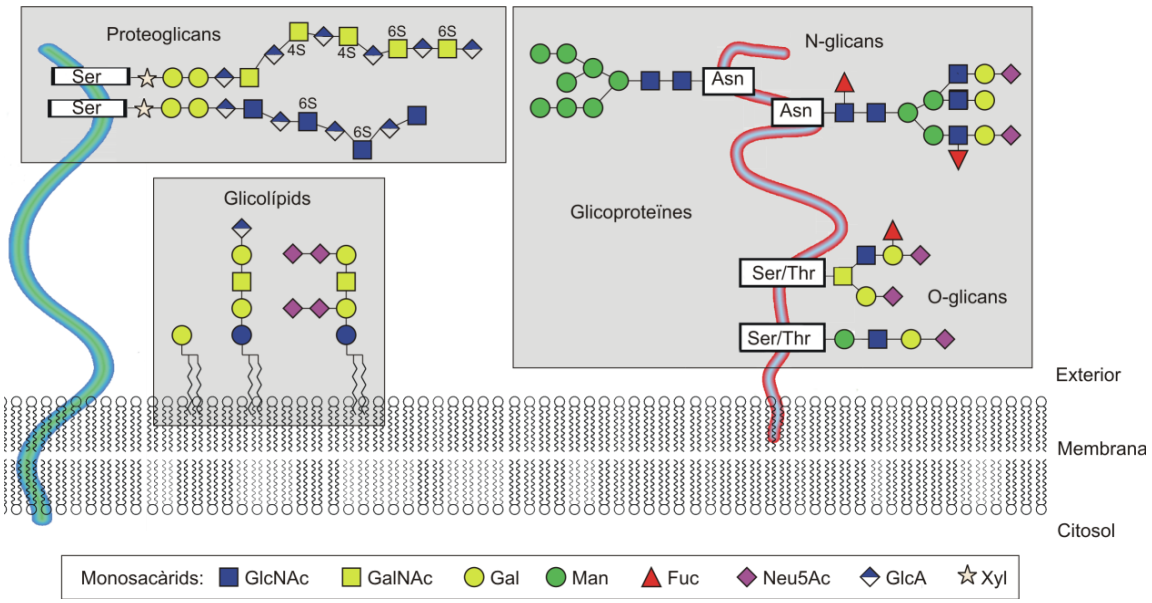


Figura i. Tipus de glicoconjugats comuns en animals.

La glicosilació, el procés enzimàtic que uneix sacàrids per produir glicans que s'adhereixen a proteïnes, lípids o altres molècules biològiques, és una modificació co-traduccional i post-traduccional present a la pràctica totalitat de components cel·lulars, on trobem glicoproteïnes, glicolípids i altres glicoderivats. Pel que fa específicament a les proteïnes, s'estima que més d'un 80% estan glicosilades i que aquests glicans són fonamentals en processos biològics com la senyalització cel·lular, el cicle infeccións de certs patògens, les respostes inflamatòria i immune, la fertilització, *etc*. Donat que es tracta d'una modificació que depèn de la disponibilitat i activitat d'un conjunt de glicosiltransferases, de la coincidència d'un substrat donador i acceptor en la dimensió espai-temps i també d'altres factors fisiològics, pot existir un elevat grau d'heterogeneïtat estructural en un determinat lloc específic de glicosilació. Això explica que el glicoma (composició total de carbohidrats en una cèl·lula o organisme) sigui molt més complex que el genoma, transcriptoma o proteoma, i que tingui també un caràcter més dinàmic, que varia considerablement no només amb el tipus de cèl·lula o teixit, sinó també amb l'etapa de desenvolupament, l'estat metabòlic, la malaltia, l'envelliment, els factors ambientals o l'evolució.

És precisament en aquesta heterogeneïtat on rau la importància de la glicosilació i dels glicans com a elements primaris de comunicació entre (glico)proteïnes, entre proteïnes i cèl·lules, o entre cèl·lules i agents exògens com ara bacteris o virus. Encara que la rellevància d'aquestes interaccions mediades per carbohidrats ha estat poc reconeguda —fins i tot hom podria dir ignorada— fins fa relativament poc temps, actualment ningú posa en dubte la seva extraordinària importància i la necessitat de tècniques adequades per al seu estudi. En efecte, la identificació del nombre, estructura i funció dels glicans en un context biològic particular, el que anomenem glicòmica, ha experimentat darrerament un seguit d'avenços substancials que faciliten la comprensió del paper dels glicans en gairebé tots els processos biològics i malalties humanes.

L'anàlisi de la glicosilació es pot abordar a diferents nivells: (i) caracterització de glicans en glicoproteïnes intactes, (ii) caracterització de glicopèptids i (iii) l'anàlisi estructural de glicans alliberats química o enzimàticament (Figura ii). En les últimes dècades el continu perfeccionament de les eines analítiques ha facilitat en gran mesura aquests estudis i actualment es disposa de nombroses tecnologies utilitzades rutinàriament en l'anàlisi de glicans, incloent l'electroforesi capil·lar, la cromatografia líquida, l'espectrometria de masses i els estudis basats en microarrays. No obstant, cal

tenir clar que no hi ha un mètode analític únic capaç de proporcionar tota la informació necessària per a la identificació i la quantificació ràpida i fiable d'una estructura en particular, i molt menys per establir també la seva funcionalitat. Aquesta tasca requereix un enfocament multidimensional que faci ús d'una combinació de diverses tècniques físiques, químiques i bioquímiques ortogonals. Així doncs, abans d'emprendre l'anàlisi de la glicosilació és important determinar: (i) quin nivell d'informació es requereix, (ii) si es necessiten dades quantitatives o qualitatives i (iii) la quantitat de material (pur) disponible. Aquestes condicions determinaran les tecnologies que millor s'adaptin al problema que ens interessa abordar i ens permetin desxifrar una part del trencaclosques.

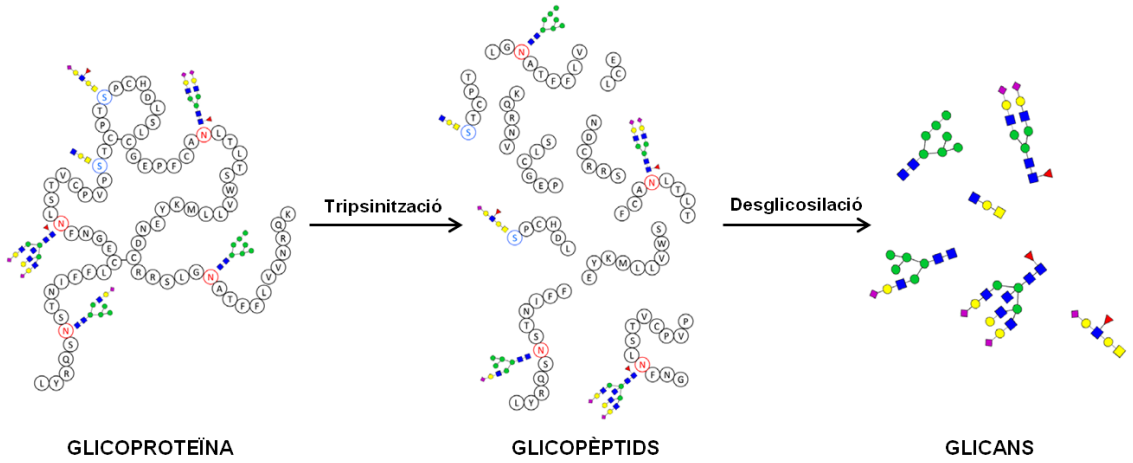


Figura ii. Aproximacions a l'estudi de la glicosilació: per mitjà de la glicoproteïna intacta (esquerra), dels glicopèptids (centre) o dels glicans alliberats (dreta).

Com a resultat d'aquest progrés analític i tecnològic, en els últims anys s'ha avançat substancialment en el coneixement bàsic de la funció de determinats epítops o cadenes glicosídiques concretes, per exemple grups sanguinis importants com ABO o Sda, o els epítops (sialil)Lewis^{a,b,x,y} o els determinants HNK-1. No obstant, es desconeixen les funcions de moltes altres estructures glicosídiques i, més concretament, de la combinació d'una determinada glicosilació en una determinada proteïna o lípid. D'altra banda, també existeix un cert desconeixement sobre les molècules que reconeixen els carbohidrats, les lectines "lectores del codi glicòmic". Aquestes proteïnes es caracteritzen per reconèixer i unir-se de forma reversible i específica a certs monosacàrids o epítops oligosacàrids donant lloc a interaccions similars a les reaccions antigen-anticòs o enzim-substrat. No obstant, les interaccions carbohidrat-lectina són relativament febles en comparació amb les anteriors, amb constants de dissociació sovint de l'ordre de mM per a monosacàrids.

El paper de les lectines en processos com l'aglutinació i la definició de grups sanguinis, la inflamació (selectines), la mielinització del teixit nerviós, la progressió tumoral, *etc*., demostra la ubiqüitat i diversitat d'activitats en què es veuen implicades aquestes proteïnes. Per això, disposar d'una eina ràpida i fiable per al seu aïllament i identificació, previ a l'estudi de les seves interaccions amb polisacàrids, constitueix un objectiu de màxim interès en l'actual investigació biomèdica.

De fet, la investigació actual en lectines es pot dividir en dues línies principals. La primera es basa en l'estudi de noves lectines, la caracterització de les seves estructures, les seves funcions biològiques i l'especificitat d'unió a carbohidrats. La segona, es centra en l'aplicació de les lectines amb especificitat coneguda com a eines per a l'aïllament, identificació, caracterització i estudi de glicans lliures, així com dels seus conjugats amb lípids i proteïnes. En particular, les lectines provinents de plantes simples constitueixen la família més gran i ben estudiada, i són les més utilitzades per a aquesta segona aproximació, donat el gran nombre de sacàrids amb què poden interactuar de forma específica (Figura iii).

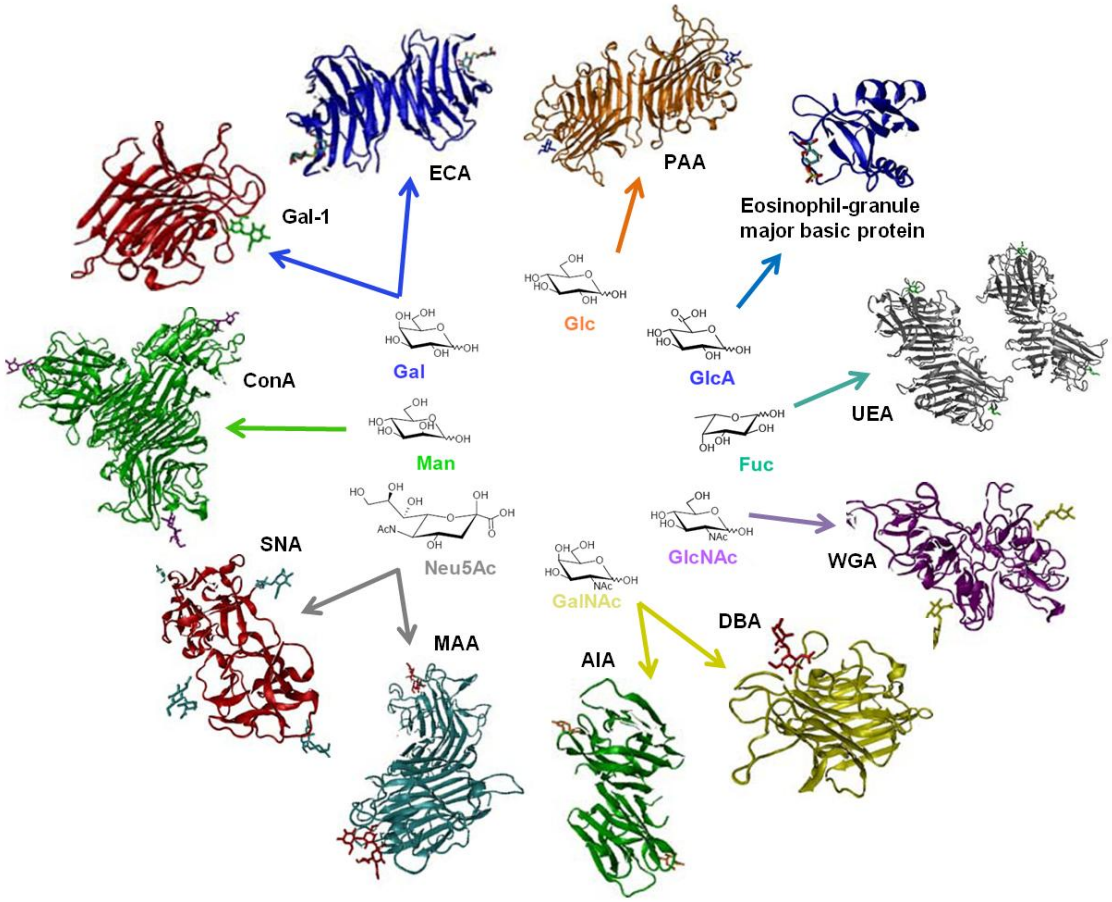


Figura iii. Els vuit monosacàrids principals que es troben normalment com a terminacions dels glicans presents en glicoconjugats (centre) i les corresponents lectines que els reconeixen.

Encara que ja es disposa d'informació estructural detallada sobre alguns complexos lectina-carbohidrat, elucidats habitualment per cristal·lografia de raigs X, calorimetria isotèrmica, espectroscòpia de fluorescència o RMN, aquestes tècniques plantegen certes limitacions: (i) són estàtiques, és a dir, requereixen que els dos components (proteïna-carbohidrat) que interactuen s'hagin aïllat i purificat prèviament en (ii) quantitats i pureses considerables (en alguns casos a escala mg), no sempre assumibles; (iii) els temps d'anàlisi són relativament llargs i en alguns casos (difracció raigs X, RMN) el tractament de les dades és molt laboriós.

Des dels anys 90 fins a l'actualitat la ressonància de plasmó superficial (SPR; *Surface plasmon resonance*) s'ha emprat per a l'estudi d'interaccions entre biomolècules, ja que presenta una sèrie d'avantatges respecte a les tècniques anteriorment esmentades. Mitjançant l'SPR s'obtenen, en temps real, tant dades qualitatives (si es dóna o no interacció) com paràmetres quantitatius (cinètics i termodinàmics) de la interacció. A més, la tècnica permet mesurar les interaccions directament, de manera que no és necessari el marcatge de les molècules per a la seva detecció, i tampoc es requereixen grans quantitats de mostra. Gràcies a aquests avantatges, la tècnica d'SPR és molt utilitzada en l'estudi de les interaccions entre biomolècules de qualsevol tipus (ADN-proteïna, ADN-ADN, proteïna-carbohidrat).

En el desenvolupament d'aquesta tesi, per a l'estudi d'interaccions carbohidrat-lectina hem fet servir la tecnologia d'SPR desenvolupada per Biacore AB (Uppsala, Suècia), on s'utilitza un xip que defineix dos medis amb diferent índex de refracció separats per una petita làmina d'or que actua com a film conductor. En un costat de la làmina hi ha un prisma de vidre que actua com a medi on hi ha la llum incident, i en l'altre costat hi ha la mostra que es vol analitzar. Per tal d'avaluar si es produeix alguna mena d'interacció, en el costat de la mostra s'hi immobilitza el lligand i posteriorment s'injecta l'analit mitjançant un sistema de canals de flux continu. Si l'analit s'uneix al lligand immobilitzat, varia la concentració del medi, i per tant també el seu índex de refracció. Aquest canvi en l'índex de refracció, provoca al seu torn una variació en l'angle del mínim d'intensitat del feix reflectit, que s'expressa en unitats de ressonància (RU). En la tecnologia SPR, es mesura el valor d'aquest angle i es representa la seva variació en funció del temps. Aquesta representació s'anomena sensograma (Figura iv). En el procediment habitual, s'injecta mitjançant un flux continu la mostra amb l'analit sobre la superfície que conté el lligand immobilitzat. Si es produeix interacció, s'obté una

corba d'associació a partir de la qual es pot obtenir la k_a. Posteriorment, s'atura la injecció d'analit, però es manté un flux continu utilitzant el mateix dissolvent. A partir d'aquest moment, s'observa com la resposta va disminuint degut a la dissociació de la interacció, fet que permet obtenir la k_d. A més a més, l'increment de resposta que s'ha produït en la injecció de l'analit dóna una idea de la quantitat d'analit que s'hi ha unit. A partir de tots aquests valors, es pot obtenir la constat d'afinitat (K_A), així com la ΔH^o i la ΔG^o. Finalment, s'injecta una solució de regeneració la qual trenca totalment el complex no covalent format pel lligand i l'analit, i s'obté novament la superfície inicial.

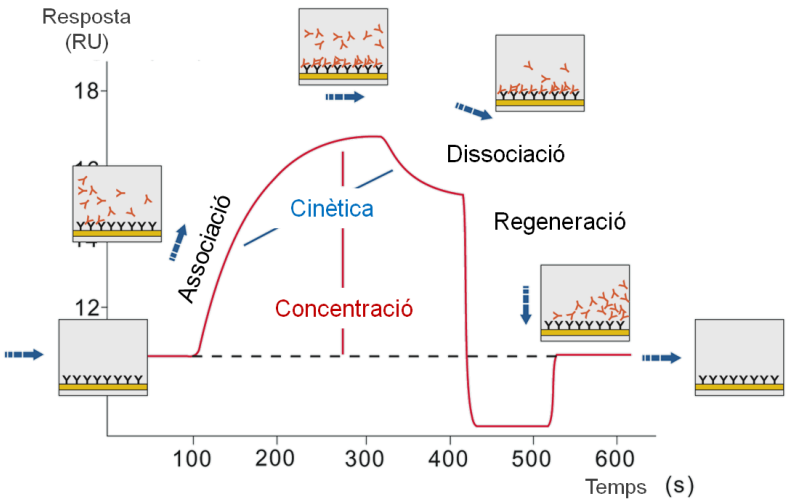


Figura iv. Perfil representatiu d'un sensograma.

Apart de l'estudi de la interacció en termes cinètics i/o termodinàmics, una anàlisi a fons del mecanisme d'interacció entre carbohidrat-lectina ha d'incloure idealment la determinació del domini de reconeixement de la lectina al carbohidrat. Tal com s'ha comentat anteriorment, tot i que la cristal·lografia de raigs X i l'RMN s'utilitzen principalment per a estudis estructurals dels complexos carbohidrat-lectina, les quantitats de mostra requerides en aquestes tècniques no sempre són fàcils d'obtenir. De fet, pràcticament tots els estudis actuals per a la caracterització de proteïnes es basen en l'ús combinat de la proteòlisi limitada amb mètodes d'espectrometria de masses (EM). Així doncs, en aquesta tesi la determinació del domini d'unió s'ha dut a terme utilitzant una columna de cromatografia d'afinitat, on els carbohidrats s'immobilitzen a una matriu de sefarosa mitjançant divinilsulfona. Aquesta metodologia, basada en la proteòlisi limitada i l'EM, anomenada CREDEX-MS (*Carbohydrate-REcognition-Domain-EXcision Mass Spectrometry*), pot esdevenir una alternativa viable per tal d'elucidar les interaccions carbohidrat-lectina.

Un dels processos biològics per excel·lència on el reconeixement de carbohidrats juga un paper crucial és la fertilització. Les dades recollides fins ara indiquen que hi ha una jerarquia d'esdeveniments d'unió mediats per carbohidrats al llarg de la ruta que porta l'espermatozoide fins a la fecundació de l'òvul. D'altra banda, també s'ha demostrat que certs carbohidrats presents en l'esperma estan implicats en la infertilitat. Així doncs, ambdós factors plantegen importants reptes per a la futura investigació en el camp de la glicobiologia en l'àmbit de la biologia reproductiva.

En els mamífers, la fecundació es produeix en el tracte reproductiu femení. Com es mostra a la Figura v, amb l'ejaculació els espermatozoides es dipositen a l'interior de la vagina i viatgen a través del coll uterí i l'úter cap a les trompes uterines. En arribar al oviducte, els espermatozoides s'uneixen a l'epiteli on són retinguts en forma de dipòsit d'espermatozoides. Aquest segrest manté la fertilitat de l'esperma, regula el transport d'esperma en l'oviducte, redueix el risc de polispèrmia i assegura que els espermatozoides estiguin presents en el moment adequat per fertilitzar l'òvul. Després del procés de capacitació (sèrie d'esdeveniments de maduració necessaris per a l'èxit de la fertilització) i coincidint amb senyals endocrins associats amb l'ovulació, els espermatozoides s'alliberen del dipòsit i segueixen camí per l'oviducte. En trobar-se amb l'oòcit, els espermatozoides penetren la corona radiada i s'uneixen a les glicoproteïnes presents en la matriu extracel·lular que encapsula l'òvul, coneguda com la zona pel·lúcida (ZP). Aquesta unió primària inicia l'exocitosi de l'acrosoma de l'espermatozoide, que resulta en l'alliberament d'enzims hidrolítics i, en conseqüència, permet la penetració dels espermatozoides a través de la ZP i en l'espai perivitel·lí. Finalment, es dóna la fusió o unió secundària de l'espermatozoide que té l'acrosoma reaccionat amb la membrana plasmàtica de l'òvul, completant la fecundació.

La formació del dipòsit d'esperma a l'oviducte i la unió espermatozoide-oòcit (en la ZP) són processos mediats principalment per interaccions proteïna-carbohidrat, que impliquen carbohidrats tant en l'epiteli de l'oviducte com en la ZP, i lectines en la superfície de l'espermatozoide. A causa de les evidents divergències evolutives, les unitats específiques implicades en aquests esdeveniments d'unió han de ser identificades de forma independent en cada espècie d'interès. Concretament, pel que fa a l'espècie bovina (model estudiat en aquesta tesi), la formació del dipòsit d'espermatozoides sembla estar regulat per la interacció entre residus de fucosa presents en l'epiteli de l'oviducte i lectines que es troben en la superfície de l'esperma, on la proteïna PDC-109

ha estat descrita com una ferma candidata en la interacció. D'altra banda, s'ha observat que l'àcid siàlic està involucrat en el procés de reconeixement inicial entre gàmetes, tot i que les lectines complementàries encara no han estat identificades.

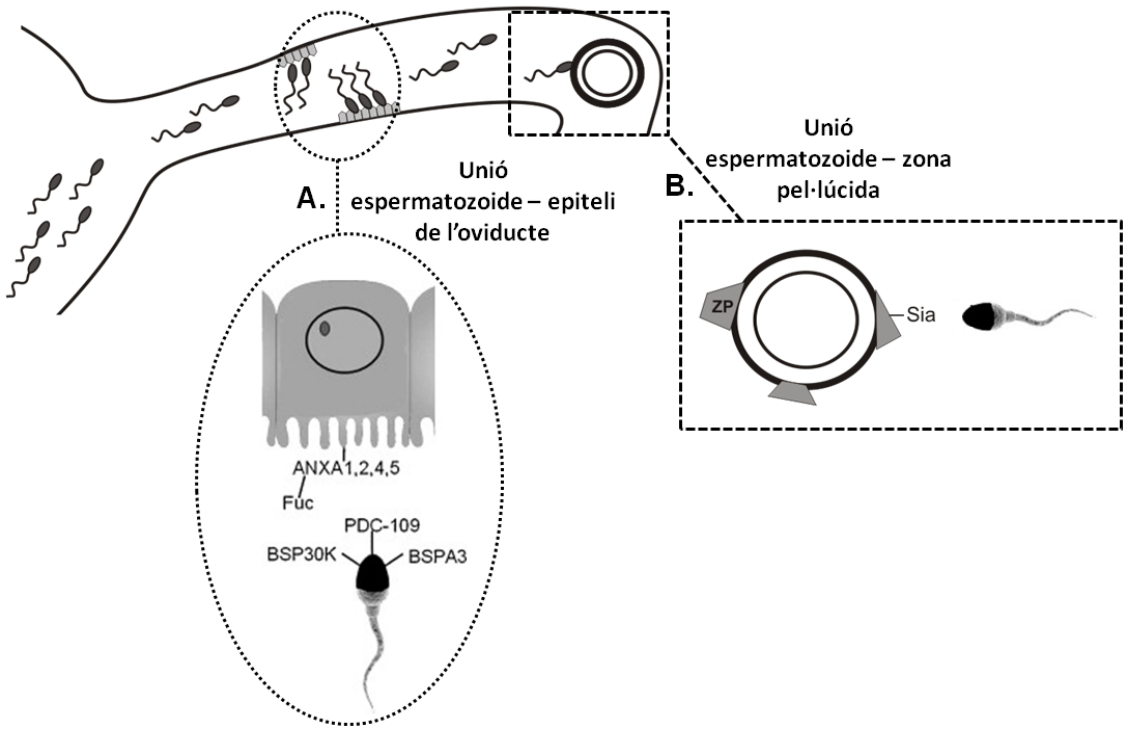


Figura v. Esdeveniments de la fertilització, on les interaccions proteïna-carbohidrat presenten un paper important. (**A**) Formació del dipòsit d'esperma en l'epiteli de l'oviducte. (**B**) Reconeixement entre gàmetes. Els carbohidrats han estat parcialment identificats, però les lectines complementàries presents en l'esperma encara no han estat consensuades.

Diverses molècules possiblement implicades en la interacció espermatozoide-oviducte o espermatozoide-oòcit han estat identificades en els espermatozoides de diferents espècies de mamífers, suggerint que possiblement són diversos els carbohidrats i lectines que participen en aquests processos específics de reconeixement. No obstant, encara no existeix un consens sobre els mecanismes d'unió i les molècules implicades. Per tant, tot i que actualment hi ha una gran quantitat d'informació disponible sobre els esdeveniments que es produeixen en la fertilització, les peces del trencaclosques encara resten per determinar. Cal encara un estudi en profunditat de les molècules presents en la superfície de l'esperma, així com en d'altres tipus de cèl·lules dels tractes genitals masculins i femenins, per aconseguir una major comprensió dels mecanismes moleculars que controlen el procés de reproducció dels mamífers. Aquest fet no només serà d'un immens valor en el camp de la reproducció assistida, sinó que també serà útil

per a la immuno-anticoncepció, on el/s receptor/s espermàtic/s de la ZP poden ser aparentment una diana clara per a estratègies de vacunació.

El present treball se situa en l'àmbit de la glicobiologia i, en concret, en el paper dels glicans en processos biomoleculars. En aquest context, l'objectiu principal d'aquesta tesi és el disseny i elaboració de neoglicopèptids d'estructura ben definida que actuïn com a glicosondes, caracteritzables mitjançant les lectines complementàries, i es puguin utilitzar per a la captura i identificació de proteïnes d'unió a glicans (és a dir, lectines) rellevants en matrius biològiques. Eventualment, donat que aquestes proteïnes sovint estan associades a certs estats patològics, la seva identificació podria proporcionar valuoses eines de diagnòstic. Per aconseguir aquest objectiu fonamental, així com per a l'estudi de les interaccions proteïna-carbohidrat, es pretén utilitzar la tècnica d'SPR explicada anteriorment. Simultàniament, fruit d'una col·laboració amb el grup del Dr. Przybylski de la Universitat de Konstanz, s'implementarà la nova metodologia CREDEX-MS per identificar mitjançant EM els pèptids que s'uneixen específicament a l'epítop glicosídic. Amb aquesta tècnica es realitzaran estudis més exhaustius del domini de reconeixement i s'estendrà el camp de l'aplicabilitat a mostres contingudes en matrius complexes no compatibles amb l'SPR. Per aconseguir aquest objectiu principal es defineixen els següents **objectius** parcials:

- Síntesi de neoglicoconjugats que exposin epítops oligosacàrids més rellevants (disacàrids amb diferents tipus d'enllaç i oligosacàrids).
- Caracterització de les interaccions biomoleculars entre aquestes glicosondes i diverses lectines utilitzant la tècnica d'SPR.
- Intersecció i complementació dels resultats d'SPR amb els de la tècnica CREDEX-MS.
- Aplicació de les anteriors tecnologies a sistemes biològics rellevants, per tal d'identificar proteïnes complementàries a carbohidrats. Ens centrarem en el tema de la reproducció molecular, en concret en la interacció entre espermatozoides i oòcits, on recentment hom ha establert que els glicans en les proteïnes de la ZP del oòcit són fonamentals.

Resultats i discussió

En la majoria d'estudis d'interacció proteïna-carbohidrat descrits en la literatura, la molècula immobilitzada sobre el suport sòlid sol ser la proteïna, perquè la immobilització del carbohidrat presenta alguns problemes importants. Atès que el carbohidrat no acostuma a tenir un grup amino acoblable a la superfície del sensor (molt sovint funcionalitzada amb grups carboxil), resulta imprescindible la seva modificació química per immobilitzar-lo sobre el suport. Aquest procés pot afectar tant el carbohidrat com la matriu, i pot per tant alterar les interaccions a estudiar. A més, en ser el carbohidrat una molècula de petita grandària, el senyal d'SPR que produeix la seva immobilització al xip sensor no és suficient per confirmar-ne inequívocament la incorporació.

Per solucionar aquest problema, al nostre laboratori s'ha desenvolupat un mètode fiable i senzill per immobilitzar el carbohidrat sobre la superfície sòlida d'un xip d'SPR. Aquesta metodologia consisteix en utilitzar un glicopèptid de fàcil unió a la superfície del sensor a través de la part peptídica. D'aquesta manera s'assegura una immobilització eficaç, fet de màxima importància degut a l'escassa quantitat disponible de la majoria de glicans rellevants; s'estalvien també etapes de derivatització sobre la superfície sensora, i a més s'aconsegueix un senyal d'SPR més pronunciat, a causa de l'augment de talla del lligand. Un altre avantatge addicional és el fet que el glicopèptid pot ser caracteritzat químicament de manera exhaustiva abans de la unió al xip, la qual cosa proporciona més garanties que els mètodes convencionals basats en la immobilització del carbohidrat.

Per al disseny del mòdul peptídic, s'ha desenvolupat la seqüència Aoa(m)-GFKKG, optimitzada tenint en compte una sèrie de factors estructurals. A l'extrem N-terminal es situa un grup aminoxiacetil (Aoa) que permet unir el carbohidrat, sense cap modificació, pel seu extrem reductor per mitjà d'un enllaç oxima amb el mòdul peptídic. S'utilitza el residu d'àcid aminooxiacètic metilat, Aoa(m), per tal de garantir la correcta exposició del carbohidrat en la superfície del xip, així com la integritat conformacional (estructura d'anell tancat) del monosacàrid pròxim a la superfície. D'altra banda, a l'extrem C-terminal del mòdul peptídic s'hi incorporen dos residus de Lys, que faciliten tant l'aproximació electrostàtica a la superfície aniònica del xip com la immobilització covalent sobre aquest, mitjançant un enllaç amida amb els grups carboxílics de la

superfície. Pel que fa als altres aminoàcids de la seqüència Aoa(m)-GFKKG, el residu hidrofòbic de Phe facilita la posterior separació i purificació del producte, i els dos de Gly augmenten de mida el pèptid –fent-lo preparativament més manejable–, intentant alhora evitar, amb el mínim volum d'aquest residu, possibles conflictes estèrics que dificultin les interaccions, tant amb la superfície com subseqüentment.

En resum, la nostra aproximació glicopeptídica es desenvolupa en tres etapes: (i) síntesi en fase sòlida del mòdul peptídic i posterior purificació per HPLC; (ii) unió quimioselectiva entre el carbohidrat i el pèptid, purificació del conjugat resultant i caracterització per HPLC i EM; i (iii) immobilització del glicopèptid sobre el xip d'SPR.

Amb aquest disseny optimitzat de glicosondes com a punt de partida, el primer objectiu d'aquesta tesi, descrit en el **capítol 1**, ha estat demostrar la utilitat i complementarietat de les metodologies SPR i CREDEX-MS en l'anàlisi aprofundida d'una interacció lectina-carbohidrat coneguda. Concretament, s'ha estudiat la interacció model entre la lectina ECA (*Erythrina cristagalli agglutinin*) i una sèrie de β-galactoses [Gal(β1-4)GlcNAc, Gal(β1-4)Glc, Gal(β1-3)GlcNAc, Gal(β1-6)GlcNAc)].

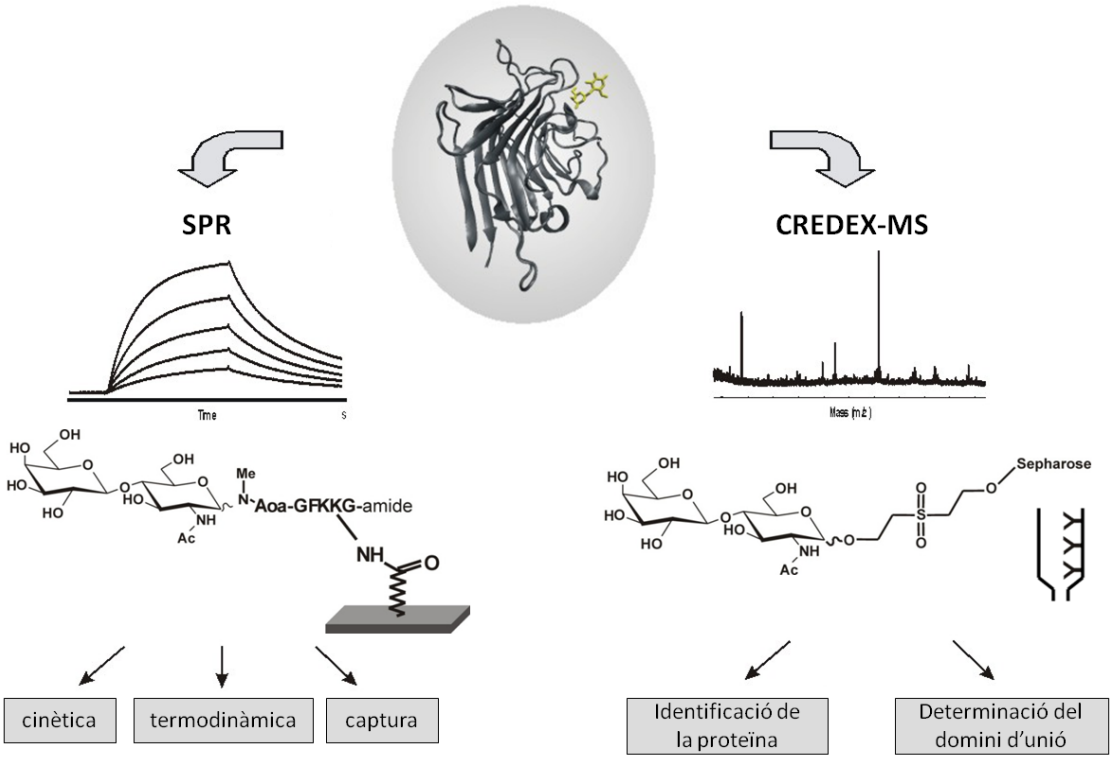


Figura vi. Esquema general de les tècniques utilitzades en aquesta tesi (SPR i CREDEX-MS) per a l'estudi d'interaccions carbohidrat-lectina. Es mostren les plataformes utilitzades en ambdues metodologies en l'exemple concret del complex ECA-lacNAc, així com la informació que es pot extreure de cadascuna d'elles.

Per mitjà de la tècnica d'SPR s'han pogut determinar les constants cinètiques de les fases d'associació i dissociació (k_a, k_d), així com les constants d'afinitat derivades (K_A) de cadascun dels disacàrids. A més dels paràmetres cinètics d'enllaç, els resultats obtinguts amb SPR han permès comprendre trets estructurals claus en els esdeveniments d'unió, en particular sobre els grups funcionals que intervenen en cada interacció. Concretament, s'ha observat que els hidroxils en C3 i C6, així com el grup *N*-acetil al C2, són crítics per a la interacció. Per tant, la unió a la lectina ECA està fortament influenciada pel tipus d'enllaç glicosídic, així com per la naturalesa del monosacàrid en l'extrem reductor; amb una clara preferència per l'epítop Gal(β1-4)GlcNAc, d'acord amb estudis anteriors. En segon lloc, els paràmetres termodinàmics de la interacció d'ECA amb aquest epítop preferencial Gal(β1-4)GlcNAc (lacNAc) es poden definir en temps real controlant la resposta en SPR a diverses temperatures i utilitzant l'equació de Van't Hoff o d'Eyring. Finalment, es demostra l'aplicabilitat de l'SPR també com a plataforma de captura de lectines amb afinitat pel carbohidrat immobilitzat i la seva posterior identificació per EM.

Complementàriament, emprant la metodologia CREDEX, prèviament validada per RMN, s'ha pogut determinar el domini de reconeixement del complex ECA-lacNAc. Per EM s'identifiquen diversos pèptids amb aminoàcids que presenten contacte directe amb el carbohidrat segons l'estructura de raigs X del complex i, per tant, poden ser assignats de forma inequívoca al lloc d'unió.

En resum, en el primer capítol hem demostrat que les metodologies d'SPR i CREDEX-MS són notablement complementàries, constituint per tant dues eines valuoses per desxifrar i estudiar amb detall les interaccions proteïna-carbohidrat.

Un cop constatada la validesa de la doble aproximació SPR i CREDEX-MS en una interacció model lectina-carbohidrat, el següent repte ha estat estendre la metodologia a una diana biològicament rellevant. En aquest context, el treball descrit en el **capítol 2** es centra en l'estudi de la principal proteïna del plasma seminal boví, PDC-109 (BSP-A1/-A2), en el seu paper de lectina participant en la fertilització.

En primer lloc, hem ampliat l'abast de la plataforma SPR amb una col·lecció de glicosondes ben definides que exposen els epítops glicosídics més rellevants dels mamífers (Taula i). Un cop sintetitzades, caracteritzades i immobilitzades als xips d'SPR, han estat validades en estudis d'interacció amb les lectines específiques corresponents als epítops exposats. En aquesta avaluació, les especificitats aparents i

constants d'afinitat determinades per SPR han estat consistents amb els resultats publicats prèviament, el que corrobora la fiabilitat de les glicosondes immobilitzades per als posteriors estudis amb la mostra biològica de PDC-109, purificada de l'esperma boví.

Taula i. Epítops glicosídics típics dels mamífers presents en les glicosondes sintetitzades i lectines complementàries utilitzades per a la seva validació.

	Carbohidrats	**Lectines**
Disacàrids amb β-galactosa terminal	Gal-β1,4-Glc Gal-β1,4-GlcNAc Gal-β1,3-GlcNAc Gal-β1,6-GlcNAc	*Erythrina cristagalli* (ECA)
Disacàrids amb mannosa	Man-α1,2-Man Man-α1,3-Man Man-α1,6-Man	*Concanavalin A* (Con A)
Disacàrids amb fucosa	Fuc-α1,3-GlcNAc Fuc-α1,4-GlcNAc Fuc-α1,6-GlcNAc	*Lotus tetragonolobus* (LTA) & *Ulex europeaus I* (UEA I)
Oligosacàrids amb àcid siàlic	Neu5Ac-α2,3-Gal-β1,4-GlcNAc Neu5Ac-α2,6-Gal-β1,4-GlcNAc Neu5Ac-α2,3/6-Gal-β1,4-Glc	*Maackia amurensis* (MAA) *Sambucus nigra* (SNA) MAA & SNA

El següent pas ha estat completar el perfil d'unió de la lectina PDC-109, avaluant la seva especificitat enfront als diferents epítops glicosídics preparats i validats. Així doncs, mitjançant l'SPR s'ha caracteritzat tant qualitativa com quantitativament la interacció de PDC-109 amb aquests carbohidrats, obtenint una major afinitat pel disacàrid Fuc-α1,4-GlcNAc. Aquest resultat concorda amb estudis anteriors, on es proposa que l'esperma boví s'uneix a un lligand oligosacarídic de l'epiteli oviductal similar a l'epítop Lewis^a (Galβ1-3[Fucα1-4]GlcNAc) en la formació del dipòsit d'esperma. També cal comentar que els sensogrames característics obtinguts de forma generalitzada en SPR per a la mostra PDC-109 han plantejat dificultats pel que fa a l'ajust cinètic, encara que han reflectit les seves propietats intrínseques, a l'hora que corroboren que aquesta proteïna es troba de forma natural com una barreja de diverses isoformes i estats d'agregació, els quals juguen un paper important en la modulació de la seva interacció amb altres biomolècules.

En vista de la funció de PDC-109 en la formació del dipòsit d'esperma per mitjà de la seva interacció amb carbohidrats que contenen residus de fucosa, amb una especificat lectina-carbohidrat corroborada mitjançant la nostra plataforma d'SPR; a continuació hem estudiat el domini de reconeixement dels complexos d'interacció entre PDC-109 i tots els disacàrids amb fucosa (Fuc-α1,(3,4,6)-GlcNAc), utilitzant la tècnica ortogonal de CREDEX-MS. En tots els experiments s'ha pogut identificar un únic fragment peptídic coincident que, com a tal, és molt probable que formi part del lloc d'unió al carbohidrat de PDC-109.

Arribats a aquest punt i tenint en compte els resultats encoratjadors obtinguts en els capítols anteriors, en el **capítol 3** hem decidit aplicar les metodologies optimitzades a l'anàlisi d'interaccions mediades per carbohidrats en sistemes més complexos, utilitzant com a mostra biològica complexa l'esperma boví. Aquest treball forma part d'un projecte d'investigació sobre aspectes moleculars de la reproducció on l'aplicació de les metodologies SPR i CREDEX-MS pretén caracteritzar interaccions carbohidrat-proteïna que regeixen dos esdeveniments claus en la fertilització, com són la formació del dipòsit d'esperma i la unió dels gàmetes en l'espècie bovina.

Com ja s'ha comentat en la introducció, existeixen certes evidències que en la unió tant a l'epiteli oviductal com a la ZP de l'oòcit intervenen carbohidrats que actuen com a receptors de lectines complementàries presents en la superfície de l'espermatozoide. Específicament, s'ha demostrat que la unió d'esperma boví a l'epiteli oviductal implica el reconeixement de fucosa i que, després de la seva capacitació, els espermatozoides són alliberats del dipòsit per avançar en l'oviducte fins a trobar-se amb l'oòcit, on la interacció de gàmetes està intervinguda a través dels residus d'àcid siàlic dels oòcits. Per tant, la identificació de la/es proteïna/es espermàtiques complementàries que participen en aquests processos específics també constitueix un dels objectius d'aquesta tesi.

Els esdeveniments d'unió i alliberament del dipòsit en l'oviducte semblen estar coordinats amb la regulació del procés de capacitació de l'espermatozoide i, a més, aquesta etapa es considera necessària perquè l'espermatozoide pugui fecundar l'òvul. Per aquest motiu, s'ha decidit treballar amb dos poblacions d'espermatozoides diferents: capacitats i no capacitats.

Per a l'estudi de les interaccions lectina-carbohidrat dels dos esdeveniments esmentats, s'ha utilitzat únicament la metodologia CREDEX-MS, ja que l'SPR presentava certes limitacions pel que fa a la utilització d'espermatozoides intactes, degut a restriccions de

mida i complexitat de mostra. Així doncs, s'han preparat diverses columnes d'afinitat amb carbohidrats immobilitzats a sefarosa, escollits d'acord amb la literatura i el seu paper en alguns dels esdeveniments abans esmentats (Figura vii). Aquestes columnes s'han carregat amb espermatozoides bovins intactes vius, en ambdues condicions de capacitació, enlloc de proteïnes d'esperma solubilitzades. Amb aquest plantejament s'elimina la incertesa relacionada amb els tractaments de solubilització o lisi cel·lular i es preserva l'estructura nativa, és a dir, la conformació i estat d'agregació de les proteïnes de superfície de l'esperma. A més, la plataforma permet emular els esdeveniments de reconeixement que ocorren *in vivo* i reduir la complexitat de la mostra resultant per mitjà de l'enriquiment només d'aquelles proteïnes espermàtiques amb afinitat pel carbohidrat exposat.

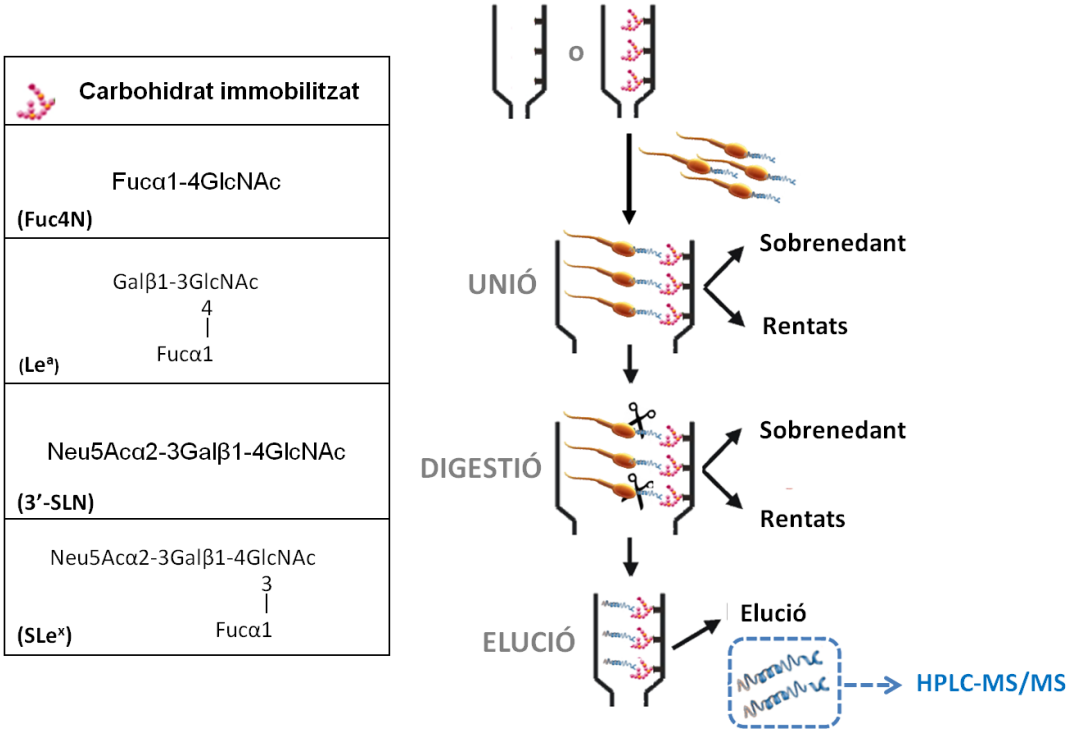


Figura vii. Experiments de CREDEX-MS amb espermatozoides bovins capacitats i no capacitats, utilitzant columnes d'afinitat amb els epítops glicosídics de la taula (immobilitzats) i columnes "blanc".

Tal com s'il·lustra a la Figura vii, els experiments de CREDEX-MS han consistit en: (i) incubació dels espermatozoides capacitats o no capacitats en les diverses columnes carbohidrat-DVS-sefarosa; (ii) eliminació de l'excés d'espermatozoides no units a la columna d'afinitat descartant el sobrenedant i realitzant diversos rentats; (iii) digestió no reductora amb tripsina en la columna dels complexos lectina-carbohidrat formats, seguida de diversos rentats; (iv) elució dels pèptids tríptics amb afinitat pels

carbohidrats immobilitzats que pertanyen al domini de reconeixement; (v) anàlisi proteòmic de la mostra resultant mitjançant LC-MS/MS. Tots els experiments s'han realitzat com a mínim per triplicat, utilitzant com a blanc columnes sense carbohidrat per tal d'avaluar unions inespecífiques o al suport sòlid. El processament i l'anàlisi estadística de les dades obtingudes ha consistit en determinar la presència/absència de proteïnes identificades en les diferents columnes que, mitjançant la tècnica de clusterització de dades, s'han agrupat segons la seva especificitat als diferents carbohidrats en ambdues condicions de capacitació. Amb aquesta aproximació basada en la plataforma de CREDEX-MS s'han pogut aïllar i identificar diverses proteïnes associades a la membrana d'espermatozoides bovins vius. Aquestes presenten afinitats específiques per diferents carbohidrats de les glicoproteïnes de la ZP i l'epiteli oviductal i, per tant, possiblement participen d'alguna manera en els corresponents esdeveniments de fertilització estudiats. A més, la comparació del nostre llistat de proteïnes amb altres llistats publicats ha revelat una nova població de proteïnes prèviament no identificades en l'esperma boví, posant de manifest l'eficàcia de la nostra metodologia i alhora corroborant la complexa composició i funció dels espermatozoides. Cal destacar que hem observat diferències entre les dues condicions espermàtiques avaluades, identificant un major nombre de proteïnes en espermatozoides capacitats. Aquest fet confirma que el procés de capacitació implica modificacions de la membrana dels espermatozoides, incloent alteracions o canvis en l'organització i/o composició de les proteïnes de la seva superfície. Finalment, l'estudi comparatiu d'homologia entre els gens bovins del nostre estudi i els del proteoma d'esperma humà recentment publicats revela la presència de certes proteïnes espermàtiques evolutivament conservades i d'altres pròpies de l'espècie bovina. A més, les proteïnes humanes amb funcions descrites en els processos de fertilització aquí estudiats han determinat com a objecte d'estudi les corresponents proteïnes ortòlogues bovines, el paper de les quals encara resta per descobrir.

Conclusions

La investigació duta a terme en el curs d'aquesta tesi doctoral ha donat lloc a les publicacions científiques incloses en els apèndixs. Les principals conclusions extretes d'aquests articles, així com dels resultats encara no publicats, es poden classificar de la següent manera:

Glicosondes optimitzades per a la plataforma d'SPR:

1. La reacció d'unió quimioselectiva entre el carbohidrat i l'*N*[Me]-Aoa-pèptid assegura una estructura nativa del carbohidrat en la glicosonda necessària pel reconeixement amb la lectina.

Plena explotació de la tecnologia d'SPR en un estudi model:

2. Mitjançant sondes glicopeptídiques d'epítops β-galactosídics preparades i immobilitzades a la superfície del sensor d'SPR, s'ha completat minuciosament l'estudi d'unió a galactosa de la lectina (aglutinina) model *Erythrina cristagalli* (ECA). Les dades cinètiques obtingudes han estat similars a les d'altres estudis publicats, i per tant es constata la validesa del nostre plantejament en l'estudi de les interaccions carbohidrat-lectina.

3. L'aplicació de la plataforma d'SPR a l'anàlisi termodinàmic de la interacció ECA–lacNAc mostra valors d'entalpia comparables als obtinguts per ITC. No obstant, s'observen diferències significatives en els valors entròpics, probablement a causa de les limitacions de transferència de massa.

4. La metodologia d'SPR és totalment compatible amb la captura per afinitat de lectines i la posterior caracterització per MALDI-TOF MS, fent aplicable la nostra metodologia a assajos d'immunoprecipitació.

Expansió de la metodologia d'SPR a altres epítops glicosídics:

5. Els protocols establerts en aquest treball són d'aplicació general per a la preparació d'una àmplia gamma de glicosondes que exposen els epítops glicosídics més rellevants dels mamífers, i que de manera eficient i reproduïble poden ser immobilitzades per a estudis d'unió amb lectines. En particular, la metodologia dissenyada permet resoldre diferències en l'enllaç glicosídic utilitzant quantitats de lectina en el rang de nM a mM.

Transferència de la tecnologia d'SPR a la tècnica ortogonal CREDEX-MS:

6. Estudis d'RMN confirmen que l'enllaç carbohidrat-DVS present en la plataforma CREDEX-MS es produeix a través del carboni anomèric del carbohidrat amb l'extrem reductor, assegurant el tancament de l'anell i deixant lliure i en estat natiu l'epítop responsable de la interacció carbohidrat-lectina.

7. La tècnica de CREDEX-MS és totalment complementària amb l'SPR. El principal avantatge de CREDEX-MS és la seva capacitat per definir el domini de reconeixement de lectines als carbohidrats.

Aplicació de les metodologies desenvolupades en sistemes biològicament rellevants:

8. La complementarietat d'SPR i CREDEX-MS s'ha demostrat amb èxit en la mostra biològica de PDC-109. Per mitjà d'ambdues tècniques s'ha avaluat el perfil d'unió d'aquesta lectina envers els epítops glicosídics més rellevants dels mamífers, s'ha corroborat la seva naturalesa polidispersa, i s'ha determinat el domini d'unió als carbohidrats, amb resultats consistents amb la bibliografia.

9. La tècnica de CREDEX-MS és adaptable a mostres biològiques complexes com ara el plasma seminal, permetent treballar amb cèl·lules intactes d'espermatozoides i proporcionant un assaig de bioafinitat que mimetitza, en la mesura del possible, la situació fisiològica. Com a resultat, les proteïnes de la superfície de l'espermatozoide amb afinitats específiques per a diferents epítops glicosídics han pogut ser enriquides, identificades i relacionades amb els esdeveniments de fertilització estudiats en l'espècie bovina.

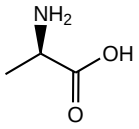
10. L'anàlisi exhaustiva de les proteïnes espermàtiques identificades, juntament amb les ja conegudes i les presents en altres tipus de cèl·lules dels tractes genitals femení i masculí, proporcionarà una major comprensió dels mecanismes moleculars que controlen el procés de reproducció en els mamífers, el que serà de gran valor en el camp de la reproducció assistida.

ABBREVIATIONS

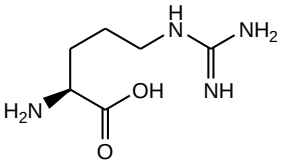
AAA	amino acid analysis
AAL	*Aleuria aurantia* lectin
Ac	acetyl
ACN	acetonitrile
aSFP	acidic seminal fluid protein
BSA	bovine serum albumin
BSP	bovine seminal plasma protein
CBP	carbohydrate binding protein/peptide
CE	capillary electrophoresis
CHCA	α-cyano-4-hydroxycinnamic acid
CID	collision-induced dissociation
CM	capacitating medium
CMD	carboxymethyl dextran
Con A	*Concanavalin A* (*Canavalia ensiformis)*
CRD	carbohydrate recognition domain
CREDEX	Carbohydrate REcognition Domain EXcision
CV	coefficient of variation
Da	dalton (atomic mass unit)
DEAE	diethylaminoethanol
DHB	2,5-dihydroxybenzoic acid
DMF	*N*,*N*-dimethylformamide
DVS	divinylsulfone
ECA	*Erythrina cristagalli* agglutinin
ECD	electron capture dissociation
EDC	1-ethyl-3-(3-diethylaminopropyl)-carbodiimide
ELISA	enzyme-linked immunosorbent assay
ELLA	enzyme-linked lectin assay
ER	endoplasmic reticulum
ESI	electrospray ionization
ETD	electron transfer dissociation
FAC	frontal affinity chromatography
Fc	flow cell
FDR	false discovery rate
FID	flame ionization detector
FTICR	Fourier transform ion cyclotron resonance
GAG	glycosaminoglycan
GC	gas chromatography
GO	gene ontology
HBS	HEPES buffered saline
HCD	high-energy collision dissociation
HDL	high-density lipoprotein
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HIA	hemagglutination inhibition assay
HILIC	hydrophilic interaction liquid chromatography

HPAEC	high performance anion exchange chromatography
HPLC	high performance liquid chromatography
IC_{50}	half maximal inhibitory concentration
IDAWG	isotopic detection of aminosugars with glutamine
ITC	isothermal titration calorimetry
IVF	*in vitro* fertilization
LC	liquid chromatography
LIF	laser-induced fluorescence
LTA	*Lotus tetragonolobus* agglutinin
M	molar – mole per liter
MAA	*Maackia amurensis* agglutinin
MALDI-TOF	matrix assisted laser desorption ionization time of flight
MES	2-(*N*-morpholino)ethanesulfonic acid
MOPS	3-(*N*-morpholino)propanesulfonic acid
MS	mass spectrometry
MVS	methyl-vinylsulfone
NCM	non-capacitating medium
NHS	*N*-hydroxysuccinimide
NMR	nuclear magnetic resonance
OEC	oviductal epithelial cells
PAD	pulsed amperometric detection
PBS	phosphate saline buffer
PC	phosphatidylcholine
pI	isoelectric point
PNGase F	Peptide-N4-(*N*-acetyl-beta-glucosaminyl) asparagine amidase
ppm	parts-per-million
PTM	post-translational modification
PVP	polivinilpirrolidone
QCM	quartz crystal microbalance
RP	reverse-phase
rpm	revolutions per minute
r.t.	room temperature
RU	resonance units
SA	sinapinic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sia	sialic acid
SN	supernatant
SNA	*Sambucus nigra* agglutinin
SPR	surface plasmon resonance
T	temperature
TALP	Tyrode's albumin lactate pyruvate medium
TFA	trifluoroacetic acid
TMSP	trimethylsilyl propanoic acid
t_R	retention time
Tris	tris(hydroxymethyl)aminomethane
UEA-I	*Ulex europaeus* agglutinin I
UPLC	ultra performance liquid chromatography
UV	ultraviolet
ZP	*zona pellucida*

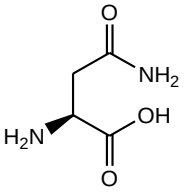
ANNEX I: AMINO ACIDS



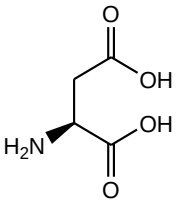
L-Alanine
Ala
A



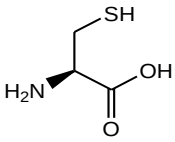
L-Arginine
Arg
R



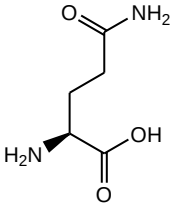
L-Asparagine
Asn
N



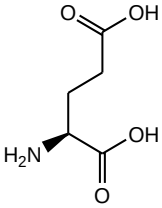
L-Aspartic acid
Asp
D



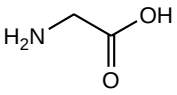
L-Cysteine
Cys
C



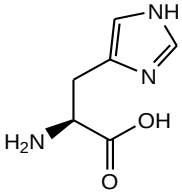
L-Glutamine
Gln
Q



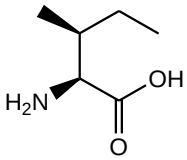
L-Glutamic acid
Glu
E



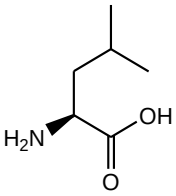
L-Glycine
Gly
G



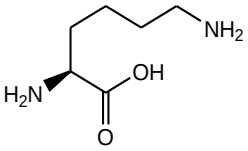
L-Histidine
His
H



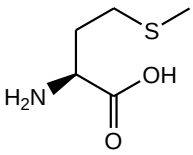
L-Isoleucine
Ile
I



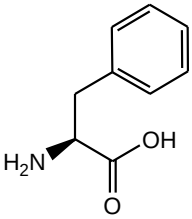
L-Leucine
Leu
L



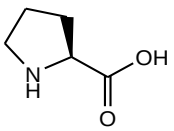
L-Lysine
Lys
K



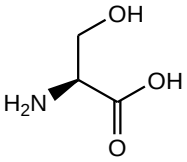
L-Methionine
Met
M



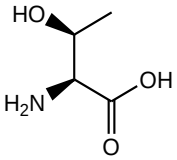
L-Phenylalanine
Phe
F



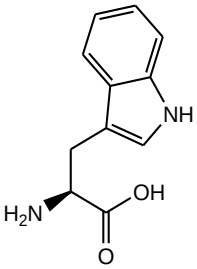
L-Proline
Pro
P



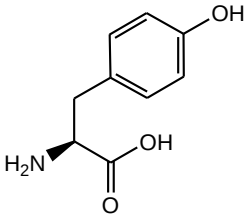
L-Serine
Ser
S



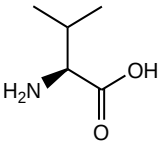
L-Threonine
Thr
T



L-Tryptophan
Trp
W

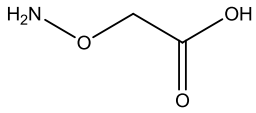


L-Tyrosine
Tyr
Y

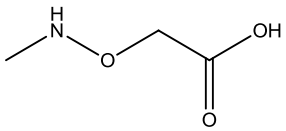


L-Valine
Val
V

Non-proteinogenic amino acids

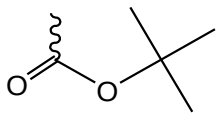
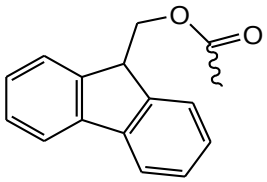
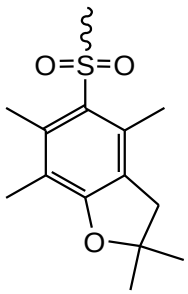
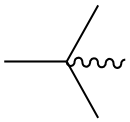
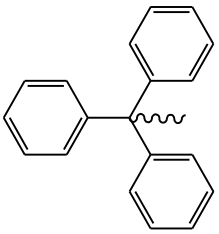


Aminooxyacetic acid
Aoa

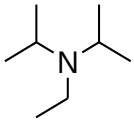
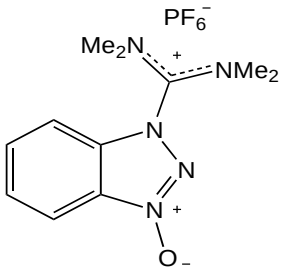


N-methyl-aminooxyacetic acid
N[Me]-Aoa; Aoa(m)

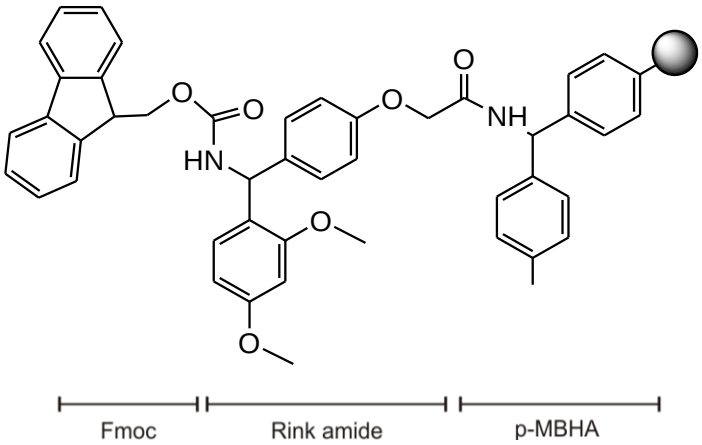
ANNEX II: PROTECTING GROUPS

Abbreviation	Name	Structure
Boc	*tert*-butyloxycarbonyl	
Fmoc	9-fluorenylmethyloxycarbonyl	
Pbf	2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl	
tBu	*tert*-butyl	
Trt	trityl	

ANNEX III: COUPLING REAGENTS AND ADDITIVES

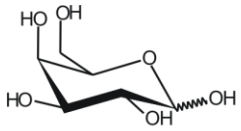
Abbreviation	Name	Structure
DIEA	*N*,*N*-diisopropylethylamine	
HBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate	
HOBt	*N*-hydroxybenzotriazole	

ANNEX IV: RESIN

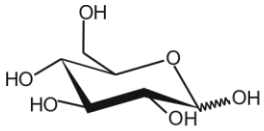
Name	Structure
Fmoc-Rink-amide-MBHA resin	

ANNEX V: CARBOHYDRATES

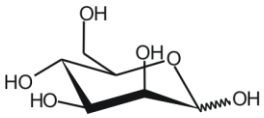
Monosaccharides



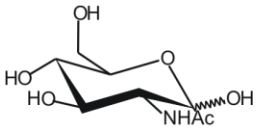
Galactose
Gal



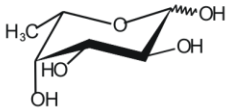
Glucose
Glc



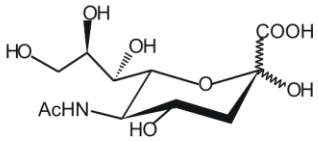
Mannose
Man



N-acetyl-glucosamine
GlcNAc

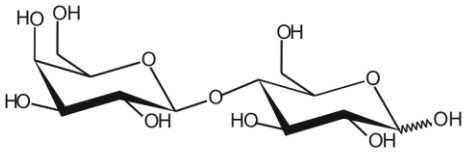


Fucose
Fuc

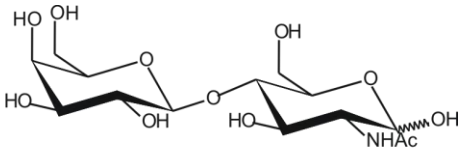


N-acetyl-neuraminic acid
Neu5Ac
(Sialic acid; Sia)

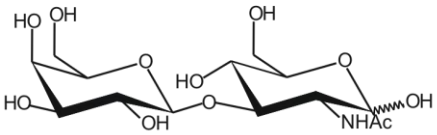
Disaccharides with terminal β-galactose



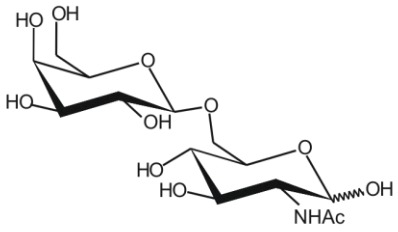
Gal-β1,4-Glc
lactose (lac)



Gal-β1,4-GlcNAc
N-acetyllactosamine (lacNAc)

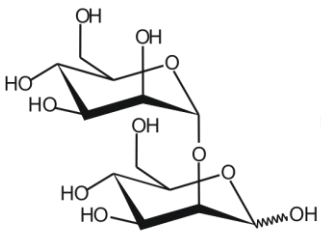


Gal-β1,3-GlcNAc

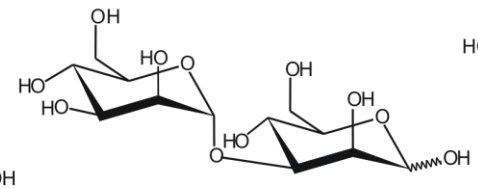


Gal-β1,6-GlcNAc

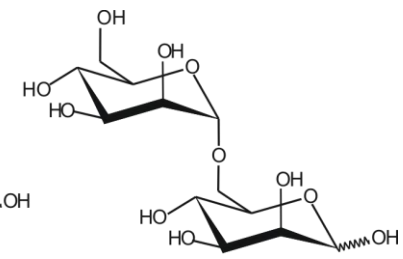
Mannobioses



Man-α1,2-Man

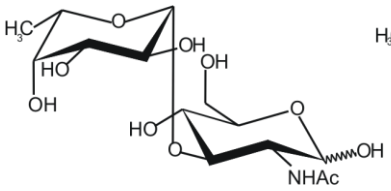


Man-α1,3-Man

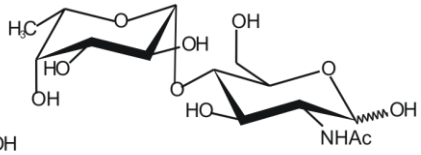


Man-α1,6-Man

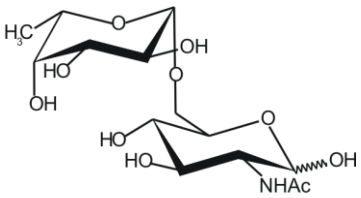
Fucosylated disaccharides



Fuc-α1,3-GlcNAc

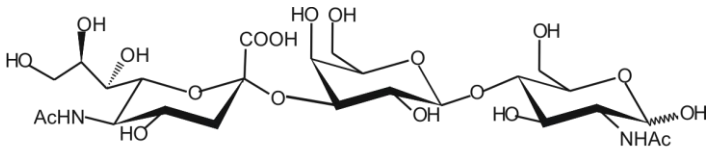


Fuc-α1,4-GlcNAc
(Fuc4N)

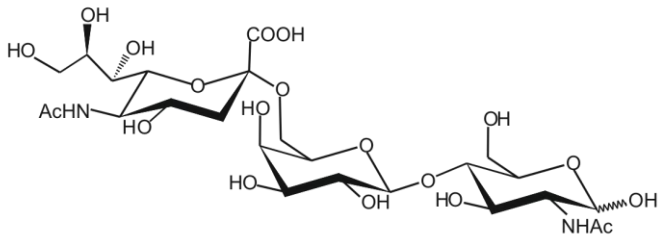


Fuc-α1,6-GlcNAc

Sialyl-containing oligosaccharides

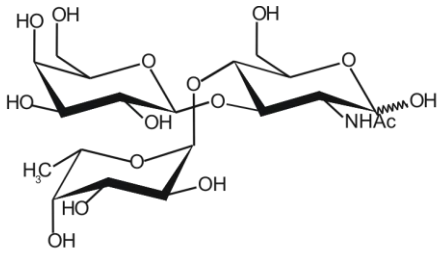


Neu5Ac-α2,3-Gal-β1,4-GlcNAc
(3'-SLN)

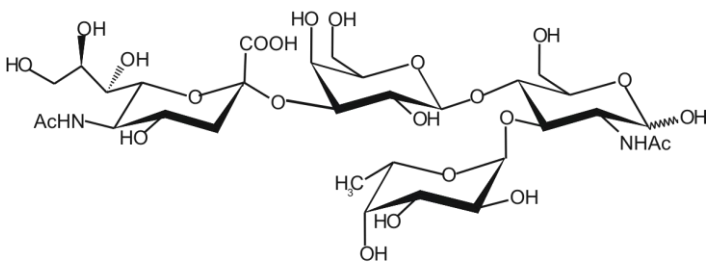


Neu5Ac-α2,6-Gal-β1,4-GlcNAc

Lewis blood group antigens



Gal-β1,3-[Fuc-α1,4-]GlcNAc
Lewis[a] (Le[a])



Neu5Ac-α2,3-Gal-β1,4-[Fuc-α1,3-]GlcNAc
Sialyl Lewis[x] (SLe[x])

GENERAL INTRODUCTION AND OBJECTIVES

GENERAL INTRODUCTION

The world of carbohydrates is extremely complex, rendering it fascinating and at the same time fearsome to those attracted to the challenge of unveiling its structural intrincacies. The term carbohydrate spans many different disciplines ranging from large-scale industrial applications to fine-tuned biomedical uses, and the science dealing with carbohydrates has experienced ups and downs over the last decades in terms of attention paid, importance attributed, and level of understanding reached. Recently, glycochemistry and glycobiology fields are enjoying renewed interest at both basic and applied (biomedical, pharmaceutical) levels, as clearly evidenced by the >500 reviews on the subject over the past 18 months. Most efforts are devoted to the study of carbohydrate-mediated biomolecular interactions and glycoprotein engineering but the structural analysis of carbohydrates, in all its aspects, remains the basis of nearly all the developments of recent times.

i. The glycan code

Carbohydrates, one of four major classes of biomolecules, along with proteins, lipids and nucleic acids, are the most abundant class of organic compounds found in living organisms and, by virtue of their diversity and coding capacity, the information carriers by excellence. Carbohydrate diversity and consequent complexity arises primarily from the structural variety at the monosaccharide level (Figure 1), where multidirectional combinations of different monosaccharide building blocks, linkages[1], anomericity, and branching generate a vast number of complex glycan structures (polysaccharides) that can be further modified by sulfation, acetylation, methylation, phosphorylation, etc., and/or linked covalently to non-carbohydrate moieties such as peptides (in different ways) or lipids forming the corresponding glycoconjugates (*i.e.* glycoproteins and glycolipids, respectively) with particular roles and localizations in living organisms.

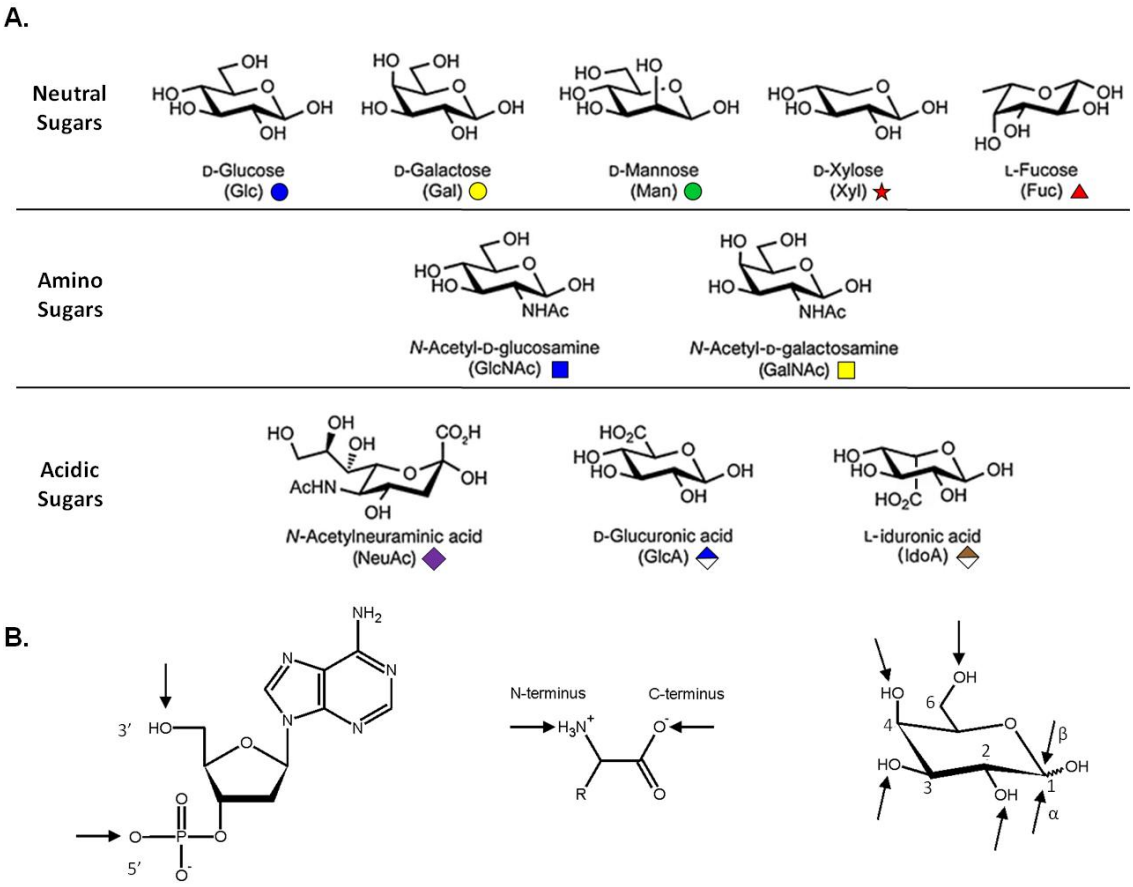


Figure 1. Structural diversity at the monosaccharide level: (**A**) Most commonly found monosaccharides in higher animals are represented using chain conformation. Corresponding abbreviation and symbolic representation accorded by NCBI and Consortium for Functional Glycomics in 2004 are depicted for each monosaccharide. (**B**) Illustration of the three coding systems and their corresponding elongation points (marked by arrows). Nucleic acids (left) and proteins (center) are almost exclusively linear and they have only a single type linkage (*i.e.* 3'-5' phospho diester bonds for nucleic acids and amide bonds between C- and N-terminus for proteins), whereas carbohydrates (right) can be elongated through hydroxyl groups at C2-C6 plus the two anomeric hydroxyl positions at C1 (α and β).

In the mammalian system, over 80% of all (membrane and secretory) proteins are glycosylated. Glycans can be attached to polypeptide structures through amide linkages to Asn side chains (N-glycosylation), through glycosidic linkages (O-glycosylation) to side chains of Ser/Thr, hydroxylysine (collagen) or Tyr (glycogenin), or through C-C linkages to the C2 position of Trp (C-mannosylation). Alternatively, they can be attached as a linker structure bridging glycosylphosphatidylinositol (GPI) anchors to protein backbones (Glypiation).

N-glycosylation sites can be predicted by the sequon Asn-Xaa-Ser/Thr (where Xaa is any amino acid except proline). For N-linked glycans, when the nascent protein enters the endoplasmic reticulum (ER) a block of sugars (*i.e.* $Glc_3Man_9GlcNAc_2$ core) is

transferred to the amino group in the asparagine side chain. The glycan is later processed and modified in the ER and Golgi, generating the three types of N-glycans found in mature glycoproteins: high mannose, complex and hybrid (Figure 2; upper panel). The vast majority of membrane and secretory glycoproteins are of the complex-type.

O-glycans are typically attached to proteins via Ser and Thr residues. Although a range of monosaccharides linked to Ser/Thr have been found (*e.g.*, GalNAc, GlcNAc, Xyl, Man, Fuc and Gal), the most common modification is the attachment of GalNAc to Ser/Thr (mucin-type O-glycans). Extension of this GalNAc residue can generate eight different core structures (Figure 2; lower panel); cores 1–4 are frequent in humans. Further outer modifications, as with N-glycans, are attached to inner cores, generating a wide structural diversity of glycan epitopes.

While the complexity and diversity of the total of glycan structures in an organism is almost impossible to calculate, some 7,000 glycan determinants (glycotopes) recognized by carbohydrate binding proteins (CBPs) including lectins, receptors, toxins, antibodies, and enzymes has been reckoned for the human glycome[2]. This value is probably underestimated but it provides an idea of the dimension generated by the approximately 700 proteins that make up the mammalian glycan repertoire, and sets the boundaries for glycan-CBP interaction studies[3] as well as establishes protein glycosylation as the most abundant, complex and structurally diverse post-translational modification.

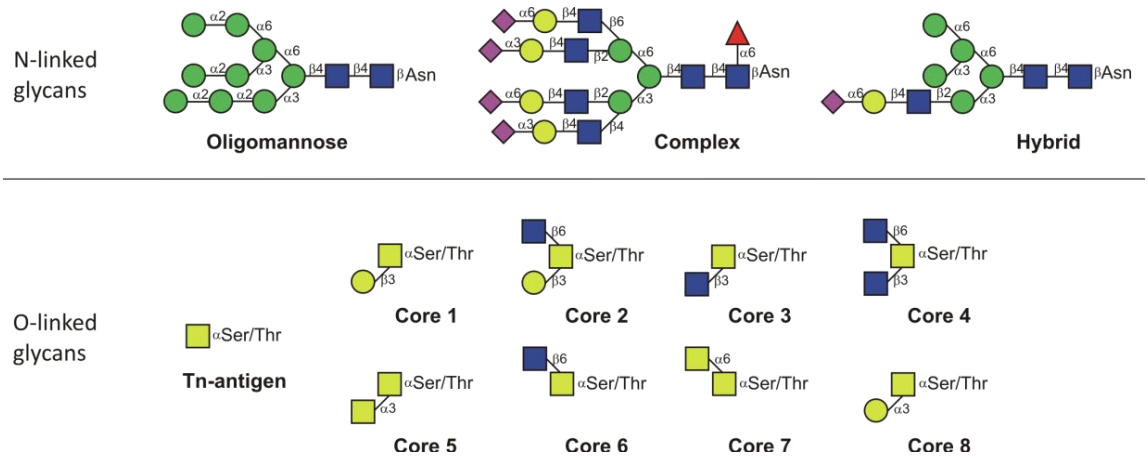


Figure 2. Structural diversity in N- and O-glycans. N-glycans (upper panel) are attached to the protein backbone by linkage to asparagine in the sequon Asn-Xaa-Ser/Thr. The N-glycan core common to complex, high mannose and hybrid structures is $Man_3GlcNAc_2$. The O-glycans (lower panel) are commonly attached to the protein via Ser or Thr residues. The eight different cores described for O-glycans are shown. Structures can be extended from the basic cores by lactosamine repeats, fucosylation and sialylation.

Additional diversification derives from the fact that, whereas the process of protein synthesis follows a well-defined, genetically encoded linear process, protein glycosylation is a non-template-driven, secondary gene event that is initiated during protein synthesis and involves a large collection of redundant and overlapping enzymes (glycosidases and glycosyltransferases) that are partially compartmentalized throughout the ER and the Golgi system[4]. Various competing reactions in the processing pathways, the need for enzyme, acceptor and substrate simultaneously, as well as other physiological factors contribute to glycan microheterogeneity, *i.e.*, glycoprotein isoforms resulting from different glycans at a given site. This heterogeneity may be relatively simple, such as for RNAse B[5], or rather complex as in the case of CD59 where at least 123 different desialylated glycan variants have been identified at a single site[6]. Macroheterogeneity results from the fact that not all N-glycan sequons (or Ser/Thr present in the glycoproteins) are necessarily glycosylated.

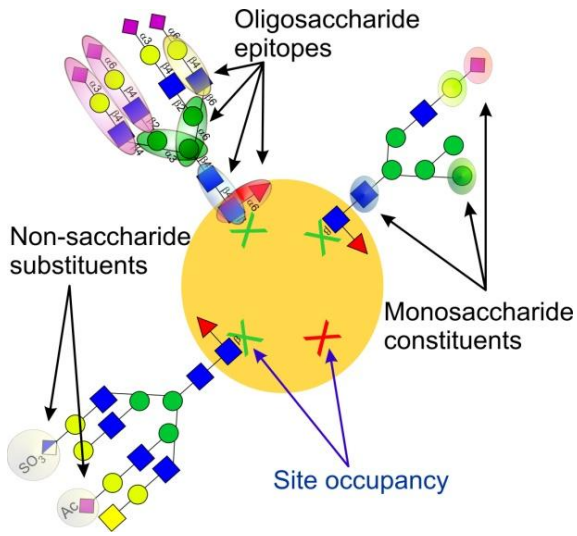


Figure 3. Different levels at which carbohydrates contribute to glycoconjugate heterogeneity: occupancy, the nature of monosaccharide constituents building up the structure, the specific epitopes composed by the monosaccharides, and non-carbohydrate substituents.

Eventually, such diversity (Figure 3) gives rise to a wide array of glycoforms, in both soluble and membrane-anchored forms, that are as essential to life as a genetic code, and constitute an evolutionary conserved feature of all living cells[7]. The identification of the number, structure, and function of glycans in a particular biological context, a task initiated decades before the "omics" boom, has been recently termed *glycomics*, and substantial progress has been made in the understanding of how glycans are directly involved in almost every biological process and in nearly every human disease[8]. Still, the glycome is far more complex than the genome, transcriptome, or proteome, given its

much more dynamic character, which varies considerably not only with cell or tissue type, but also with developmental stage[9], metabolic state, or with changes such as disease[10], aging[11,12], environmental factors[13], or evolution [14,15]. For instance, epigenetic regulation may induce novel glycan structures that make the organism fitter in a specific environment without altering genetic information[16]. It is therefore of the utmost importance to know what carbohydrate structures decorate which glycoproteins under particular conditions.

ii. Structural analysis of glycans

Accurate analysis of glycosylation is still considered an extremely challenging task even by those specialists dedicated to this fascinating bioorganic conundrum. This is because of the many different physical parameters that must be established before a structural characterization can be considered complete (Figure 4). As a consequence, there is no single analytical method capable of providing all the necessary information for rapid and reliable identification and quantification of a particular structure, let alone for establishing its particular functionality. The job usually requires a combination of several orthogonal, physical, chemical, and biochemical techniques in a multidimensional approach as depicted in Scheme 1.

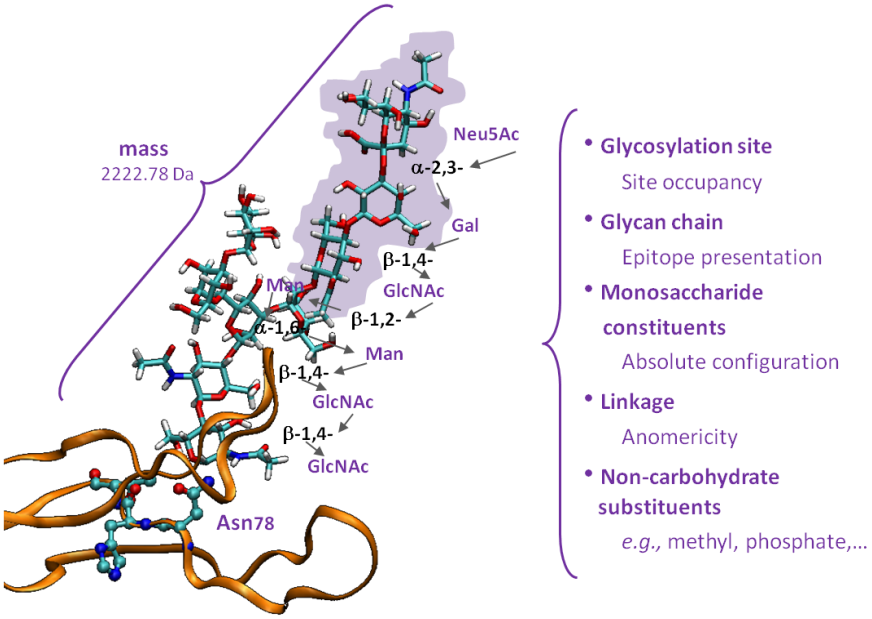
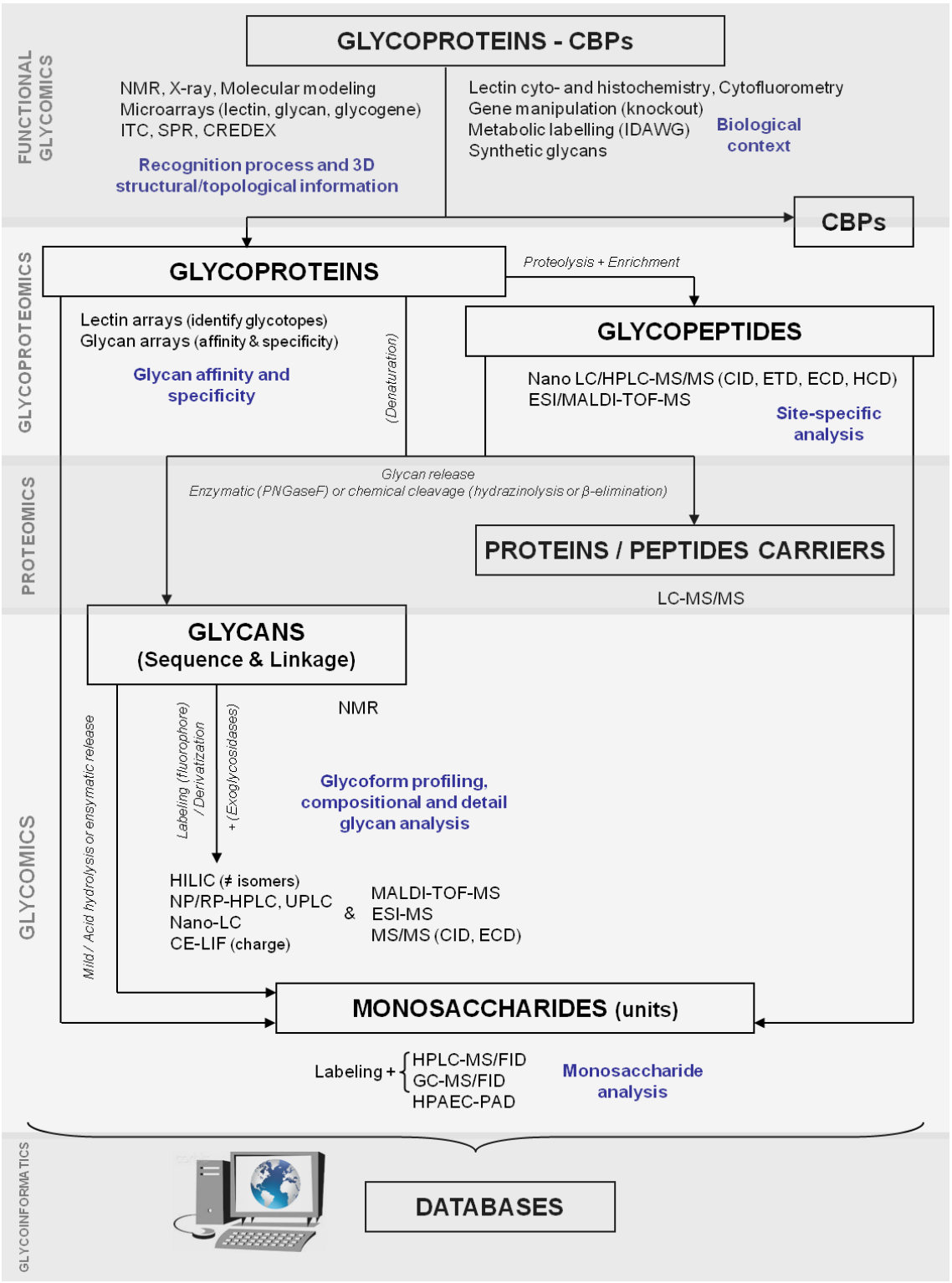


Figure 4. Structural elements in the full structural elucidation of glycoproteins. As an example, one of the N-glycans in the human chorionic gonadotropin (hCG) glycoprotein α-chain is represented. Elements to be specified are listed on the right and some of them displayed on the structure representation. The shaded part represents the potential epitope to be recognized in a carbohydrate-driven interaction.

Scheme 1. Different levels of glycan analysis including compositional glycan analysis, highly detailed glycan structural analysis, glycan affinity and specificity, glycoform profiling, site-specific analysis and 3D structural and topological information studies. Moreover, experimental approaches to determine possible carbohydrate binding protein (CBPs) structures as well as to characterize glycan-CBP recognition processes and complexes formation are required, particularly in biological contexts. Finally, advanced glyco-informatic resources are essential for the annotation, collection of analytical results and proper analysis of the large-scale data generated. The abbreviations are listed in the specific section included in this thesis.



Over the last 2-3 decades, the continuous refinement of analytical tools has greatly facilitated glycan analysis; numerous reviews[17-22] and research articles cover the main technologies routinely used today for N- and O-linked glycan analysis, including capillary electrophoresis (CE)[23-25], liquid chromatography (LC)[26,27], mass spectrometry (MS)[28-31] and microarray-based[32-36] approaches to glycomics and glycoproteomics[29,37,38].

Glycosylation analysis can be tackled at different levels, namely: (i) characterization of glycans in intact glycoproteins, (ii) characterization of glycopeptides and (iii) structural analysis of chemically or enzymatically released glycans. The diverse technologies available for each approach have been summarized in Scheme 1. It is important to stress that in all these techniques a compromise exists between analytical sensitivity and the degree of structural detail provided. Hence, before embarking on glycosylation analysis it is important to ascertain: (i) which level of information is required, (ii) whether quantitative or qualitative data are needed and (iii) how much (purified) material is available, as these factors will determine the technologies that are best suited to address the problem and decipher that part of the puzzle.

At the operational level, glycomics can be addressed through a variety of strategies and technologies that turn out to be orthogonal rather than parallel. While all of them rapidly generate very large amounts of data, differences between platforms can turn data analysis into a complex, time-consuming task requiring bioinformatics tools and databases to facilitate data processing and interpretation. Most of these glycoinformatic tools have particular focuses, *e.g.*, data from HPLC[39], MS[40], NMR or microarray[41,42] experiments. Initiatives for cross-linking data from different techniques and integrating multiple data sets are prospering and extremely useful[43,44], although in the use of database search outputs critical interrogation is advisable.

iii. Functional analysis of glycans

The chemical and biological diversity of carbohydrates gives rise to a structural complexity that underlies their functional variety. Thus, glycosylation is not only important for protein folding and stability[45,46] but also in various biological processes and recognition events (Figure 5). These roles may be unrelated to the close structural environment where glycosylation occurs or, to the contrary, very stringent in terms of glycotope structure and protein localization. Also, the functions exerted are very diverse

including: i) structural, organizational and stabilizing roles, ii) protective or barrier functions, iii) provision of specific receptors for microorganisms, toxins or antibodies to attack, shield or lure, iv) modulation of protein functions in a glycosylation-dependent manner, v) intra- and intercellular trafficking roles, and vi) mediation of cell-matrix or cell-cell interactions[47,48]. Therefore, no particular function can or should be attributed to a given oligosaccharide, so that general statements on the subject are practically impossible. The only common general principle emerging from the numerous functions is that glycans generate important functional diversity required for the development, differentiation, and crosstalk in complex organisms as well as for their interactions with other organisms in the environment.

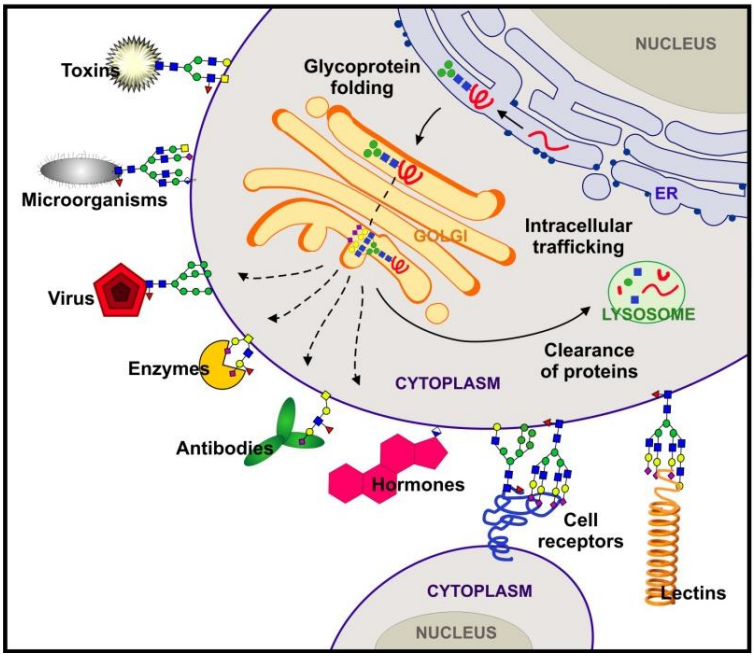


Figure 5. Glycans participate in multiple mechanisms of cellular regulation. The general functions of glycans span from nascent protein folding and intracellular trafficking to roles in extracellular compartments such as cell-cell communication, providing specific receptors for noxious agents, protect from microorganisms and antibodies or regulate myriad receptor-ligand interactions.

Apart from being directly involved in almost every biological process, glycans certainly play a major role in nearly every human disease, cancer very much included. Many pathological states are characterized by changes in the carbohydrate structure of cellular glycoproteins (*i.e.* aberrant glycome composition), and in some cases the modifications have been traced to alterations in the activity of specific glycosidases and/or transferases[49]. Post-glycosylational modifications of specific sites within the glycan chain such as sulfation, acylation, phosphorylation, methylation and epimerization can

also modulate carbohydrate biological function and play a critical role in many normal and disease processes[50]. From this, it easily follows the considerable interest in the identification of glycan profiles of particular glycoproteins, body fluids or tissues under healthy or disease conditions. For instance, since in most cancers fucosylation and sialylation levels are significantly modified, such aberrant glycan structures can become useful glyco-biomarkers[51-55].

A variety of feasible approaches have been extensively used to unravel the biological role(s) of the carbohydrate units of glycoproteins. They include enzymatic or chemical removal of completed sugar chains, modification of glycans by purified glycosidases and transferases, use of inhibitors of glycosylation or glycoprotein processing and of cell mutants with known defects in glycosylation and, more recently, techniques of molecular genetics. For a small number of glycoproteins, modulation of the physicochemical properties or biological activities by their glycans has been demonstrated, but for the great majority of these compounds, the definite role of the carbohydrate remains an enigma. One thing is clear: glycosylation can have markedly different effects on different proteins. This means, that each glycoprotein must be examined individually and meticulously for the possible functions of the glycans it carries.

On the other hand, the accumulation of evidence on the role of carbohydrates as recognition molecules is most rewarding. Over time it has been demonstrated that glycoproteins are fundamental in many important biological processes including fertilization, immune response, inflammation, viral replication, parasite infection, cell growth, cell-cell adhesion, or clearance of glycoproteins. In this context, one must also bear in mind that identifying the carbohydrate binding entity is as important as deciphering the cognate sugar epitope. The vast majority of studies conducted today are performed within the constraints dictated by either physiological or technical boundaries. In an ideal situation, the analysis of biological interactions with glycoprotein participation and directly triggering a physiological response would be performed *in situ*, in real time, and without external intervention. However, this goal is as yet unattainable and state-of-the-art approaches still require the use of chemical and/or biological labeling strategies or the analysis under *in vitro* conditions where the biological context is greatly reduced to the cellular level.

Therefore, one of the major challenges that glycoscience has faced since its very beginning remains: handling the glycoproteome at the endogenous level, addressing the complexity in an automated high-throughput mode, analyzing glycoproteins in complex samples with simultaneous characterization of both the glycan moieties and the corresponding protein carriers. Novel instrumental developments or the intelligent hyphenation of orthogonal existing techniques such as combining front-end biomolecular interaction analysis with in-line mass spectrometric evaluation, will be required to meet this challenge which will always constitute the first step in understanding the biological function of a glycoconjugate. In this respect, the integration of glycomics with other –omics fields such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics[56] will certainly rank glycomics according to its merits. No doubt the current efforts towards systems glycobiology modelling, *i.e.* coupling biochemical knowledge and mathematics into *in silico* models of the cellular glycosylation system, will largely contribute in this respect[57]. Evidently, a broad picture of how glycosylation is regulated through omics-data acquisition and systematic integration will be an enormously valuable asset to gain understanding of glycan functions as well as to develop clinical diagnostics and glyco-biomarker discoveries[54]. Such systems-level studies will help establish novel quantitative and mechanistic links between gene expression, protein expression, enzyme activity, carbohydrate structure and glycoconjugate function.

For a more detailed discussion on some of these matters, the interested reader is referred to the review *Mammalian protein glycosylation – structure versus function* provided as Appendix I.

GENERAL OBJECTIVES

There is a growing awareness of the biological importance of carbohydrates on cell surfaces and circulating proteins, on account of their role in cellular communication and cell-pathogen recognition. However, only a number of these interactions have been elucidated and their interacting partners characterized. For this reason, there is considerable interest in devising powerful analytical tools to study these molecular recognition events in detail.

The work described in this thesis builds on previous research from our group, where a glycopeptide probe was introduced as a tool for displaying carbohydrate epitopes on sensor surfaces. From that starting point, the present work has evolved with a view to extend the application field of the initial approach to other carbohydrates relevant in biological processes. Thus, the following goals were set:

- Design, elaboration and characterization of new glycoprobes, documenting in detail their binding properties using well-defined lectins (chapters 1 and 2).
- Expanding the approach beyond the proof-of-principle stage, in order to detect, characterize and identify carbohydrate-binding proteins (lectins) in biological systems, with a focus on the bovine model and fertilization (chapters 2 and 3).

CHAPTER 1:

Carbohydrate-driven biomolecular interactions

1 Carbohydrate-driven biomolecular interactions

1.1 INTRODUCTION

Molecular recognition events are central to living processes. Of the many interacting molecules existing in nature, it has become clear that carbohydrates and their conjugation to proteins or lipids play key roles in a variety of biological phenomena, and that many carbohydrate-mediated processes are associated with cell-cell/matrix interactions, cell proliferation and other important biological events such as fertilization or brain development[58]. Furthermore, carbohydrate-mediated interactions are also implicated in diverse disease mechanisms, from inflammatory processes to bacterial/viral infections[59] or cancer metastasis[60].

The ubiquity of these carbohydrate-mediated interactions is not surprising given that nearly all secreted and membrane-associated proteins are glycosylated, and a survey of the make-up of the cellular surface reveals that it is abundant in carbohydrate-modified species (*i.e.* glycolipids, glycoproteins, and glycosaminoglycans).

Therefore, the extended presence and biological importance of carbohydrate-containing molecules clearly indicates the need to develop novel –and to improve existing– chemical, biochemical and instrumental analytical techniques for the reliable detection, identification and structural characterization of glycospecies. In this context, the detection principles of carbohydrates and glycoforms are often based on selective capture of sugar moieties using specific receptors, *i.e.* lectins, followed by the characterization of sugar-lectin interaction.

1.1.1 **Lectins: the glycocode readers**

Nature appears to have taken full advantage of the vast diversity of glycans expressed in organisms by evolving protein modules to recognize and differentiate discrete carbohydrate epitopes that act as recognition determinants and mediate specific physiological or pathological communication processes. These ubiquitous proteins of non-immune origin (*i.e.* other than enzymes or antibodies), and capable of decoding the biological information contained in carbohydrates are called lectins (lat. *legere* meaning "to choose" or "select")[61]. They recognize and reversibly bind to specific

monosaccharides or oligosaccharide structural epitopes, and their interaction resembles the antibody-antigen and enzyme-substrate reactions. However, lectin-carbohydrate interactions are relatively weak when compared to others in nature, with dissociation constants often of the order of mM for monosaccharides. This means that they are deliberately transient in nature, though sufficient for communication purposes.

The beginnings of lectinology date back to 1888, when Herrmann Stillmark described the agglutination properties of ricin[62]. Then in 1919, the first pure lectin, concanavalin A, was isolated by Sumner and Howell, who also demonstrated its sugar specificity[63]. Subsequently, lectins played a crucial role in elucidating the molecular basis for blood group specificity[64]. During recent years[65], lectins have come a long way and many primary and three dimensional structures of lectins have been elucidated. It was observed that lectins from distinct sources lacked primary sequence similarity but shared similarities in their tertiary structures. Structural studies conducted on animal lectins also suggested that the carbohydrate binding activity of most lectins resides in a limited polypeptide segment, designated as the carbohydrate recognition domain (CRD)[66]. The CRD typically recognizes the terminal non-reducing carbohydrate residues of cell membrane glycoconjugates, and lectin CRDs also may discriminate between anomeric isomers as a function of their specificities.

Lectin structural advances have allowed the replacement of more traditional divisions according to lectin origin (*e.g.*, plants, animals and microorganisms) by a classification based on common structural features[67]. Thus, most lectins are now thought to belong to three classes: (1) simple, (2) mosaic (or multi-domain) and (3) macromolecular assemblies. Simple lectins consist of a small number of subunits with molecular weights below 40 kDa. Each monomeric unit contains a carbohydrate-binding site. This class comprises practically all known plant lectins (also called legume lectins) and most members of the galectin family. Mosaic lectins are composite molecules consisting of several kinds of protein domains, only one of which possesses a carbohydrate-binding site. This class includes several proteins from different sources: viral hemagglutinins and animal lectins of C-, P- and I-types. Macromolecular assemblies, consisting of polymeric subunits arranged helically and assembled in the form of *fimbriae*, are common in bacteria. Within each class, proteins can be grouped into families, with similar sequences and structural properties. There are typically three classical lectin families—legume lectins, C-type lectins, and galectins—as well as other plant lectins,

viral proteins (*e.g.*, influenza hemagglutinin), toxins (*e.g.*, cholera toxin), anti-carbohydrate antibodies, and pentraxins.

Current research on lectins can be divided into two major parts. The first is based on the characterization of their molecular structures, their biological roles and carbohydrate-binding specificity. The second is focused on the application of lectins with known specificity as potential tools for isolation, identification, characterization and study of either free forms of glycans or their conjugates with lipids and proteins, both in solution and on solid surfaces[68]. In particular, simple plant lectins[69], the largest and most thoroughly studied family, are more extensively used for these latter purposes in an analytical environment, given the fact that they can be purified in sufficient quantities, are well-characterized, and cover nearly all known monosaccharides[70] (see Table 1.1).

Table 1.1. Examples of plant lectins classified according to their carbohydrate binding affinity.

Glycan ligands	Plant lectin source	Abbreviation
Glc/Man	*Canavalia ensiformis*	Con A
Gal	*Erythrina cristagalli*	ECA
	Ricinus communis	RCA
	Griffonia simplicifolia	GSA
GlcNAc	*Triticum vulgaris*	WGA
GalNAc	*Dolichos biflorus*	DBA
	Psophocarpus tetragonolobus	PTA
Neu5Ac	*Maackia amurensis*	MAA
	Sambucus nigra	SNA
Fuc	*Lotus tetragonolobus*	LTA
	Ulex europaeus	UEA
	Aleuria aurantia	AAL

It should be noted that, despite the high sequence homology, the legume family displays enormous diversity in carbohydrate specificity. Specific lectins towards almost every monosaccharide present in mammals (Glc, Gal, Man, Fuc, GlcNAc, GalNAc, Neu5Ac), have been isolated from natural sources (seeds, roots, bulbs, bark and leaves; *i.e.* storage organs of plants) and characterized.

1.1.1.1 *Lectin - carbohydrate interactions*

Carbohydrates interact with lectins through hydrogen bonds, metal coordination, van der Waals and hydrophobic interactions[71] (Figure 1.1). Hydrogen bonds are common in saccharide structures due to the presence of hydroxyl, amine, and carboxyl groups. The availability of large numbers of hydroxyl groups on carbohydrates enables the formation of a complex network of hydrogen bonds in which the hydroxyl group can act both as a donor and an acceptor either directly or water-mediated. In particular, for lectins that recognize neutral hexoses (such as Glc, Man or Gal), the hydroxyl at the C4 position seems to be a decisive player, as differences in the spatial orientation of C4 (axial in Gal, equatorial in Glc and Man) cause significant changes in recognition[72]. Furthermore, a characteristic bidentate hydrogen bond is often formed between two adjacent carbohydrate hydroxyls and the two oxygens of aspartic or glutamic acid residues[73] (see Figure 1.1).

Despite the hydrophilic character of carbohydrates, these interactions through hydrogen bonds are not dominating in carbohydrate–lectin binding. Non-polar (hydrophobic interactions) play also a major role in stabilizing sugar–lectin complexes. The steric arrangement of hydroxyl groups creates hydrophobic patches that can contact with hydrophobic amino acids, such as phenylalanine, tyrosine, or tryptophan, on the protein surface[67]. Moreover, numerous van der Waals interactions contribute significantly to binding. Additionally, divalent cations, such as Ca^{2+} and Mn^{2+}, are involved in carbohydrate recognition either indirectly, by correctly orientating the amino acid chain[74], or by acting as a bridge between the protein and the sugar moieties through the direct interactions with sugar hydroxyl groups.

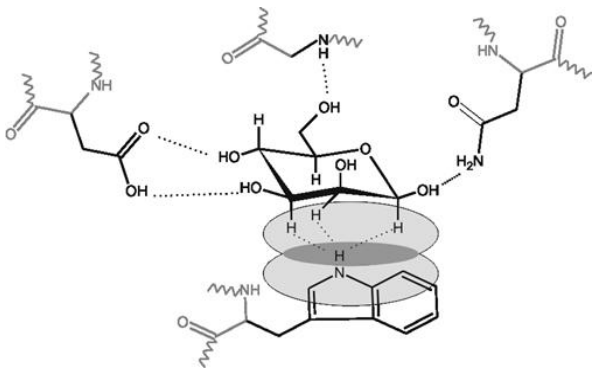


Figure 1.1. Schematic representation of carbohydrate-lectin interactions. Hydrogen bonds between sugar hydroxyl groups and side chains as well as with backbone amide groups are depicted. Hydrophobic interactions (shaded areas) occur between a sugar B-face rich in C-H bonds and aromatic residues of the lectin. Wavy lines depict the rest of the protein backbone.

Even taking into account all the forces involved in carbohydrate-lectin binding, it remains a typically weak interaction. In fact, carbohydrate–protein interactions are much weaker than protein–protein interactions, by about a factor of 10^2-10^3 compared to typical antibody equilibrium dissociation constants (K_D). The K_D value of lectin binding with a simple monosaccharide is usually in the milimolar range. Higher affinities towards di- and trisaccharides (in the micromolar range) are possible through secondary binding sites that fit extra monosaccharides. The reason for this weakness lies in the solvent-exposed nature of the lectin binding sites, which are shallow pockets making few direct contacts with the ligands[75]. However, in biological context, this limitation has been overcome by multivalent interactions[76], *i.e.*, simultaneous contacts between the clustered carbohydrates on cell surfaces and protein receptors that contain multiple carbohydrate recognition domains. This enhancement due to oligo- or polymeric interactions is called "glycoside cluster effect"[77] and invariably involves an increased binding affinity (binding strength and specificity) relative to that of the monomeric ligand. Lectins often have more than one carbohydrate binding site or domain, and it has been shown that the occurrence of two simultaneous binding events can increase the avidity of interaction by more than 100-fold[78], and even as high as 10,000-fold[79].

In summary, carbohydrate-mediated recognition among cells can be envisioned as consisting of several hierarchical levels of molecular information. First, the primary oligosaccharide sequence might be recognized by a protein. Second, several oligosaccharides may be presented and recognized along a polypeptide. Third, multiple glycoconjugate molecules, *e.g.* glycoproteins or glycolipids, may be displayed on the surface; and their density, distribution, and relative orientations may contribute to even greater specificity. Thus, multivalency and cooperative recognition are very important phenomena in glycobiology[80].

1.1.2 **Tools for measuring carbohydrate-mediated biomolecular interactions**

The potential to monitor and control biomolecular interactions has become a new frontier in chemistry and biochemistry, expanding to cellular and molecular biology. For instance, understanding how biomolecules interact is crucial to improving human health, since protein-carbohydrate interactions are also emerging as factors of medical importance in tumour and infection-related research. Thus, from a biological point of

view, identifying the carbohydrate binding entity is as important as deciphering the cognate sugar epitope. However, learning about biomolecular interactions requires sensitive and specific methods; and although numerous tools are available, they have drawbacks and are often limited regarding amount and/or purity of sample needed, sensitivity, cost, and difficult to modify for various uses.

There are many different analytical methods for studying carbohydrate-lectin interactions; typical ones can be classified into three categories: (1) indirect methods such as inhibition of hemagglutination (HIA) and enzyme-linked lectin assay (ELLA), (2) structural methods such as X-ray crystallography and nuclear magnetic resonance (NMR), and (3) other biophysical methods such as isothermal titration calorimetry (ITC), fluorescence spectroscopy, quartz crystal microbalance (QCM) and surface plasmon resonance (SPR).

Historically, carbohydrate-protein interaction affinities have been evaluated by HIA. The test is based on the ability of a soluble hapten to block the ability of a lectin to aggregate blood cells. The concentration of hapten causing 50% inhibition of precipitation is defined as the inhibitory concentration, IC_{50}. As absolute IC_{50} values can vary widely among experiments, HIAs are mainly used to compare binding affinities of different soluble ligands. ELLA, another technique for determine relative binding affinities, monitors the competition between soluble and immobilized carbohydrates for binding to a horseradish peroxidase-conjugated lectin. Similar to some ELISA formats, enzyme-conjugated lectin and soluble ligand are added to microtiter plate wells coated with a polymeric saccharide and, after a washing step, the amount of retained lectin is determined spectrophotometrically to estimate the corresponding IC_{50}. ELLA avoids HIA-associated problems such as aggregation; however, it has disadvantages of its own such as non-specific binding to the microtiter plate, the requirement for lectin-enzyme conjugates, and reduced reproducibility of IC_{50} values due to glycoprotein microheterogeneity. The latter problem can be avoided by covalent immobilization of defined glycoconjugates instead of complex structural glycoproteins[81]. Although HIA and ELLA were very useful in the early stages of lectinology to determine relative carbohydrate affinities[82], they have been gradually replaced by techniques such as microarrays that can measure carbohydrate specificity in a more straightforward manner[83]. However, although microarrays possess desirable features such as nanoscale

and high throughput screening, the need to label at least one of the interacting partners substantially reduces their appeal.

To study carbohydrate-lectin interaction at the molecular and atomic levels, two main biostructural methods are used: X-ray crystallography[84] and NMR[85]. X-ray crystallography involves three steps: obtaining a sufficiently large, pure crystal; subjecting a rotating crystal to X-rays and measuring the intensity of the X-rays diffracted by the electrons in the crystal; and processing and refining the obtained data computationally to yield a final structure for the crystal. Disadvantages of X-ray crystallography include the difficulty in obtaining a sufficient amount of a pure crystal, not to mention the crystallization process alone. NMR overcomes some of these difficulties as it allows structural information about intermolecular interactions to be obtained in solution. Data from NMR studies not only allows the structural determination of the actual entities in the sample, but also their conformational changes and interactions. Also, it allows to determine equilibrium association constants by 1D ^{1}H NMR titration experiments where chemical shift variations induced by increasing amounts of ligand are monitored[86]. Given the intrinsic chemical nature of sugars, in particular their flexibility, it is well established that NMR data should be complemented by computational methods (molecular simulations and modelling) in attempts to unravel unambiguously the structural and conformational features that govern the recognition of saccharides by receptors[87]. Despite the powerful possibilities of NMR, its application for routine analysis is limited by the high cost of equipment and the high expertise required to interpret NMR data. Also, as with X-ray crystallography, NMR has the main disadvantage of requiring substantial amounts of highly pure sample (both lectin and carbohydrate), often difficult to obtain from natural sources.

Thermodynamic parameters of carbohydrate-lectin interactions are usually determined by means of ITC, QCM and SPR, with the latter two techniques providing also kinetic data. ITC is based on the fact that every chemical reaction generates or requires heat, whose changes can be measured with high sensitivity and accuracy by this method. In the binding of a glycan to a lectin, for instance, ITC evaluates the free energy change of the interaction[88]. Most ITC instruments use a continuous power compensation design, where heat evolved from the binding is measured as the voltage necessary to return the cell to its initial temperature. ITC provides information on binding constant, enthalpy and stoichiometry of binding. Its advantages include the ability to monitor interactions

without the need for immobilization, modification or labelling of binding partners. Disadvantages are the limited accuracy with which temperature changes can be determined, and the prerequisite that the interaction has a measurable enthalpic effect. Moreover, in practice ITC poses problems such as high sample amount (> 10 mg), plus solubility problems when low-affinity interactions are studied, as high lectin concentrations are needed for reliable data. Microcalorimetry, a recent improvement, still requires amounts in the micromolar and millimolar range for lectin and ligand, respectively[89]. Fluorescence spectroscopy is also used to evaluate carbohydrate–protein interaction. However, fluorescent derivatives of carbohydrates or proteins require additional labeling steps, and the labels themselves may interfere with the binding process. In addition, as with ITC, data obtained by these techniques usually refer to free, relatively small, glycans (mono-, di-, trisaccharides at most) interacting with a native protein or a modified version of it, *i.e.* situations rather unlike native scenarios. These limitations have prompted the development of biosensor techniques, especially those with label-free detection such as SPR and QCM, for studying carbohydrate-protein interactions.

In recent years, SPR and QCM surface-based analytical techniques have been widely applied for carbohydrate-protein interaction studies[90-93]. Both techniques are based on the immobilization of one molecule on a surface whilst the interacting partner is flown across as a free molecule. Interactions are monitored directly, with no need for labelling, with significantly lower sample requirements and with the added benefit of allowing to determine kinetics parameters[94]. In both techniques binding is detected in real time by means of the difference in mass occurring at the surface upon formation of interacting complexes; detection, however, is based on different physical principles. QCM has a piezoelectric detection mode that measures the frequency shift resulting from the change in mass on the quartz crystal resonator surface. In addition, changes in viscosity are also measured to help analysis. Unlike SPR, QCM is affected by both the water that may be associated with the adsorbed layer and by conformational changes in the adsorbed species. SPR uses an optical phenomenon to detect refractive index changes at a gold surface, which are proportional to the mass change at the very same surface. In SPR, both association and dissociation rate constants can be determined separately, hitherto the only technique that allows so. Thermodynamic parameters, such as enthalpies and entropies, can be also determined by calculating the temperature

dependence of the binding constants[95]. In addition, in SPR, the use of a flexible hydrophilic polymer such as carboxymethylated dextran as surface provides a three-dimensional scaffold that may mimic the extracellular matrix. On the down side, the matrix also enhances phenomena such as mass transport-limited binding, causing low diffusion rates from bulk solution to surface-immobilized ligand, a problem particularly important for large analytes.

1.1.3 **Surface Plasmon Resonance (SPR)**

Surface plasmon resonance monitors biospecific interactions, which occur on a surface of a metal (Au) layer between the immobilized ligand and the free analyte flowing over in solution, by measuring changes in resonance angle due to an increase in analyte mass concentration in the vicinity of the surface. To immobilize biomolecules, the Au layer is coated with a matrix of carboxymethylated dextran (CMD) that provides a three-dimensional hydrophilic matrix, where biomolecules can be immobilized covalently by using different well-characterized chemistries.

Since the first biosensor based on SPR was developed almost two decades ago[96], the application of this technique has increased rapidly, from surface studies to bioanalytical applications[97] (see Table 1.2).

Table 1.2. Application areas of SPR[a].

Biophysical	**Bioanalytical**	**Others**
Biomolecular interaction studies	*Medical diagnostics*	*Surface-enhanced spectroscopies*
· Affinity	· Cancer biomarkers	· Raman
· Kinetic	· Hormones	· Fluorescence
· Thermodynamics	· Pathogens (virus and bacteria)	*Other*
· Specificity	· Allergy markers	· Plasmon energy transfer
Biopharmaceutics and drug discovery	· Single nucleotide polymorphism	· Cancer photodiagnostics
· Cell membrane studies	*Bioanalysis*	· Cancer photothermal therapy
· Antibody development and immunogenicity	· Doping agents	· Photovoltaic cells
· Pathways elucidation	· Drugs	· Light harvesting devices
Proteomics		
· Structure-activity studies		

[a] Table reproduced from [97] with permission from the PCCP Owner Societies.

SPR-based instruments have made it possible to obtain kinetic and thermodynamic data for a large number of interactions, where the interacting molecules may be proteins, peptides, lipids, viruses, nucleic acids, or small organic molecules such as fragments or drug candidates. Although several SPR-based instruments are available, so far Biacore analyzers (GE Healthcare, United Kingdom) are the most widely used[98,99].

In this section a brief explanation of the SPR phenomenon and main applications will be given, as this technique has been widely employed in this thesis to study carbohydrate-protein interactions.

1.1.3.1 *SPR working principle*

SPR instruments use an optical method to measure the refractive index near a sensor surface (within ~ 300 nm) (see Figure 1.2). When a beam of light with certain angle of incidence (θ_{spr}) propagates from a material with a high refractive index (*e.g.*, glass; μ ~ 1.26 - 1.38) into another with a low refractive index (*e.g.*, aqueous buffer; μ ~ 1) the light is refracted towards the interface. At higher angles of incidence, all light is reflected inside the medium of higher refractive index, and the phenomenon of total internal reflection occurs. However, if the surface of the glass is coated with a thin film of metal, usually gold, internal reflection is not total and reflected light interacts with the surface mobile electrons causing their excitation, generating surface plasmon wave-like oscillations (energy transfer/loss) and thereby causing a dip in the intensity of the reflected light. In other words, at a certain incidence angle the momentum of incoming light (photons) is equal to the momentum of the surface plasmons, and the electrons "resonate"; this gave a name to the phenomenon of surface plasmon resonance (Figure 1.2A). This certain angle, at which surface plasmon resonance as well as maximum loss of reflected light intensity occurs is called resonance angle or SPR angle (θ_{spr}). The SPR angle is dependent on the optical characteristics of the system, *e.g.,* on the refractive indices of the media at both sides of the metal. While the refractive index at the prism side is not changing, the refractive index in the vicinity of the metal surface will change, for instance, when accumulated mass adsorb on it. This change in the chemical composition of the medium near to the gold film leads to a shift of the reflected angle (Figure 1.2AB, characteristic shift from angle I to angle II; $\Delta\theta_{spr}$) and the signal is recorded by the optical detector. Therefore, as binding of biomolecules results in an

increase in the refractive index and θ_{spr} is sensitive to changes in refractive index, binding events are detected as a shift in the θ_{spr} (Figure 1.2B).

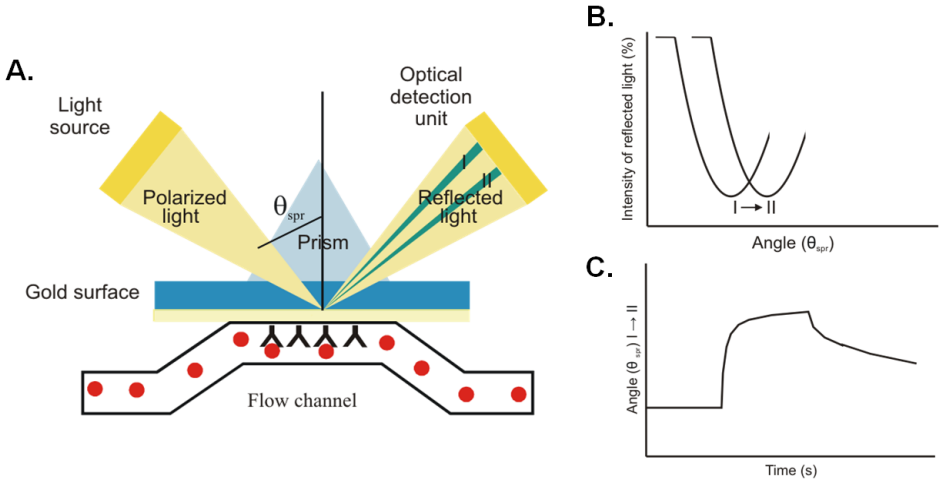


Figure 1.2. (**A**) Schematic representation of the SPR principle with a prism as one medium and a solution of analyte (●) as the other. The ligand (人) is immobilized on the gold surface in contact with the solution. (**B**) Analyte-binding is measured by monitoring changes in the SPR angle (θ_{spr}), at which the intensity of reflected light reaches a dip. (**C**) In a sensorgram, the angle at which the dip is observed is plotted *vs* time.

In SPR-based instruments θ_{spr} is monitored in real time. Therefore, as depicted in Figure 1.2C, the result of an interaction between two biomolecules can be plotted as changes in θ_{spr} in resonance units or response units (RU) *vs* time, a so-called sensorgram. The binding response in RU (1 RU is approximately equivalent to a shift in θ_{spr} of 10^{-4} degrees) depends on the bound analyte, its molecular mass, and to a certain extend the shape. Empirical measurements have shown that binding of 1 pg globular protein to the sensor surface leads to a response of approximately 1 RU.

In a conventional SPR interaction experiment, a target component (analyte) is captured by the capturing molecule (ligand), which is previously immobilized on the sensor surface, giving rise to a measurable signal (direct detection). As shown in Figure 1.3, different phases can be distinguished in the corresponding sensorgram. (1) Each measurement starts with conditioning the sensor surface with a suitable buffer solution. It is relevant to have a stable baseline before the capturing events starts. (2) Then, the free analyte is injected over the sensor surface in an aqueous solution (sample buffer, running buffer) under continuous flow accumulating at the surface upon binding, *i.e.* association process. (3) Next, buffer is injected over the sensor and the non-specifically bound components are flushed off. Also, dissociation of the analyte from the formed biocomplex [L-A] takes place and leads to a decrease of analyte in the vicinity of the

sensor surface and, as a result, in a decrease of the SPR response. (4) Finally, it is important in order to recuperate the surface for the next interaction experiment using a regeneration solution which breaks the specific binding between ligand and analyte.

Appropriate interpretation of the data displayed by a sensorgram, during association and dissociation steps, provides quantitative information on active concentration of molecule in a sample, as well as on the affinity and kinetics of the interaction.

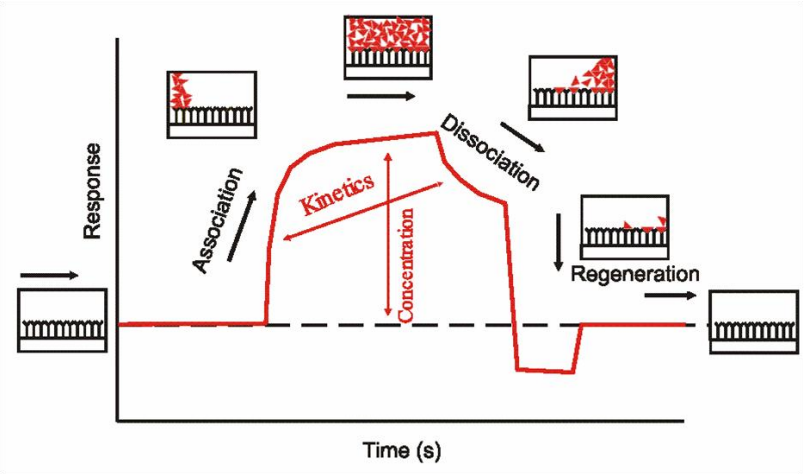


Figure 1.3. A representative sensorgram showing the four steps of an interaction experiment. In the baseline phase, only buffer flows across. In the association phase, the analyte in running buffer is flown over the surface. In the dissociation phase, the running buffer alone is passed over the surface. Finally, a regeneration step is applied to recover the initial surface before running a new cycle. Whereas the shape of the curve gives information about kinetics, the height of the curve quantifies the binding capacity. Reproduced from the BIAcore website (www.biacore.com).

1.1.3.2 *SPR applications*

Real-time measurements with SPR biosensors provide important biomolecular information such as binding affinity, association and dissociation constants and the concentration of the analyte (Figure 1.4).

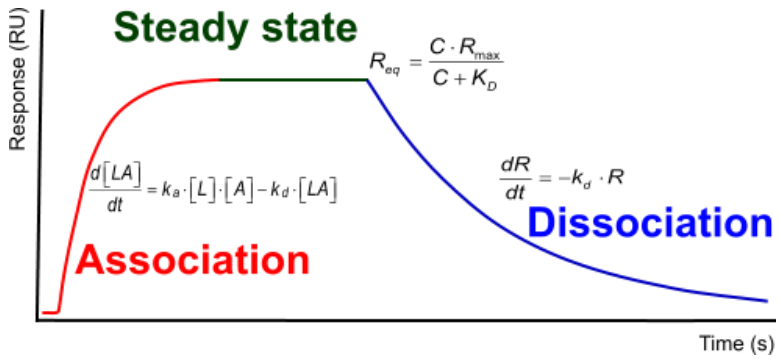


Figure 1.4. Three phases of the interaction kinetics that contain information over the interaction between the molecules. Reproduced from the BIAcore website (www.biacore.com).

The best known benefit of direct detection using SPR is the determination of kinetics of biomolecular interactions. In theory, a simple 1:1 interaction between the analyte (A) and the immobilized ligand (L) is described by:

$$[L] + [A] \underset{k_d}{\overset{k_a}{\rightleftharpoons}} [LA]$$

The reaction between immobilized ligand (L) and an analyte (A) can be assumed to follow a pseudo first order kinetics. During the association phase, the complex [LA] increases as a function of time. When the analyte in the flow over the surface of the chip is replaced by the flow buffer, the free concentration of the analyte suddenly drops to zero and the complex will start to dissociate. The following differential rate equation describes the relation between the parameters:

$$d[LA]/dt = k_a [L] [A] – k_d [LA] \qquad (i)$$

The *association rate constant* (k_a) describes the rate of complex formation, *i.e.* the number of LA complexes formed per second in a one molar solution of L and A. The units of k_a are M^{-1} s^{-1} and are typically between $1 \cdot 10^3$ and $1 \cdot 10^7$ in biological systems.

The *dissociation rate constant* (k_d) describes the stability of the complex, *i.e.* the fraction of complexes that decays per second. The unit of k_d is s^{-1} and is typically between $1 \cdot 10^{-1}$ and $1 \cdot 10^{-6}$ in biological systems.

Equilibrium is reached when the rates of the association and dissociation reactions are equal and equilibrium constants are determined by measuring the concentration of free interactants and the complex at equilibrium:

$$K_A = [LA]/ [L][A] = k_a / k_d \qquad K_D = [L][A] /[LA] = k_d / k_a \qquad (ii)$$

Equilibrium constants are in upper case, *i.e.*, K_A for association (M^{-1}) and K_D for dissociation (M). As can be seen, the *equilibrium association and dissociation constants*, which represent the affinity of an interaction, have a reciprocal relationship with each other ($K_A = 1/K_D$). Thus, a high K_A describes an interaction with a high affinity to associate, whereas a high K_D means low stability of the LA complex. However, it must be taken into account that interactions with the same affinity can result from different kinetics. As an example, a K_D of 10^{-9} M can result either from an interaction with k_a 10^6 $M^{-1}s^{-1}$ and k_d $10^{-2}s^{-1}$ or from another with k_a 10^3 $M^{-1}s^{-1}$ and k_d 10^{-5} s^{-1}. Thus K_A and K_D must be used with care when describing complex formation, as they

describe the system at equilibrium but not the dynamics. For the dynamics, refer to the association and dissociation rate constants.

In SPR, the affinity constant can also be measured directly by steady state binding experiment. Typically, duplicate series of analyte concentrations varied over 2-3 orders of magnitude between $10 \cdot K_D$ to $100 \cdot K_D$ are injected and the level of binding in equilibrium (R_{eq}) is measured. At equilibrium, the concentration of the complex is directly proportional to the response (R_{eq}) and the concentration of the analyte is the concentration injected (C). The concentration free ligand (R_{max} - R_{eq}) can be derived from the complex if the total surface binding capacity (R_{max}) is known. The association equation (i) can be written as:

$$dR/dt = k_a\, C\, (R_{max} - R_{eq}) - k_d\, R_{eq} = 0 \qquad \text{(iii)}$$

Dividing equation (iii) by k_d and introducing the affinity constant $K_A = k_a/k_d$ gives:

$$R_{eq} = (K_A\, C\, R_{max}) / (1 + K_A\, C\, n) \qquad \text{(iv)}$$

where n specifies the number of binding sites per analyte molecule.

and (iv) can be rearranged to:

$$R_{eq}/C = K_A\, R_{max} - K_A\, R_{eq} \qquad \text{(v)}$$

Thus, a plot of R_{eq}/C as y-values *vs* R_{eq} as x-values results in a straight line with a slope of K_A and an y-intercept of $K_A R_{max}$. Therefore, the association constant K_A can be determined by linear regression from transformed data points at equilibrium. It is important to ensure that the analyte injections reach equilibrium. Therefore, an approximate time is calculated (time to reach 90% of equilibrium), when the analyte concentration is equal to the K_D:

$$t_{0.9} \approx 1/k_d \qquad \text{(vi)}$$

Additionally, under saturation conditions, if it is assumed that all the ligand is oriented correctly and 100% active, the stoichometry of the interaction (n) can be calculated by the equation:

$$n = (M_L/M_A)\, (R_{max}/L) \qquad \text{(vii)}$$

where M_A and M_L specify the molecular weight of analyte and ligand respectively; L is the ligand immobilization level; and R_{max} the maximal response obtained experimentally at saturation conditions.

Apart from this linearization method, data from interaction plots of SPR biosensors can be analyzed by the integrated rate method. Although the differential rate equation (i) describes the one-to-one Langmuir reaction accurately, it is not useful as a model to fit to a curve. By integrating the differential rate equation, sensorgrams can be analyzed directly using non-linear methods. This is because integrated rate equations describe the whole curve as opposed to the differential rate equations, which describe the slope of a curve. Therefore, the classical Langmuir surface adsorption can be described with the following integrated rate equation:

$$R_t = R_{eq} (1 - e^{-(k_a \cdot C + k_d)(t - t_0)}) \quad \text{(viii)}$$

Typically, duplicates of a two-fold dilution series of analyte concentration ranging from $8 \cdot K_D$ to $0.25 \cdot K_D$ are injected onto surfaces with different immobilization levels and at different flow rates. Then, SPR experimental data are analyzed by this non-linear method and fitted to different pre-programmed kinetic models in order to determine the rate constants of the interaction by using different commercial available software packages (*e.g,*. BIAevaluation, CLAMP99, Scrubber). Deviations from the ideal bimolecular interaction can be analyzed by the integrated rate analysis, whereas the linearization method is used to visualize possible deviations from the ideal bimolecular interaction case, and to determine the start values of the parameters used in the nonlinear regression analysis of the integrated rate method. Furthermore, this non-linear regression can fit complex models by an iterative process adjusting the variables in the chosen equation to minimize the sum-of squares. The given Chi^2-value shows information about the goodness of fit.

Apart from the affinity and kinetic parameters that describe the strength and speed of an interaction, SPR provides thermodynamic information. The binding free energy described by equation (ix) is experimentally determined according to the Van´t Hoff equation (x).

$$\Delta G^o = \Delta H^o - T\Delta S^o \quad \text{(ix)} \qquad \Delta G^o = - R\, T \ln K_A \quad \text{(x)}$$

ΔH^o and ΔS^o parameters can be calculated by measuring the dissociation constant over a range of temperatures and fitting them to a more rigorous integrated form of the Van´t Hoff equation[100].

1.1.3.3 *Integration of SPR biosensors with mass spectrometry (SPR-MS)*

Although SPR allows the sensitive direct label-free detection and study of biomolecular interactions, mainly between two pure biomolecules, it does not provide any structural data about the interacting species. At the same time, SPR is non-destructive technique, thus the same sample from the binding event can be further characterized by a complementary technique.

In this context, combination of SPR with mass spectrometry offers unique capabilities in terms of not only the quantitative and qualitative analysis of the sample but also of gaining the exhaustive structural information about the SPR-captured analyte through the measurement of its molecular weight and fragmentation patterns[101]. The fact that the quantities that can be bound to the sensor chip are of the same order of magnitude (*e.g.*, 100 fmol) as those typically needed for MS analysis suggests that coupling of these two analytical techniques is possible[102]. In tandem SPR-MS approaches, molecules are initially captured from the solutions by ligands covalently immobilized on the SPR sensor surface followed by their further analysis by MS. MALDI-TOF MS was the first MS format to be coupled with SPR, both because of its high sensitivity and relative tolerance for contaminants such as salts, detergents and buffer components.

Different strategies for integrating the sensor chip with MS instruments are possible. The first one is based on collecting the analyte eluted from the recovered (regenerated) SPR sensor chip and directing it into mass spectrometer. This is convenient when the SPR sensor chip is integrated into a flow system (*e.g.,* Biacore instruments). However, it is time consuming, with sample losses, and quantitative elution may be difficult in the case of high-affinity binding. In order to achieve more flexibility and sensitivity, Sönksen *et al.* developed a sandwich micro-recovery method to elute subpicomole levels of affinity-bound molecules from the sensor surface in few microliters for subsequently analysis by MALDI-TOF MS[103]. Briefly, after cleaning the whole SPR system to remove any non-specifically bound proteins, the elution step was carried out between two air bubbles that have two main functions: (1) recording the recovery event in the sensorgram, and (2) allowing a switch to a volatile elution buffer (*e.g.*, ammonium bicarbonate). Recently, as a result from the collaboration between Biacore and Bruker Daltonics, SPR-MS functionality for the BIAcore 3000 instrument has been developed in order to automate and increase the capacity of this process. Three special commands (MS_INJECT, MS_WASH and MS_RECOVER) were created in order to

minimize the contamination of the recovered material with components of the running buffer, by washing of the fluidic system after sample injection (MS_WASH) and recovery the bound analyte in a small volume between injections of air bubbles (MS_RECOVER)[104]. To elute the specifically bound analyte in the smallest possible volume, 2 µL of recovery solution is placed over the flow cells and left in contact with the surface for 30 sec. The solution is then returned back to the injection needle and deposited either into a vial or directly on a MALDI target. Throughout the course of such experiment, SPR detection is used to monitor and quantify the capture and recovery of the analyte. Figure 1.5 shows a typical micro-recovery sensorgram obtained after the sample injection (A), wash of the fluidic system (B) and recovery of the bound material (C). In most examples, the experimental system was based on well characterized, high-affinity analyte-antibody interactions, or used samples at higher concentrations than those of native biological fluids. In addition, it must be taken into account that using this "sandwich elution" there is inevitable protein loss during post-elution steps that decreases sensitivity that can be critical for subsequent MS measurements, especially noticeable when protein concentrations are low. However, micro-recovery procedures can be repeated because the elution procedure is not destructive to the sensor chip, thus increasing sample amount. Another clear advantage of micro-recovery is that it allows subsequent handling of the sample for identification. Thus, subsequent tryptic digestion followed by peptide mass mapping by MALDI-TOF MS or sequencing by ESI-MS/MS is possible.

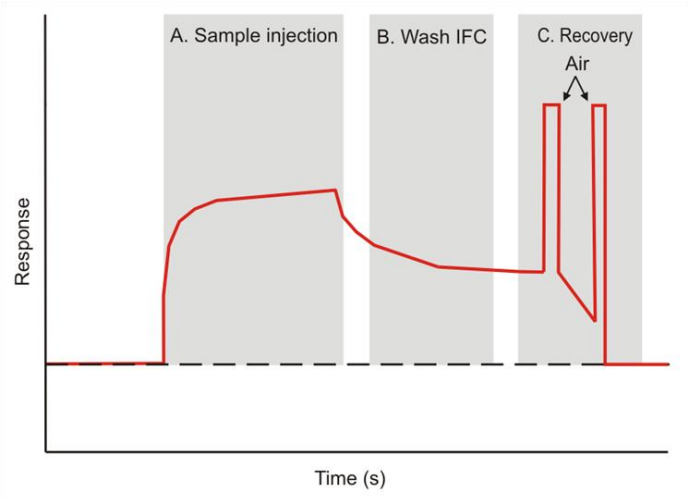


Figure 1.5. Example of a micro-recovery sensorgram where bound material is eluted for subsequent MS analysis.

Another strategy is based on MALDI MS being directly performed on the SPR chip after capturing the analyte of interest. After the analyte has bound to the sensing surface the SPR slide is inserted directly in an appropriate MALDI-MS plate holder. Finally MALDI matrix is dropped on each localised spot and mass spectrometry measurements are performed. Bellon and co-workers[105] developed a hyphenated SPR-MALDI technique for the detection of glycoproteins β-lactoglobulin and ovalbumin directly from a biochip containing immobilized antibodies. Femtomole amounts of specifically bound analytes determined by SPR were sufficient to obtain good quality mass spectra. This strategy has two main advantages with respect to the former one. First, the array format of the sensor chip allows carrying out the fast screening of multiple interactions, in contrast to conventional flow SPR, *e.g.*, Biacore instruments, which allow monitoring only 1-4 interactions at a time. Secondly, the sensor chip can be easily transferred to MALDI support, allowing both direct SPR and MALDI-MS measurement from the same surface, which minimizes sample contamination and sample loss. However, this approach suffers also from two main disadvantages, namely the need for an appropriately configured mass spectrometer, and cost, as sensor chips are typically used just once. In addition, MALDI-TOF analysis of an entire protein is not a realistic method of protein identification, given the resolution of current instruments, as well as the fact that post-translational modifications alter the molecular mass rendering straightforward identification very difficult.

For unequivocal identification of an SPR-retained protein molecule, a further step of peptide mass fingerprint analysis is necessary. In this context, Natsume and Nelson developed an elegant approach that integrates proteolytic digestion. After specific capture of the protein on the first flow cell, it can either be digested on-chip by a proteolytic enzyme delivered to the same flow cell[106], or be routed into another, protease-containing flow cell where digestion occurs[107]. Both accurate molecular masses of the native protein and the proteolytic fragments provide information for the rigorous protein identification.

The different approaches of the coupling of SPR and MS show that this hyphenation has the potential to address many different types of biological questions, allowing quantitative and qualitative binding assessments in a variety of biological samples. SPR-MS makes it possible to link protein detection, capture, and kinetic analysis with the measurement of protein and peptide masses and protein identification.

1.1.3.4 *Glycopeptide probes for carbohydrate-lectin SPR interactions studies*

Although SPR-based technology was initially designed for the study of protein-protein interactions, its field of application has been extended to other biomolecular interactions (*i.e.* peptide-protein[108], DNA-protein[109], lipid-protein[110], sugar-protein[111] and sugar-sugar) in recent times. Carbohydrate-protein interaction studies by SPR have been addressed in several ways, with either the carbohydrate or the protein being immobilized. The initial choice was to immobilize the lectin moiety, for two obvious reasons: (1) immobilization chemistry already well-established for proteins (2) immobilizing the larger molecule avoids potential mass transport problems[79].

Protein immobilization can be carried out either covalently through amine or thiol groups (Figure 1.6AB), or indirectly through capturing systems (*e.g.*, biotinylated lectins on streptavidin-activated surfaces) (Figure 1.6C). Generally, the non-covalent immobilization is preferred in front of classical covalent chemistry because it creates a more homogenous surface and preserves ligand native structure, although the coupling of biotin to the protein is a random process[112]. This immobilized lectin approach has allowed reliable monitoring of numerous binding events, but only complex glycans[113] or glycoprotein[114] analytes have been successfully monitored by SPR. Interactions with small carbohydrates (mono- or disaccharides; < 500 Da) are difficult to quantify because of their low MW and refractive index, requiring high levels of immobilized lectins that are both difficult to obtain and may cause mass transport limitations in kinetic measurements. Other aspects to consider when working with immobilized lectins are non-specific binding due to their large sizes as well as their overlapping substrate specificities, which complicate relative quantification.

The alternative approach, carbohydrate immobilization, is chemically demanding and not universally applicable. Direct attachment of unmodified carbohydrates via aldehyde chemistry (possible in all reducing sugars), with spontaneous reaction between the aldehyde group and, *e.g.*, a hydrazide on the chip surface[115] (Figure 1.6D), is successful only with large polysaccharides[116] but not with small sugars (mono- di- and trisaccharides). Thus, the approach is mainly advantageous for complex carbohydrates from natural sources (*e.g.*, glycoproteins) which unfortunately are often unavailable in the large amounts required.

Sugar epitopes have also been immobilized as glycopeptides derived from proteolysis of native glycoproteins via the amino groups of the peptide moiety[78]. However, the macro- and microheterogenity found in glycoproteins often makes unequivocal interpretation of the results unviable, and purifying large amounts of well-defined glycoproteins from natural sources remains a nearly impossible task. For these reasons, well-characterized, chemically synthesized glycoconjugates are in several ways the best alternative to prepare sensor surfaces of precisely defined glycosylation. We have recently reviewed in detail current developments in glycoconjugate synthesis. This review is included in this dissertation as an Appendix.

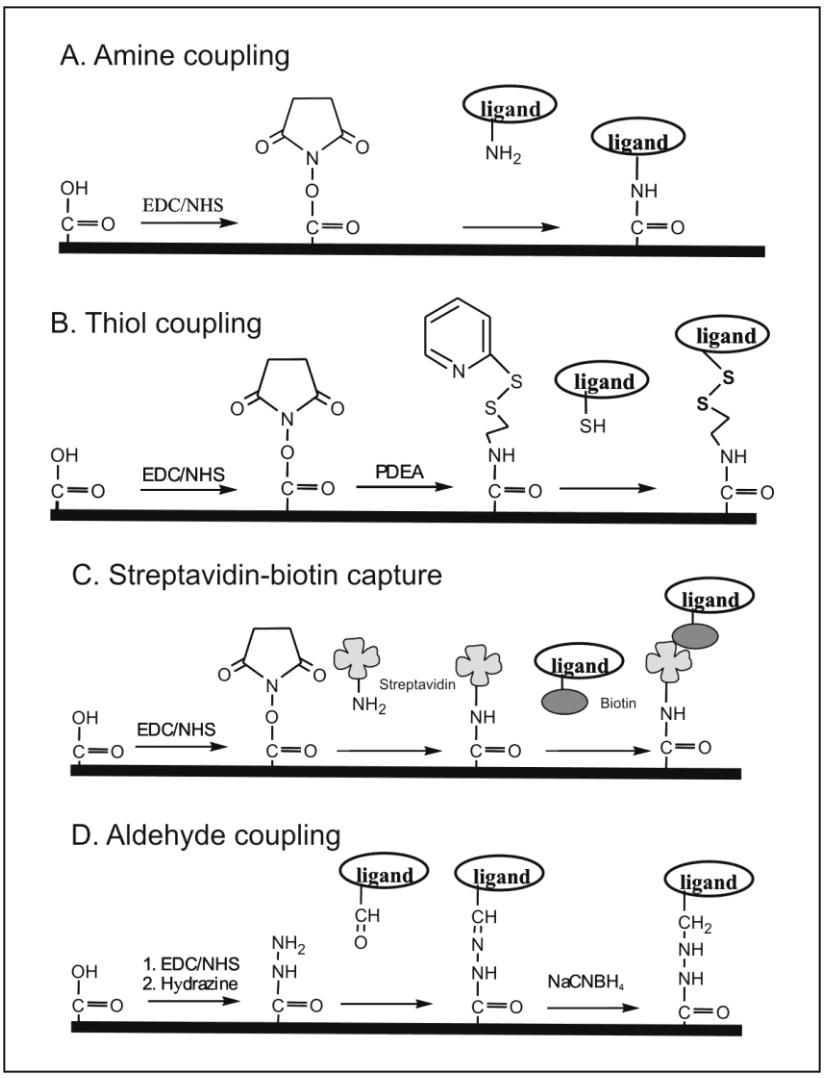


Figure 1.6. Common SPR-immobilization techniques. (**A**, **B**) The amine (-NH_2) and thiol (-SH) coupling chemistries for immobilizing covalently proteins to a carboxymethylated dextran coated surface are well established procedures. (**C**) Streptavidin-biotin capture is a typical capturing approach to immobilize non-covalently proteins. (**D**) For immobilization of oxidized glycoproteins, the aldehyde coupling chemistry is an alternative.

Carbohydrates can be conjugated to different kinds of molecules, including whole proteins such as human serum albumin[117], biotin[118], lipids[119], peptides[120] or non-natural compounds[121], and this non-glycosidic part may be used to improve the sugar immobilization and disposition onto the surface. A number of coupling methods have been developed to link carbohydrate to aglycons in order to facilitate immobilization. Reductive amination was one of the first methods to be implemented in the preparation of glycoconjugates (Figure 1.7A). Under suitable conditions glycans with a free reducing end can be covalently attached to protein amine groups (usually lysine side chains) forming an iminium ion that can be reduced to an amine. However, this condensation reaction is slow and inefficient, requiring large excesses of sugar and long reaction times. In addition, reductive amination leads to the opening of the reducing (carbonyl) end sugar ring, rendering it different from the natural counterpart, a fact that may have consequences for its recognition.

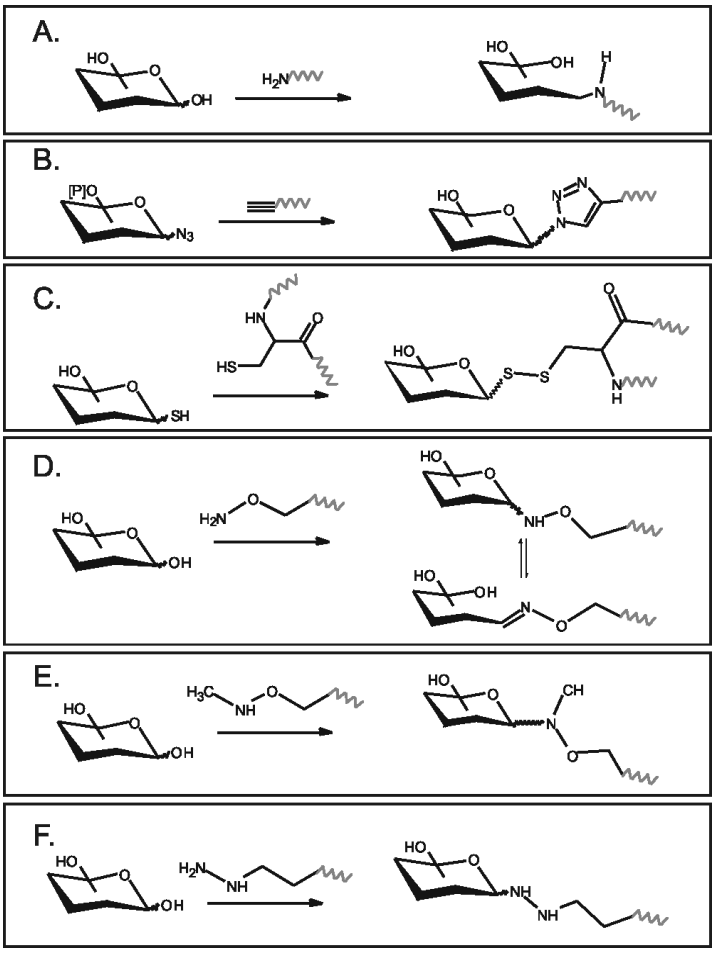


Figure 1.7. Examples of chemoselective ligation reactions used in the synthesis of glycoconjugates. (**A**) Reductive amination. (**B**) Huisgen 1,3-dipolar cycloaddition. (**C**) Thiol-based ligation. (**D,E,F**) Carbonyl-based ligation with: (**D**) aminooxy, (**E**) *N*-methyl-aminooxy and (**F**) hydrazide group.

Recently, a growing number of ligation reactions with primary application in synthetic organic chemistry have been successfully extended to biological molecules. Thus, new approaches taking advantage of chemoselective ligation reactions[122] such as the Huisgen 1,3-dipolar cycloaddition of azides and alkynes (Figure 1.7B), commonly known as click chemistry[123]; or the reaction of thiol groups with a variety of electrophiles (Figure 1.7C), or the reaction of carbonyl groups (ketones or aldehydes) with strong nucleophiles (Figure 1.7DEF), have been applied to glycoprobe elaboration. In the latter group, the so-called alpha effect renders the nitrogen atom of the strong non-natural nucleophile (*e.g.*, hydroxylamine, hydrazide, thisemicarbazide) less basic but more nucleophilic than that of a primary amine and thus favors the condensation. Under mildly acidic conditions (pH 4-5.5) these highly reactive nucleophiles are deprotonated while other potentially competing amino, guanidine or imidazole groups are protonated and therefore non-nucleophilic.

Amongst carbonyl-based chemoselective ligations, conjugation of unmodified carbohydrates with hydroxylamines has been widely used for the synthesis of glycopeptides. In contrast to reductive amination, in oxime-linked glycopeptides the peptide-bound sugar is in equilibrium between closed and open ring forms (Figure 1.7D). To overcome this limitation and drive the equilibrium exclusively towards the closed ring form, the use of *N*-methyl-hydroxylamine instead of regular hydroxylamine has been proposed[124], and succeeded in retaining the linked sugar in cylic pyranose form (Figure 1.7E). Another approach to maintaining the cyclic form of the attached sugar is through condensation of unmodified carbohydrates with a hydrazide group (Figure 1.7F). NMR analysis of hydrazine-linked glycoconjugates showed that the cyclic β-anomeric products were generated predominantly[125]. However, the lower stability at low pH and the presence of *syn*- and *anti*-isomers in hydrazine-linked conjugates still limit the broad utilization of these procedures[126].

All these methodologies have advantages such as being highly efficient and chemoselective under aqueous mild conditions, and allowing conjugation of unprotected glycans. Moreover, the immobilization process is chemically well defined so that a homogenous surface may be prepared. However, in some cases, the required derivatization of the glycan component plus the use of non-native structures in close proximity to the glycan may limit applicability and arguably affect the interaction. As reported elsewhere, although lectins and other carbohydrate-binding proteins are highly

specific to their sugar epitopes, the aglycon, particularly its conformation, may play an important role as well[127].

Several other issues such as oligosaccharide density, spacing, and orientation achieved upon immobilization, as well as the nature, flexibility and length of the linker should not be overlooked when working with immobilized glycoconjugates. Moreover, since many lectins achieve their specificity and affinity through multivalent interactions with glycans (*i.e.* the cluster effect), glyco-based surfaces should aim at faithful representation of a multivalent sugar display, and at capturing the physiological avidity of such interactions in as native-like fashion as possible.

1.1.4 **Mass spectrometry (MS)**

Arguably one of the more powerful and versatile analytical techniques for all sorts of compounds, including carbohydrates, MS has become the cutting-edge technology for glycomics, linking mass with composition and providing precise characterization of complex structures. A wide range of MS equipment are available for glycan analysis, as well as a number of methods (see section 1.1.2) for identifying and quantifying carbohydrate-protein interactions *in vitro*, each with particular strengths and weaknesses. During the last two decades, MS and its associated techniques have proven its value, not only for the study of covalent complexes, where MS is well established, but also –and significantly for protein-ligand interactions– for non-covalent assemblies. In particular, ESI-MS[128] has been recently used by protein chemists to study non-covalent protein interactions, since the mild ionization process allows detection of intact, weakly bound complexes without causing molecular fragmentation. The strength of this technique lies in the ability to directly probe binding stoichiometry, to measure multiple binding equilibria simultaneously, and, when combined with gas phase techniques such as collision-induced dissociation or ion-mobility separation, to interrogate the composition and structure of the complexes. Similarly, ESI-MS has been applied to a variety of carbohydrate-protein interactions[129,130] as well as other protein interactions (*e.g.*, protein-cofactor, enzyme-substrate and protein-DNA), and the measured affinities generally agree well with values measured using other established binding assays, such as ITC. Furthermore, the CaR-ESI-MS approach has been shown to be very effective for screening carbohydrate libraries for specific interactions with target proteins[131].

1.1.4.1 *Carbohydrate-Recognition-Domain-EXcision-MS (CREDEX-MS)*

Alternative MS-based approaches have also been used to evaluate carbohydrate–protein binding interactions. Frontal affinity chromatography coupled online to ESI-MS (FAC-MS)[132] allows high-throughput screening of compound mixtures for specific protein interactions and determination of dissociation constants. The method involves continuous infusion of a sample with potential ligands through a microcolumn wherein the protein target (an antibody, lectin, or enzyme) is immobilized. The ligands, detected ESI-MS, are eluted according to their binding affinities for the target protein, thereby allowing the relative affinities to be easily established.

Carbohydrates immobilized to solid surfaces have been also employed in formats slightly different from SPR and with various other objectives, such as affinity-based systems to map the interacting domains of sugar-lectin complexes (*i.e.* carbohydrate recognition domains (CRD)). Though not providing the level of structural detail of NMR and X-ray crystallography, these approaches, originally developed for antigen-antibody interactions, have the advantages of high sensitivity, rapid analysis time and low sample consumption. The general method, which provides molecular epitope characterization by combining limited proteolysis and MS[133], is based on the principle that formation of an antigen-antibody complex shields some amino acid, particularly those in regions involved in interaction, from protease attack. Based on this hypothesis, two different strategies have been developed: (1) an epitope *excision* approach that involves (i) antibody immobilization, (ii) affinity binding of the antigen followed by limited proteolytic digestion of the immune complex, (iii) elution and MS analysis of affinity-bound peptides (Figure 1.8A), and (2) an epitope *extraction* approach where the antigen first undergoes enzymatic digestion, then digested fragments are captured by binding to the antibody (Figure 1.8B). Several epitopes have been successfully characterized by means of both approaches[134,135].

Starting from these successful results, this epitope excision/extraction methodology was extended to carbohydrate-protein interactions. For instance, a novel glycan-affinity chromatography method combining proteolytic digestion of protein-glycan complexes and mass spectrometry called CREDEX-MS (Carbohydrate REcognition Domain EXcision Mass Spectrometry)[136,137] has proven useful in the structural definition of the CRDs of both human galectin-1 and galectin-3 with lactose as the immobilized binding

partner[138]. The affinity-bound peptides, eluted and identified by MS, were in perfect agreement with the crystal structure for galectin-3 in complex with lacNAc[139].

In contrast to antigen-antibody, in carbohydrate-protein interactions only one interacting partner, the lectin, can be proteolytically digested, so the obvious experimental format in this case is to have the carbohydrate moiety (epitope) immobilized and pass through the lectin in solution. Among several chemistries for covalent immobilization of sugars onto surfaces[140,141] (see 1.1.3.4), attachment to divinylsulfone (DVS)-activated sepharose is the preferred one for short sugar epitopes, because it appears to maintain the glycan unit next to the divinylsulfone group mainly in its native closed-ring form, while non-anomeric sugar hydroxyl groups are randomly linked to the divinylsulfone-activated surface[142]. Consecutive enzymatic reactions on the bound epitope can be performed to further fine-tune the epitope mapping process[143]

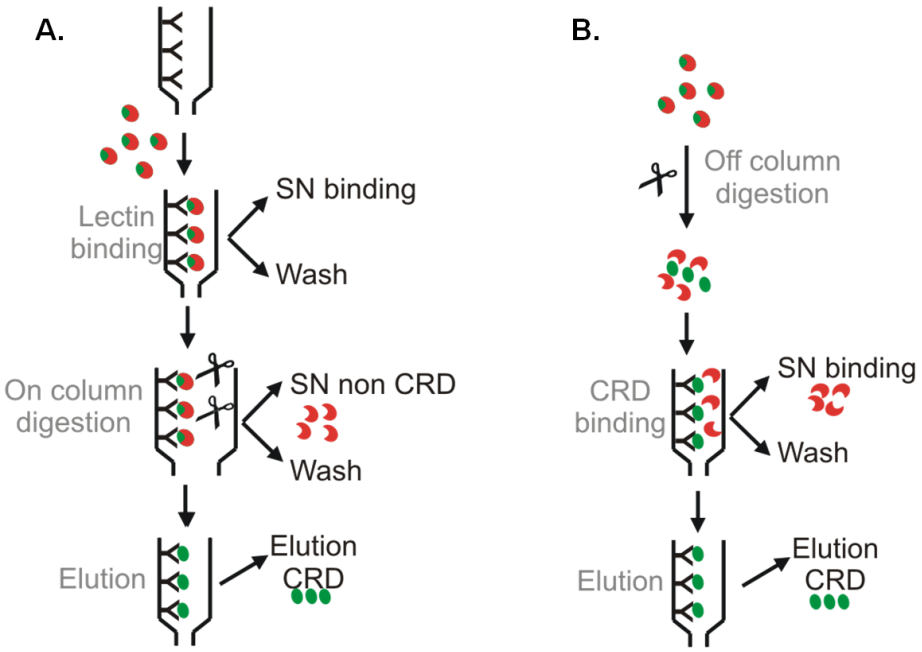


Figure 1.8. Epitope analysis methods based on limited proteolysis and MS. (**A**) Excision experiment. (**B**) Extraction experiment.

1.2 OBJECTIVES

In previous work of our group, an optimized synthesis of aminooxy-peptide glycoprobe precursors for surface-based sugar-protein interaction studies was developed. With this original work as a starting point, the first goal in this chapter was to further expand this approach to key glycan epitopes.

To that end, a meticulous study of galactose-binding using *Erythrina Crista-galli* agglutinin (ECA) as a lectin model was completed using both SPR and MS. The suitability of SPR for carbohydrate-protein interaction studies has been amply discussed before, as well as, the CREDEX-MS approach to identify the interacting domains of sugar-lectin complexes. Combined use of these complementary tools provided detailed information on carbohydrate-lectin interactions, and is presented here as a proof of principle.

This main objective had been further elaborated as follows:

- Preparation of SPR glycoprobes: Aoa-peptide synthesis, conjugation to different mono-, di-, and trisaccharides, and MS characterization of glycopeptide end products.
- SPR kinetic, affinity and thermodynamic studies of interactions between well-known lectins and different sugar epitopes.
- Interfacing SPR with MS for lectin capture and characterization.
- Preparation of glycan affinity columns.
- NMR characterization of the sugar-DVS linkage to be used in the CREDEX methodology.
- Characterization of lectin CRDs by CREDEX-MS.

1.3 RESULTS AND DISCUSSION

1.3.1 **Glycoprobe synthesis**

As already mentioned, a reliable method for immobilizing glycan displaying probes on SPR chips for sugar-protein interaction studies was developed in our group[93]. In this approach, the sugar moiety was immobilized through a tailor-made peptide module on the sensor surface and the interacting lectin was passed through. In this way, the sugar epitope is immobilized in a well-defined manner and since the aglycon is a peptide, it can in principle be readily adapted if a sequential or conformational feature needs to be reproduced. Oxime ligation[144] was used to attach the unmodified glycan via the aminooxy functionality (Aoa) incorporated at the N-terminus of the peptide module, Aoa-GFKKG-amide[145], whereas the methylated version, *N*[Me]-O-Aoa-GFKKG-amide, ensured correct exposure of the carbohydrate on the chip surface as well as the conformational integrity (*e.g.,* as a pyranose ring) of the monosaccharide unit proximal to the surface. Given our interest in glycoprobes that display small sugar epitopes (*e.g.* mono- to trisaccharide) for lectin recognition, this NMR-confirmed well-defined (native-like, *i.e.*, closed ring structure) conformation of the sugar unit linked to the peptide was of crucial importance[92]. Finally, the resulting glycopeptide probes were attached to the dextran matrix through the peptide module through classical amine coupling chemistry. As shown in Figure 1.9, the first step in amine coupling is to activate the carboxymethyl groups of SPR surface with *N*-hydroxysuccinimide (NHS), creating a highly reactive succinimide ester which reacts with primary amines on peptides and proteins (usually Lys side chains). After immobilization, the remaining activated carboxymethyl groups are blocked by injecting a high concentration of ethanolamine.

The length and GFKKG sequence of the peptide module in the glycoprobe used to immobilize sugar epitopes onto the sensor surface was designed and optimized with several criteria in mind: (1) Two mandatory amino acids, the N-terminal Aoa for chemoselective sugar ligation and two Lys residues near the C-terminus to ensure both proper guiding to the negatively charged sensor surface and efficient covalent (amide bond) binding to the activated carboxyl group on the SPR chip surface were required. (2) In addition, an aromatic Phe residue was included to increase hydrophobicity and facilitate purification by HPLC and quantification by UV-spectrophotometry, while (3)

Gly residues were added to provide flexibility and distance between the dextran surface and the glycan epitope. Thus, at the end, the peptide moiety of our neoglycoconjugates was large enough (MW > 700) to provide a substantial mass enhancement effect on the refractive index near the dextran surface compared to that of a simple sugar epitope, thereby facilitating the monitoring of the immobilization.

With this original work and optimized glycoprobe design as a starting point, the first goal of this thesis was to demonstrate the usefulness of the approach through the analysis of a well-known lectin-sugar interaction. Specifically, the interaction between the β-galactose-specific legume lectin *Erythrina cristagalli* agglutinin (ECA) and a series of related β-galactosides [Gal(β1-4)GlcNAc, Gal(β1-4)Glc, Gal(β1-3)GlcNAc, Gal(β1-6)GlcNAc)] was studied by both SPR and MS, and the carbohydrate-binding site of the lectin was identified by CREDEX-MS.

For this purpose, all the glycopeptides probes mentioned were prepared as detailed in materials and methods section. Briefly, *N*[Me]-O-Aoa-GFKKG-amide peptide was synthesized by Fmoc-based solid phase synthesis and directly conjugated to several galactose-containing disaccharides utilizing 25 mM sugar and 20 mM peptide in 0.1M sodium acetate, pH 4.6 at 37ºC for 72 h. Remarkably, for *N*Ac-disaccharides, the conjugation yield could be improved by acidifying the reaction mixture to pH 3.5. The resulting disaccharide-*N*[Me]-O-GFKKG-amide glycopeptide conjugates were purified by HPLC and characterized by MALDI-TOF MS.

1.3.2 Interaction studies using Surface Plasmon Resonance (SPR)

A further step in characterizing the carbohydrate specificity of ECA with our SPR platform required a sensor chip where the three above disaccharides, only differing in the type of glycosidic linkage, were compared. To this end, conjugates of the *N*[Me]-Aoa-peptide with *N*-acetyl-lactosamine (Gal-β1,4-GlcNAc, type II glycans), lacto-*N*-biose (Gal-β1,3-GlcNAc, type I glycans) and β1,6 galactosyl-*N*-acetyl glucosamine (Gal-β1,6-GlcNAc) were immobilized. In addition, a glycoprobe with lactose (Gal-β1,4-Glc) was also tested, to evaluate the importance of the *N*-acetyl group at position C2. A non-glycan ligand, namely unconjugated *N*[Me]-Aoa-peptide, was also immobilized as reference surface.

A first concern of this study was efficient ligand immobilization (Figure 1.9). For this purpose, a pre-concentration of the ligand (in our case, the glycopeptide probes) close to

the activated surface is necessary and will be favoured by a low ionic strength and a pH of the solution lower than the pI of the ligand. At this pH, the ligand has enough positive charge to be concentrated on the surface, but still preserves some neutral amine groups capable of reacting with the activated NHS-esters. In our case, considering that the theoretical pI of our glycopeptides was ~ 10, significantly higher than most proteins, a pH 6.0 buffer was used satisfactorily for ligand immobilization. In fact, immobilizations at lower pH such as pH 4.0 carried out in previous work of our group caused embedding of the glycoprobe within the dextran matrix and poor lectin recognition.

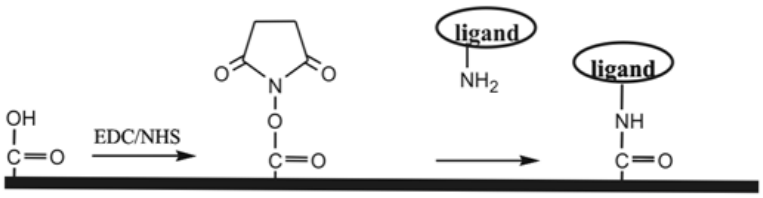


Figure 1.9. Scheme of a ligand immobilization to the dextran surface sensor chip using the amine coupling procedure.

An equally important goal of immobilization chemistry is obtaining appropriate levels at all four flow cells. For kinetic measurements, low ligand density is desirable in order to avoid mass transport effects or steric hindrance but still giving a good response. For this purpose an accurate estimation of glycopeptide concentration is of utmost importance. To this end, small glycopeptide amounts (*i.e.* 100-200 µg) were quantified by UV at 258 nm (λ_{max} of a Phe residue[146]) in a nanodrop system. An experimental extinction coefficient of ε=0.1438 mM^{-1} cm^{-1} for the non-conjugated *N*[Me]-O-Aoa-GFKKG-amide peptide was in this way obtained and used for UV quantification of all glycoprobes before immobilization.

Finally, immobilization levels of ~ 500 RU and ~ 120 RU were obtained for reference and glycoprobes, respectively. ECA lectin, extensively studied in terms of carbohydrate binding, was flown at different concentrations over the sensor chip surfaces.

1.3.2.1 *SPR-based kinetic and affinity analysis*

The medium-throughput screening capacity of our Biacore 3000 instrument allowed simultaneous analysis of ECA binding to the four β-galactoside containing epitopes: Gal(β1-4)GlcNAc, Gal(β1-4)Glc, Gal(β1-3)GlcNAc, and Gal(β1-6)GlcNAc. Kinetic

rate constants (k_a, k_d) for both association and dissociation phases, and the derived affinity constants (K_A) could be determined for the first three disaccharides (Figure 1.10, Table 1.3); for Gal(β1-6)GlcNAc the response was too low for reliable quantitative data to be derived. In addition to binding parameters, SPR results provided helpful structural insights into the binding events, particularly about the functional groups involved in each interaction. Thus, binding to ECA was strongly influenced by the type of glycosidic linkage as well as by the nature of the monosaccharide at the reducing end, with a clear preference for Gal(β1-4)GlcNAc, in agreement with previous studies[147].

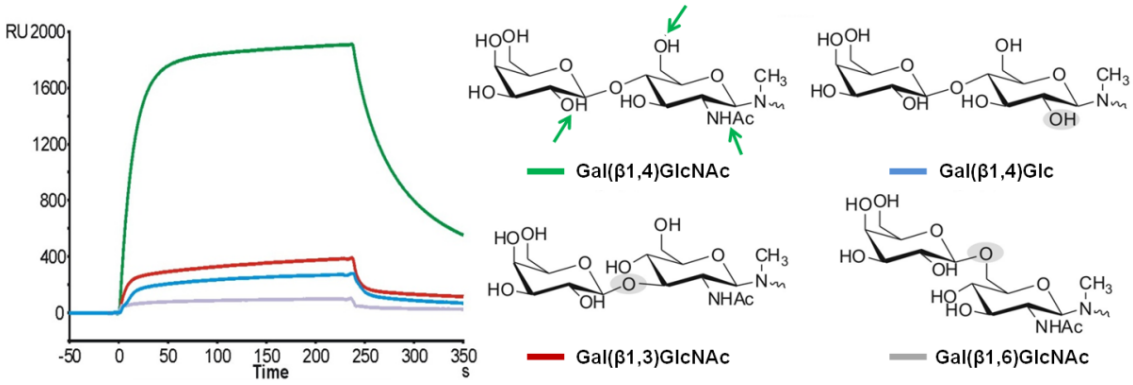


Figure 1.10. SPR analysis of ECA binding to four β-galactosides. Glycopeptide probes displaying the epitopes were coupled to the sensor surface at similar immobilization levels. The sensorgrams show the differential curves after subtracting a reference channel with no epitope immobilized. Gal(β1-4)GlcNAc functional groups crucially involved in ECA interaction are marked with an arrow. In the other disaccharide structures, functional groups whose modification causes loss of ECA affinity are shaded.

Table 1.3. Kinetic rate constants (k_a, k_d) and the derived affinity constant (K_A) of ECA to the four different glycoprobes exposing terminal β-galactosyl-disaccharides.

Glycoprobes	$k_a(M^{-1}s^{-1})$	$k_d(s^{-1})$	K_A (M^{-1})
Gal(β1-4)GlcNAc	4.7×10^4	5.3×10^{-3}	8.9×10^6
Gal(β1-4)Glc	3.5×10^3	5.8×10^{-3}	6.0×10^5
Gal(β1-3)GlcNAc	2.3×10^3	3.2×10^{-3}	7.2×10^5
Gal(β1-6)GlcNAc [a]	-	-	-

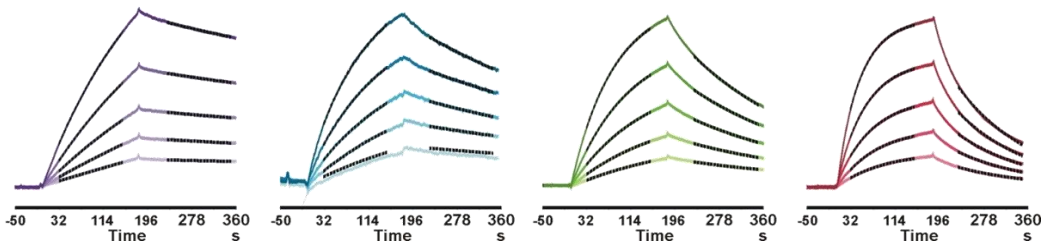
[a] The binding response obtained for the Gal(β1-6)GlcNAc was not enough to determine affinity constants.

For the most favourable Gal(β1-4)GlcNAc epitope, the 15-fold higher affinity over Gal(β1-4)Glc underlines the significant role of the *N*-acetyl group at position C2 in lectin binding. Also, by comparing the responses of the three *N*-acetyl-disaccharides,

the relative role of the hydroxyls of the non-terminal sugar in the interaction can be ascertained. Thus, the decreased binding of the β1-3 isomer (8% relative to the β1-4), or the even lower affinity of the β1-6-linked disaccharide suggest that both C6 and C3 hydroxyls are significantly involved in the canonic binding of Gal(β1-4)GlcNAc to ECA, so that when either of these hydroxyls is obliged to engage in glycosidic bond formation impaired affinity ensues. In summary, straightforward inspection of SPR data highlights a key role of the *N*-acetyl group and, to a lesser extent, of the C3 and C6 hydroxyls, in sugar-ECA recognition, in good agreement with X-ray data showing the O3, N2, O6, O_{NAc} atoms to be directly involved in the interaction[148].

1.3.2.2 *SPR-derived thermodynamic parameters*

In addition to the kinetic and structural information discussed above, thermodynamic parameters for the preferential Gal(β1-4)GlcNAc *i.e.* (lacNAc)-ECA interaction could also be determined in real time by monitoring the SPR response at various temperatures. Figure 1.11 shows ECA binding profiles in the 10 to 25 ºC range, and how temperature rise affected (*i.e.*, accelerated) both association and dissociation steps.



T (ºC)	10	15	20	25
$k_a(M^{-1}s^{-1})$	0.59×10^4	0.99×10^4	1.46×10^4	2.82×10^4
$k_d(s^{-1})$	0.68×10^{-3}	1.75×10^{-3}	3.50×10^{-3}	7.13×10^{-3}
$K_A (M^{-1})$	8.68×10^6	5.66×10^6	4.17×10^6	3.96×10^6

Figure 1.11. Effect of temperature on the interaction between ECA and Gal-β1,4-GlcNAc. Sensorgrams obtained for five concetrations of ECA (from 100 to 700 nM) and derived kinetic data at five different temperatures are shown. At each temperature, colored lines represent experimental data and black lines the local fitting to a 1:1 Langmuir model. The values of kinetic contants were calculated by the mean of the five concentrations measured by duplicate.

Both k_a and k_d rate constants were determined at each temperature by locally fitting the sensorgrams to the five concentrations used. As values shown, ECA underwent approximately 5- and 10-fold increases in association and dissociation rates with temperature, respectively, which in turn caused a 2-fold decrease in the K_A. These

derived equilibrium association constants (K_A) at each temperature were then used to calculate thermodynamic parameters by means of the Van't Hoff equation:

$$\ln K_A = -\Delta H°/RT + \Delta S°/R \quad \text{where } R = 8.314 \text{ J K}^{-1} \text{ mol}^{-1}$$

Alternatively, thermodynamic parameters could also be determined from each rate constant (k_a, k_d), independently, by means of the Eyring equation:

$$\ln (k/T) = -\Delta H/RT + -\Delta S/R + \ln(k'/h)$$

where k is the appropriate rate constant (association or dissociation), and k´ and h are the Boltzmann ($k' = 1.380 \times 10^{-23}$ J K^{-1}) and the Planck constants ($h = 6.626 \times 10^{-34}$ J s), respectively.

Both Van't Hoff and Eyring plots (Figure 1.12) could be fitted to a linear model. As depicted on Table 1.4, ΔH values derived from each analysis (-41.9 and -42.3 kJ mol^{-1}, respectively) were in good agreement with the -45.6 and -54.5 kJ mol^{-1} values from previous reported ITC and NMR studies, respectively[149,150].

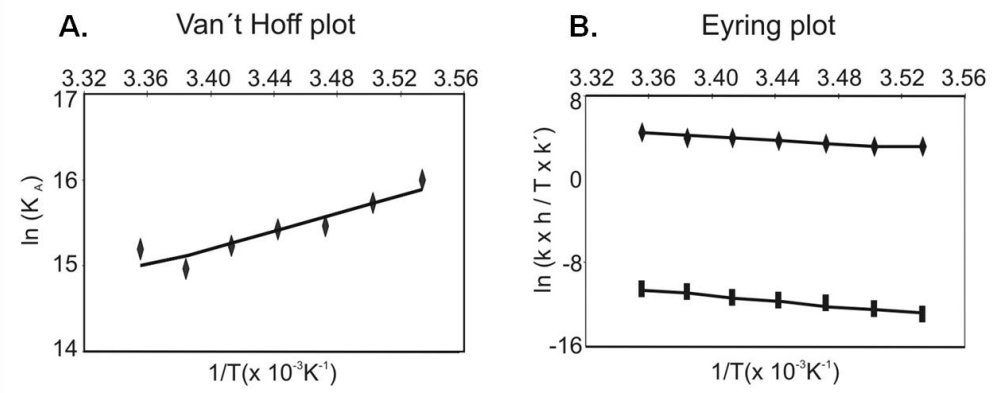


Figure 1.12. (**A**) Van't Hoff plot for the binding of ECA to lacNAc. Dots showing affinity constants determined at 2.5 ºC intervals in the 10-25 ºC range could be linearly fitted to derive ΔH values. (**B**) Eyring plots for the same interaction. Dots and squares correspond to association and dissociation rate constants, respectively; both series could be linearly fitted to derive ΔH values.

Table 1.4. Thermodynamic parameters for ECA-lacNAc interaction determined using SPR, ITC and NMR. In the SPR block the thermodynamic parameters were derived from the Van´t Hoff or Eyring equations.

	SPR		ITC	NMR
	Van´t Hoff	Eyring		
ΔH (kJ mol^{-1})	-41.9	-42.3	-45.6	-54.5
ΔS (J mol^{-1} K^{-1})	-14.2	-15.7	-75.3	-102
K_A (M^{-1})	3.9×10^6	4.0×10^6	1.1×10^4	6.0×10^3

In all these technologies, entropic values were calculated from the Gibbs free energy equation ($\Delta G^{o}=\Delta H^{\circ}-T\Delta S^{\circ} = - R T \ln K_A$) employing the equilibrium association constant and experimental enthalpy as recommended[100]. The SPR-derived equilibrium constant for lacNAc ($K_A \sim 10^6 M^{-1}$ at 25ºC, Table 1.4) was about two orders of magnitude higher than previously reported ITC- and NMR-derived values for the same disaccharide in free form ($K_A \sim 10^4 M^{-1}$, Table 1.4). Similar differences between surface- and solution-based methods had been already observed for other carbohydrate-lectin interactions[119,151], probably because lectin multivalency in addition to the rather high glycan surface density required for SPR experiments may increase the apparent affinity through secondary interactions. In any event, such differences in equilibrium constants, compounded with the slightly different ΔH values from SPR and ITC/NMR methods, can explain the discrepancy in entropic values found in Table 1.4.

1.3.3 **Characterization by Mass Spectrometry (MS)**

1.3.3.1 *Interfacing SPR with MS for lectin capture and identification*

Although the primary application of the SPR-based technology is to characterize biomolecular interactions through the direct monitoring of binding and the determination of quantitative kinetic data, recently SPR-based instruments have been also used as an affinity capture/purification platform to capture interacting proteins from complex mixtures for subsequent identification by mass spectrometry. This additional SPR-based application was tested in our BIAcore 3000 instrument through a standard ligand-capture experiment for the interaction of ECA with Gal(β1-4)GlcNAc, the glycoprobe with the highest affinity (Figure 1.13).

Briefly, after injecting the analyte onto the glycoprobe-immobilized surface, the flow cell and the fluidic system were automatically washed with a MS-compatible buffer (*e.g.* 25 mM NH_4HCO_3) in order to clean the system and avoid carry-over of unreacted protein material to the recovered elution. After washing, the surface-captured lectin, a 55 kDa dimer, gave a 790 RU readout (*i.e.*, an estimated 790 pg of surface-bound material). Then, the captured lectin was specifically eluted with excess lacNAc disaccharide through a recovery function, named MS recover. This command results in the collection of bound lectin in very small volume (2 µL), separated by air bubbles to avoid sample diffusion or cross-contamination and returning the captured protein automatically into a specific vial or even a MALDI target. Meanwhile during the

capturing experiment, the SPR signal was used to monitor and quantify both the capture and recovery of the bound interacting partner (Figure 1.13A). In this particular case, although the fast dissociation of ECA (k_d=5.3 × 10^{-3} s^{-1}) made recovery and subsequent analysis more complicated, especially at room temperature and even injecting higher lectin concentrations; concentration by vacuum centrifugation furnished enough material (~15 fmol) for molecular weight determination (Figure 1.13B). One partial solution to this problem, considering the high dissociation rate of this lectin, was to perform the experiment at a lower temperature. Indeed, at 10ºC, both association and dissociation rates were diminished (Figure 1.11). However, the slower dissociation required higher sugar concentration for the elution step, and this in turn complicated observation of the released lectin by MALDI-TOF MS.

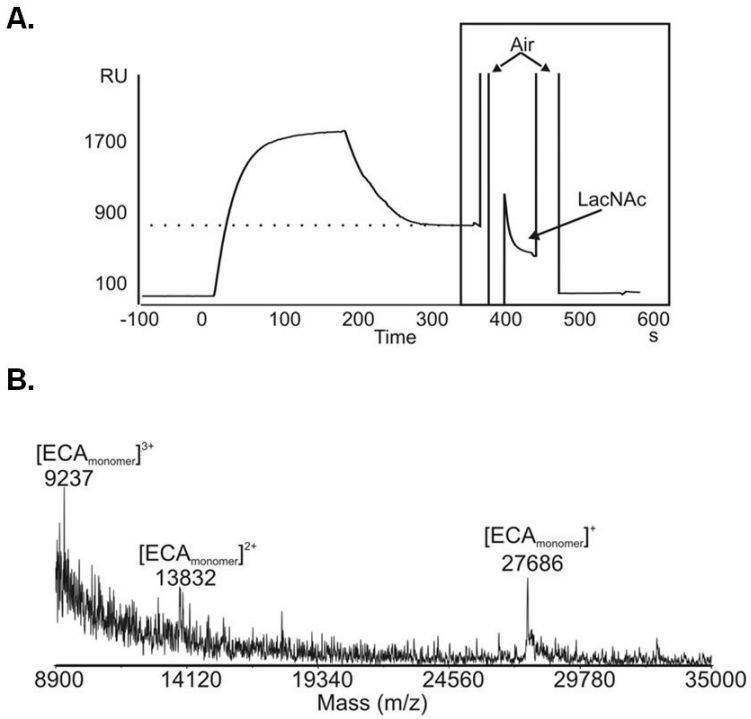


Figure 1.13. (**A**) SPR sensorgram of the recovery experiment over lacNAc-glycoprobe. ECA at 1 μM was passed on the glycosylated surface and the captured material (~15 fmol) was recovered with 2 μL lacNAc. (**B**) MALDI-TOF MS spectrum of the recovered protein.

In summary, the combination of SPR and MALDI-TOF MS molecular weight determination allows successful characterization of sugar-lectin interactions, provided the dissociation rate of the complex, the desorption ability of the lectin and the ionization capacity of the lectin are favourable enough.

1.3.3.2 *Carbohydrate-binding site determination by CREDEX-MS*

A thorough study of the carbohydrate-lectin interaction mechanism must ideally include determination of the carbohydrate recognition domain (CRD) of the lectin. As mentioned in introduction, although X-ray crystallography and NMR are primarily used for structural studies of carbohydrate-lectin complexes, the sample amounts required in these techniques are not always easy to obtain. In addition, practically all current approaches to protein characterization rely on the combined use of limited proteolysis and MS methods. However, our SPR platform does not allow direct, on-chip tryptic digestion, because the free Lys residue in the lectin-capturing glycan-*N*[Me]-O-Aoa-GFKKG module (*i.e.*, the one not used for anchoring the glycoprobe to the chip surface) is trypsin-susceptible and cleavage at this site would cause the affinity-bound lectin material to be lost. Thus, in order to cover this limitation, the CRD determination was performed by off-line proteolysis in a divinylsulfone-based carbohydrate affinity column, using the CREDEX-MS approach (section 1.1.4.1). This novel methodology based on limited proteolysis and mass spectrometry, can become a viable alternative for elucidating carbohydrate-lectin interactions.

First, given our interest in glycoprobes that display sugar epitopes for lectin recognition, a well-defined structural and conformational characterization of the sugar-DVS linkage in CREDEX methodology was of crucial importance. Thus, a simplified model of the sugar immobilization step was performed in solution instead of the solid phase role reaction used in the standard CREDEX protocol. In this way, desired products were easily isolated by HPLC and analyzed by MALDI-TOF, as shown in Figure 1.14. MVS, an activating agent as DVS but monofuncionalized, and the sugars lac (lactose) and lacNAc (*N*-acetyllactosamine) as non-acetylated and acetylated disaccharides models were chosen for the experiment.

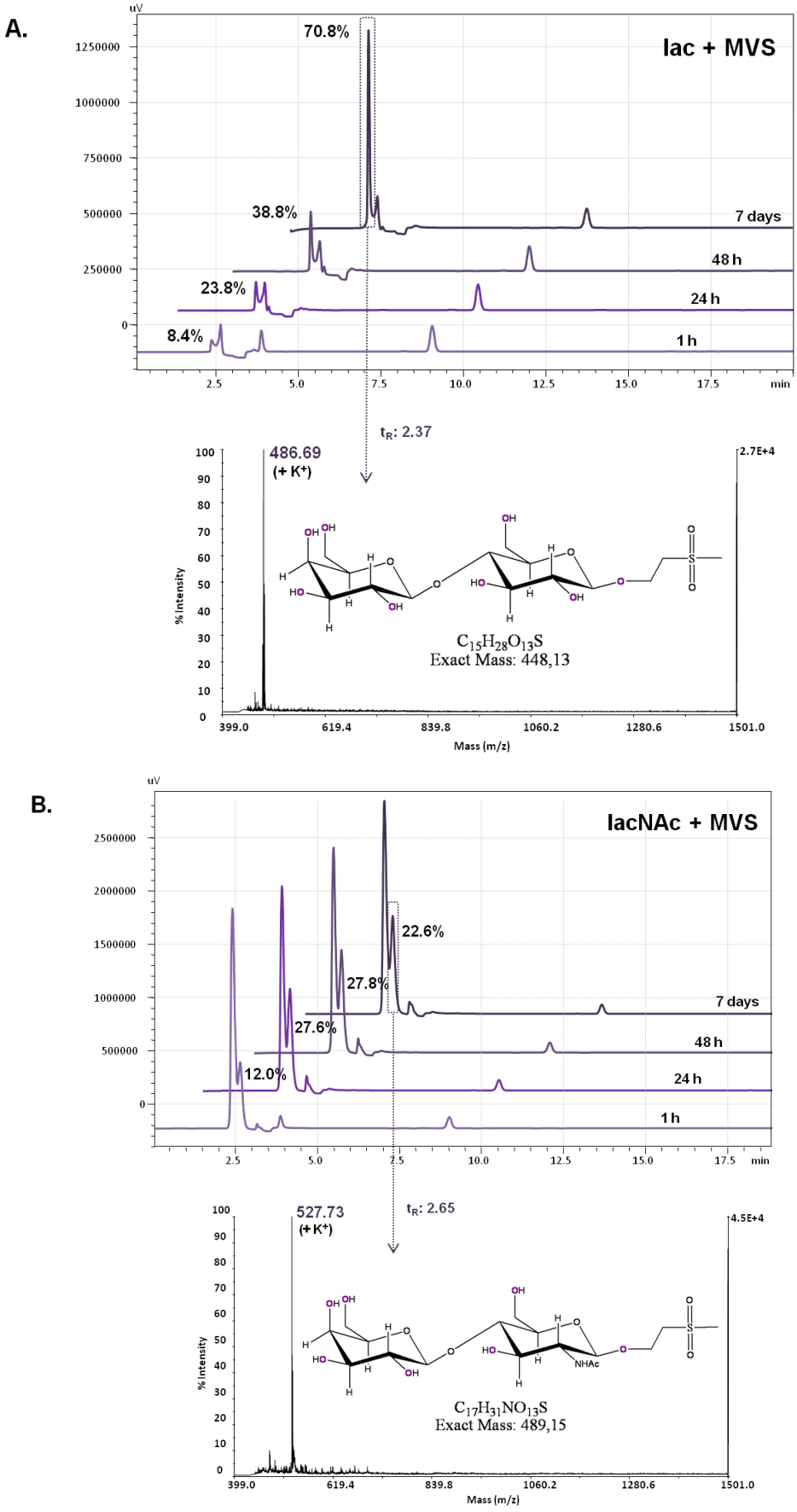


Figure 1.14. HPLC time course and MALDI-TOF MS of purified end product of (**A**) lac + MVS, (**B**) lacNAc + MVS reactions.

Then, ^{1}H-NMR experiments of the products obtained in both trials were conducted and compared with those obtained with commercial sugars that were not functionalized with MVS, in order to observe the induced changes. One of the relevant questions, namely if binding between the sugar moiety and DVS involved the anomeric carbon or an open/closed form, could be answered with these NMR experiments. Experimental ^{1}H-NMR spectra are shown in Figure 1.15 and summarized on Table 1.5.

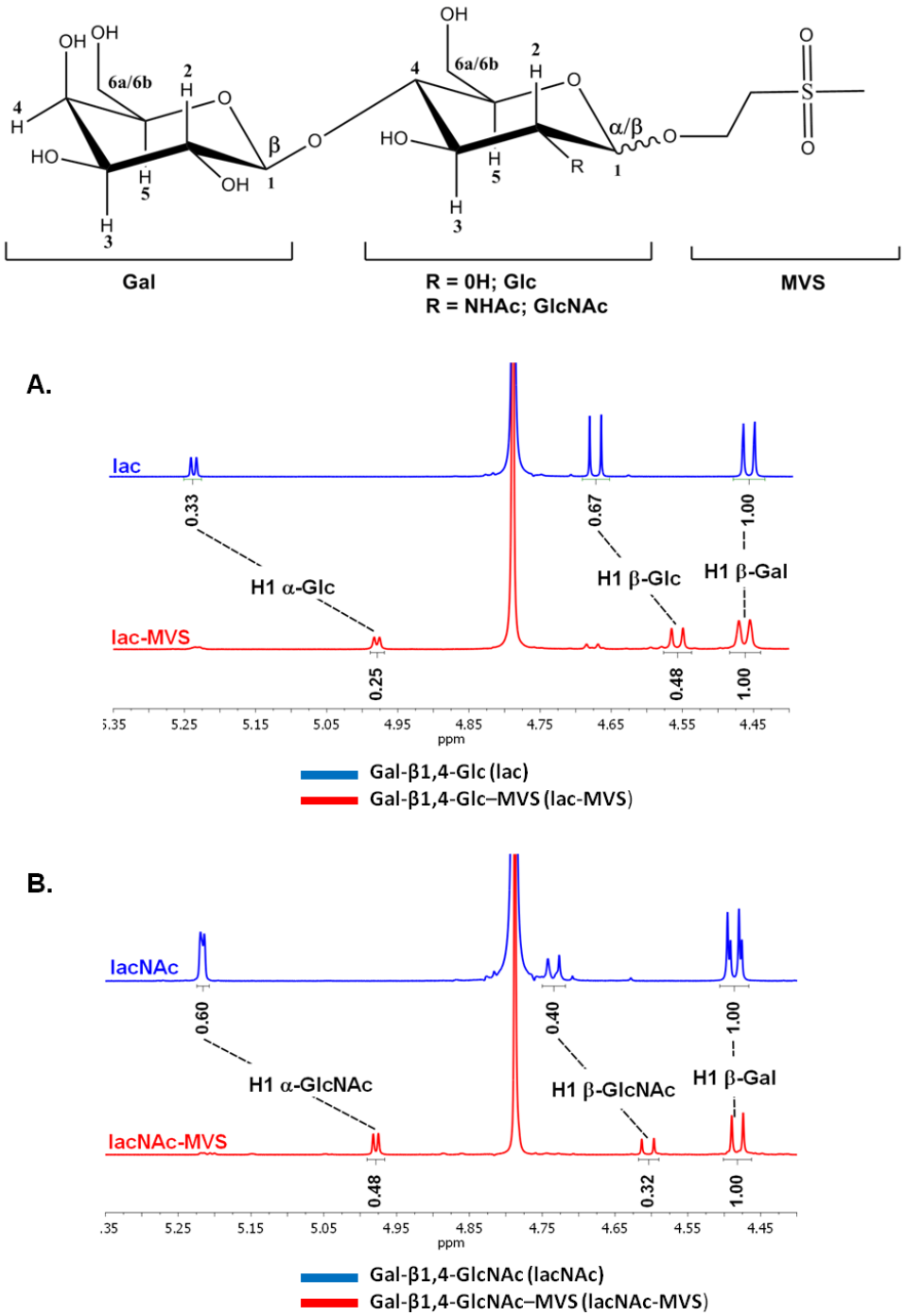


Figure 1.15. Expanded anomeric (H1) proton region in the ^{1}H-NMR spectra of: (**A**) lac and lac-MVS, (**B**) lacNAc and lacNAc-MVS. H1 signal integrations indicated under each peak.

Table 1.5. ^{1}H-NMR chemical shifts* of the most relevant exchangeable protons in lac *vs* lac-MVS (top) and lacNAc *vs* lacNAc-MVS (bottom).
* relative to the signal of TMSP (δ 0 ppm) in D_2O, 500 MHz, at 25°C

	lac		lac-MVS		
Proton	δ(ppm)	$^{3}J_{1,2}$(Hz)	δ(ppm)	$^{3}J_{1,2}$(Hz)	Δδ(ppm)[a]
Glc 1α	5.23	3.80	4.97	3.76	-0.26
Glc 1β	4.67	7.99	4.55	7.99	-0.12
Gal 1β	4.46	7.82	4.46	7.80	0.00
	lacNAc		**lacNAc-MVS**		
Proton	δ(ppm)	$^{3}J_{1,2}$(Hz)	δ(ppm)	$^{3}J_{1,2}$(Hz)	Δδ(ppm)[a]
Glc 1α	5.22	2.56	4.98	3.58	-0.24
Glc 1β	4.73	7.75	4.61	8.36	-0.12
Gal 1β	4.49	7.81	4.48	7.82	-0.01

[a] Chemical shift differences relative to lac (top) or lacNAc (bottom).

The key features of the spectra were the three signals in the 4.40-5.40 ppm region, where the anomeric protons of both Glc and Gal residues appear. Since the α and β-anomeric protons of the Glc residues are easily differentiated by their characteristic *J*-coupling constants (Glc: $^{3}J_{1\alpha,2}$ ~3.6 Hz; $^{3}J_{1\beta,2}$ ~7.8 Hz), the equilibrium α:β ratios can be readily established by signal integration. Although the presence of an *N*-acetyl group instead of a hydroxyl at position 2 of the glucopyranose ring affected only (slightly) the chemical shift of the H1 proton in the β anomer, it did substantially alter the α:β ratio, from 0.5 to 1.5 for lac and lacNAc, respectively (Figure 1.15). On the other hand, Gal had the β-configuration fixed in the disaccharide.

Replacement of an OH by an O-alkyl group in the reaction with MVS induced predictable changes (more than -0.20 ppm upfield, Table 1.5) in the chemical shifts of both α and β H1 protons of Glc (dotted lines, Figure 1.15), with practically no effect on the α:β anomer ratios. At any rate, the NMR-analysis suggested that the linkage between the sugar and MVS (a reasonable extrapolation to the DVS used in the CREDEX methodology) involves the OH at the anomeric carbon of the Glc residue, hence leaving unaffected the Gal epitope responsible for carbohydrate-lectin interaction. MVS reaction on a different hydroxyl group was not observed and di-substitution was negligible.

With these evidences, CREDEX-MS technology was validated as a working tool for the study of several sugar-protein interactions and was subsequently used to determine CRD of the complex ECA-lacNAc, hence testing the complementarity of CREDEX-MS with the SPR-based glycoprobe approach.

The ECA lectin is a 55 kDa protein, active as a dimer and N-glycosylated at both Asn17 and Asn113[152]. From its crystal structure in complex with lactose, residues Leu86, Ala88, Asp89, Tyr106, Gly107, Phe131, Asn133, Ala218 and Gln219 have been identified as directly interacting with the sugar moiety(Figure 1.16)[153].

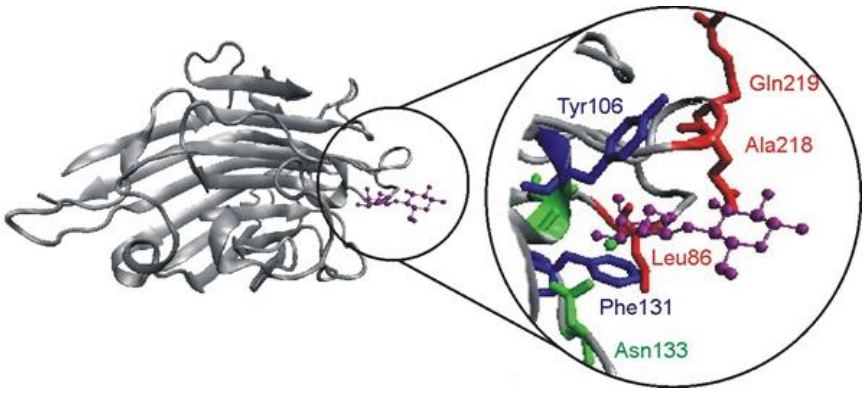


Figure 1.16. Ribbon representation of the crystal structure of ECA in complex with lactose (PDB 1GZ9). The CRD is expanded at right. Key residues in the interaction are colored: those involved in weaker hydrogen bonds are shown in red, those in strong hydrogen bonds in green, and those in hydrophobic interactions in blue. Lactose is shown in purple.

In this type of interactions, the CRD is generally made up of amino acid residues close in space though not necessarily in sequence. Thus, optimal digestion conditions to achieve highest sequence coverage are mandatory. Among three standard proteases, trypsin [1:20 (w:w) ratio] was chosen because it gave 86 % coverage (13 peptides, Figure 1.17) *vs* 30% and 26% for chymotrypsin and Glu-C, respectively.

In order to identify the carbohydrate-binding site by CREDEX-MS, the sugar epitope *N*-acetyllactosamine was immobilized onto DVS-activated sepharose (DVS-S) using conventional coupling chemistry in a microcolumn. Then, an excision experiment (section 1.1.4.1, Figure 1.8) combining interaction, proteolysis and MS was performed with ECA lectin. Briefly, ECA was incubated with the lacNAc-DVS-S for 24 h, the column was washed until no protein was observed by MALDI-TOF MS and then the lectin-sugar complex was digested with trypsin at 37 ºC overnight.

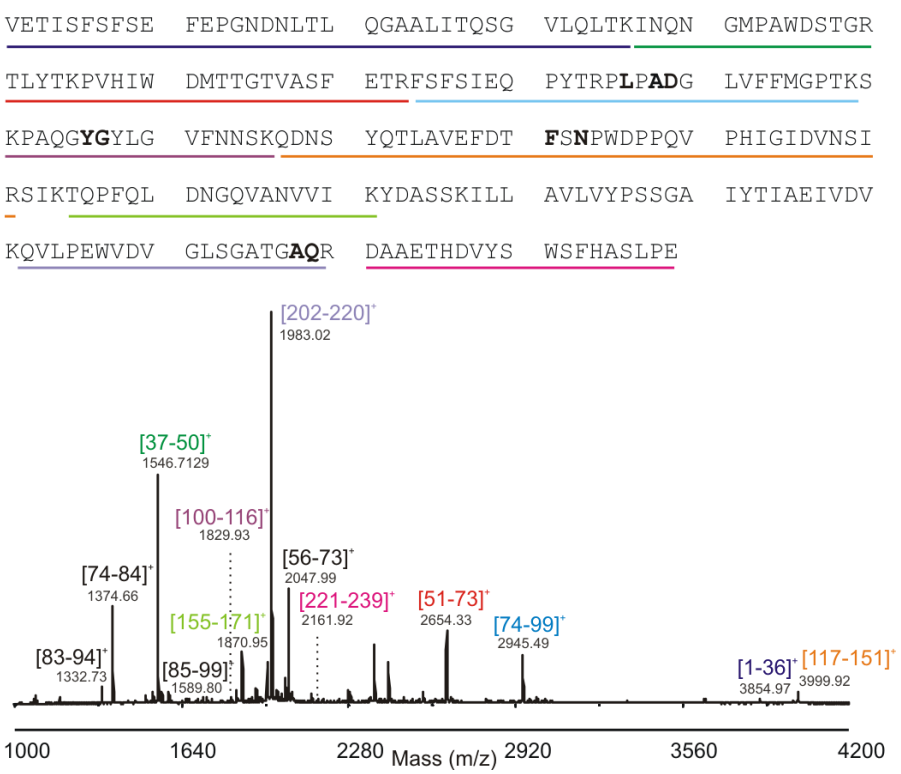


Figure 1.17. Peptide mass fingerprint of ECA (digestion in solution). Above, primary sequence of ECA[153] where amino acid residues reported to participate directly in carbohydrate-lectin interactions are in bold.

Comparison of the standard (Figure 1.17) with the flow-through from the on-column digest fraction of the excision experiment (Figure 1.18A) revealed essentially identical mass fingerprint peaks with only minor differences in intensity. The column was next washed until no peptide signals were observed (Figure 1.18B) and then the glycan-interacting peptides were eluted with 60 % acetonitrile in water and analyzed (Figure 1.18C). In this fraction, several peptides {[74-99], [75-94], [83-94], [100-116], [117-151] and [202-220]} (Figure 1.18C) that contain amino acids displaying direct contact with the carbohydrate in the X-ray structure of the lac-ECA complex (Figure 1.19A) could be unequivocally assigned to the binding site. In addition to these specific peptides, three peptides found in low abundance {[37-50], [51-73] and [74-84]} showed no contacts with the sugar in the X-ray structure (Figure 1.19B), hence no relation to the binding site. Their presence in the elution fraction was explained as a case of "riding" (via β–strand interaction) with the spatially close, glycan-binding peptide [202-220].

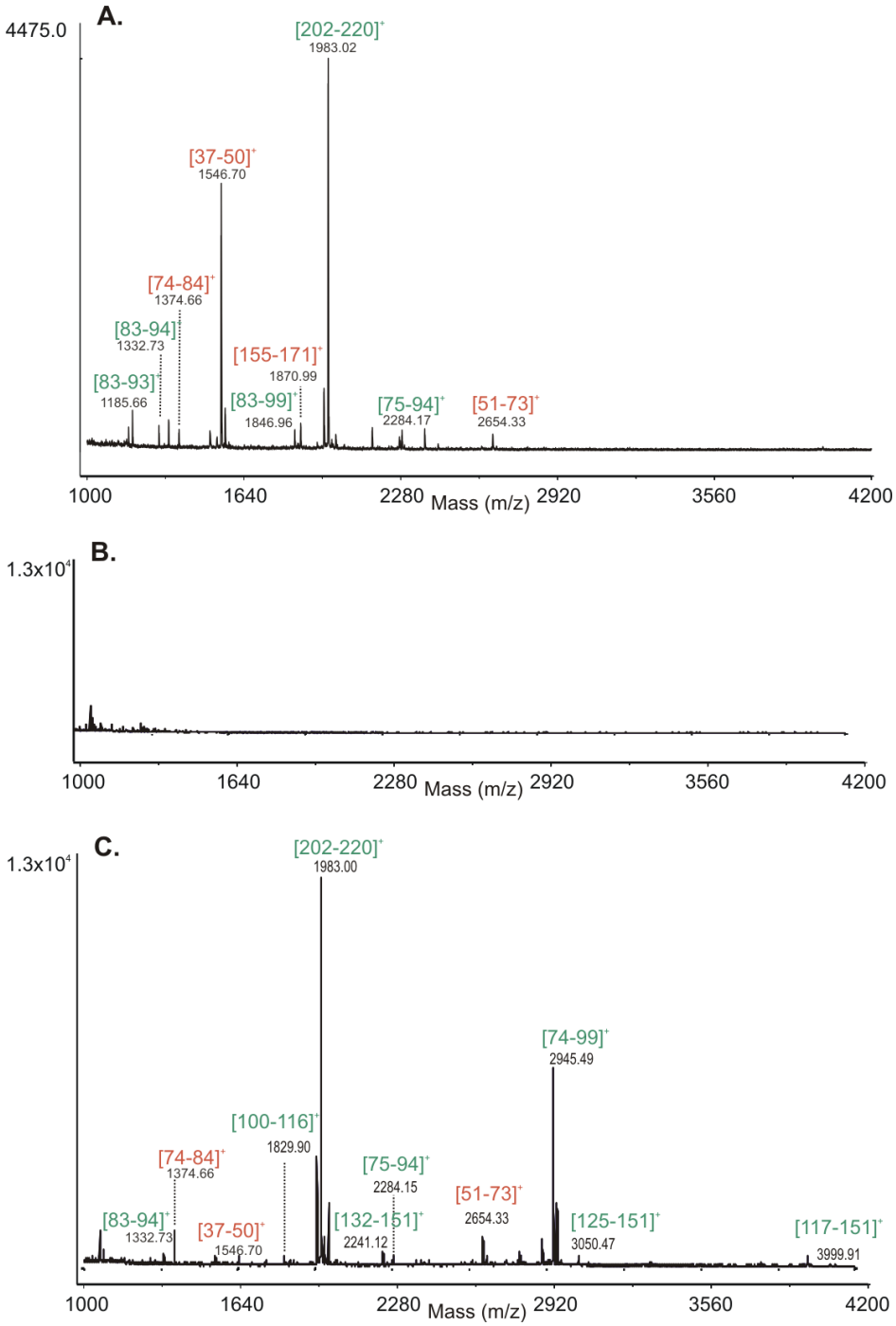


Figure 1.18. MALDI-TOF MS spectra corresponding to different fractions of the excision experiment: (**A**) On column digestion with trypsin of the complex lacNAc-ECA. (**B**) Supernatant after washing. (**C**) Elution fraction. Sugar-interacting peptides are colored in green whereas peaks in red show peptides with no direct contact with the sugar (peptide-peptide interaction).

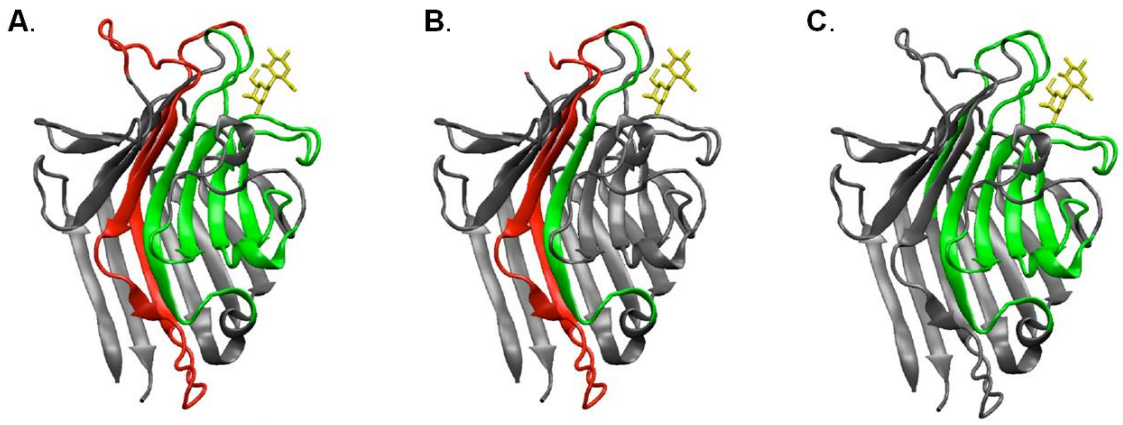


Figure 1.19. X-ray crystal structure (PDB 1GZC) of ECA in complex with lactose (in yellow). (**A**) Peptides identified in the elution fraction of an excision experiment with trypsin are shown in green (sugar-peptide interaction) and red (peptide-peptide interaction) in the ribbon representation. (**B**) Peptides [37-50], [51-73] and [74-84] (in red), non-covalently bound with spatially close [202-220] (in green), "ride" with this peptide in the elution fraction. (**C**) In an excision experiment with two consecutive digestions, only peptides involved in sugar-peptide interactions (in green) are detected[153].

In order to confirm the binding site identification, an additional chymotrypsin digestion subsequent to the trypsin excision was performed. Figure 1.20 shows that, after sequential trypsin-chymotrypsin digestions, the longest peptide [202-220] is split into shorter fragments and the "riding peptides" [37-50], [51-73] and [74-84]) are no longer observed in the elution fraction, while the specific binding peptides [74-99], [117-151], as well as non-digested [202-220] remain present (Figure 1.20E and Figure 1.19C). This result proves that the Gal(β1-4)GlcNAc-ECA interaction can withstand prolonged, sequential digestion with two proteases, despite its relatively low affinity (see Table 1.3), hence making possible accurate molecular definition of the interaction partners and removal of unspecific peptides. This is valuable when proteolytic excision results in long peptides enhancing the likelihood of peptide-peptide interactions that may hamper the identification of the carbohydrate binding site.

Additional unequivocal identification of the binding site came from an extraction MS experiment (section 1.1.4.1, Figure 1.8), where ECA was first digested with trypsin, the digest was passed through the affinity column, and only peptides [202-220] and [74-99] representing the binding site were observed (Figure 1.21).

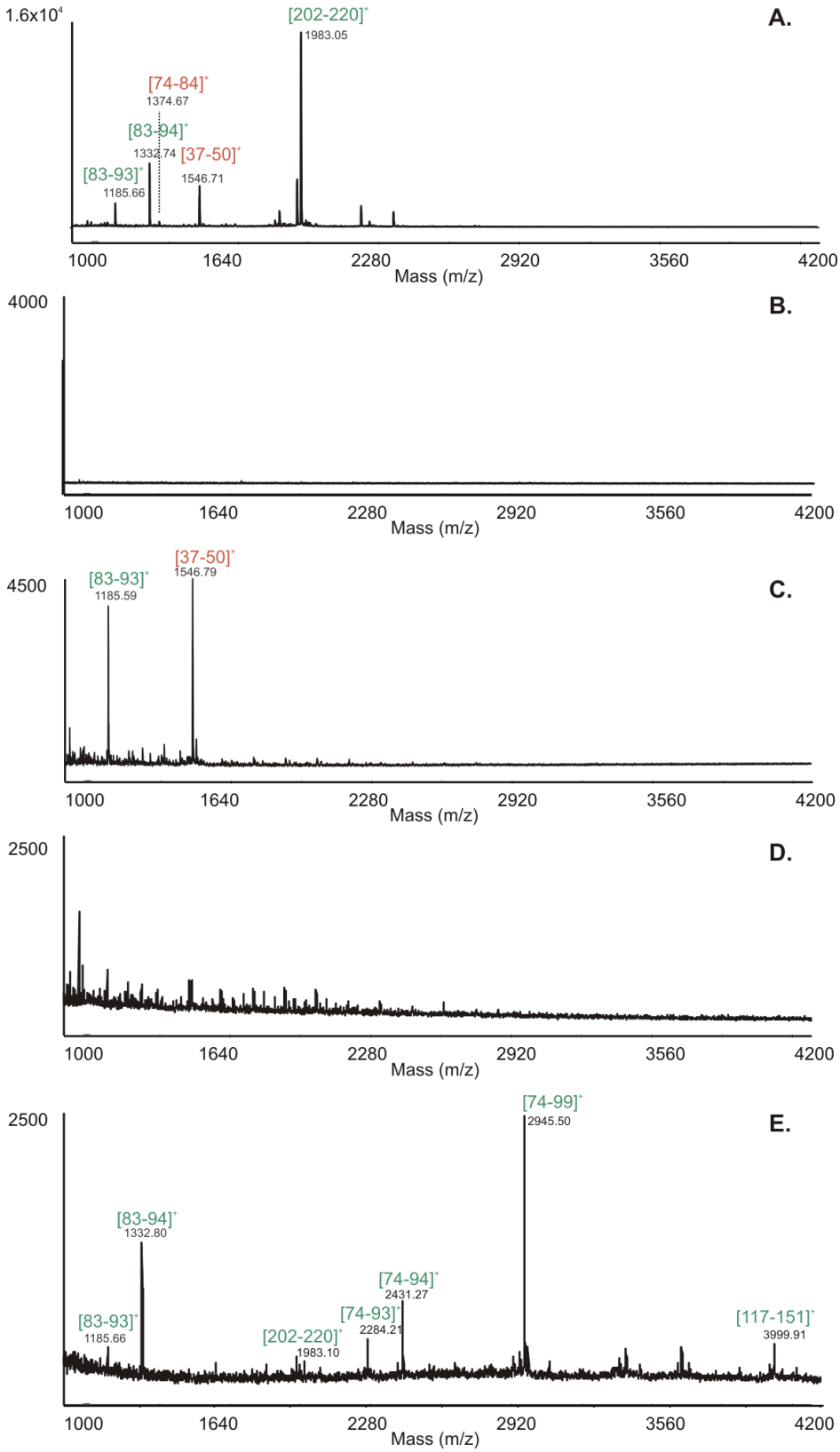


Figure 1.20. MALDI-TOF MS spectra corresponding to different fractions of excision experiment with two consecutive proteolytic digestions. (**A**) First on-column digestion with trypsin of the complex lacNAc-ECA. (**B**) Supernatant after washing. (**C**) Second on-column digestion with chymotrypsin. (**D**) Supernatant after washing. (**E**) Elution fraction. Sugar interacting peptides are colored in green whereas peaks in red show the peptides with no direct contact with the sugar (peptide-peptide interaction).

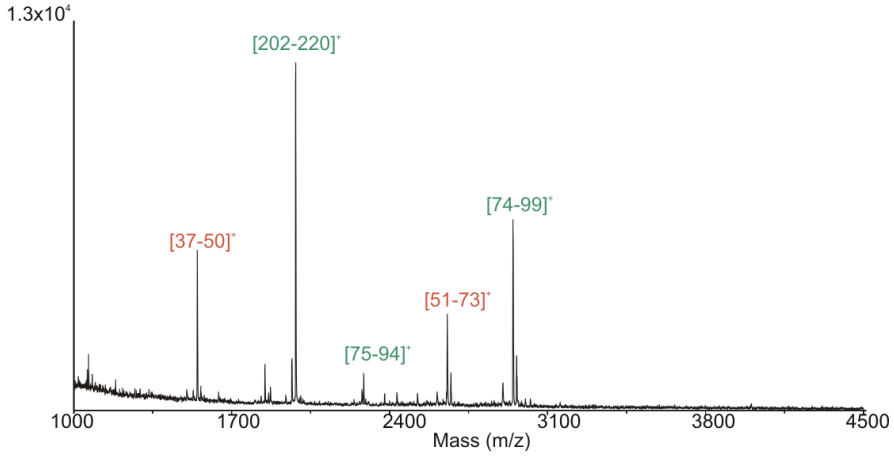


Figure 1.21. MALDI-TOF MS spectra corresponding to the elution fraction of the ECA extraction experiment. Sugar-interacting peptides are colored in green whereas peaks in red show the peptides with no direct contact with the sugar (peptide-peptide interaction).

From the CREDEX-MS experiments, in combination with the information obtained by X-ray crystallography, three specific-binding peptides were selected (the reported sugar binding residues are underlined in the sequences):

1. [74-99] FSFSIEQPYTRP**L**PA**D**GLVFFMGPTK (26-mer)

2. [117-151] QDNSYQTLAVEFDT**F**S**N**PWDPPQVPHIGIDVNSIR (35-mer)

3. [202-220] QVLPEWVDVGLSGATG**AQ**R (19-mer)

Figure 1.22. X-ray crystal structure (PDB 1GZC) of ECA in complex with lactose (in yellow). Peptides identified in the elution fractions of lacNAc-ECA excision and/or extraction experiments that were selected as candidates for further studies are shown in red (**1**), green (**2**) or blue (**3**) in the ribbon representation.

From each candidate (Figure 1.22), peptides of different length (Carbohydrate Binding Peptides, CBPs) were synthesized by Fmoc-based solid phase synthesis, in order to find the shortest CBP that binds specifically to the sugar and determine the effect of peptide sequence, length and structure in the recognition of the carbohydrate. They were satisfactorily purified by HPLC and characterized by MALDI-TOF MS:

CBP1a: [84-91] RP**L**PA**D**GL (8-mer)
CBP1b: [82-93] YTRP**L**PA**D**GLVF (12-mer)
CBP1c: [80-95] QPYTRP**L**PA**D**GLVFFM (16-mer)
CBP1d: [78-97] IEQPYTRP**L**PA**D**GLVFFMGP (20-mer)
CBP1e: [76-99] FSIEQPYTRP**L**PA**D**GLVFFMGPTK (24-mer)

CBP2a: [129-135] DT**F**S**N**PW (7-mer)
CBP2b: [127-137] EFDT**F**S**N**PWDP (11-mer)
CBP2c: [125-139] AVEFDT**F**S**N**PWDPPQ (15-mer)

CBP3a: [214-220] GATG**AQ**R (7-mer)
CBP3b: [212-220] LSGATG**AQ**R (9-mer)
CBP3c: [210-220] VGLSGATG**AQ**R (11-mer)
CBP3d: [206-220] EWVDVGLSGATG**AQ**R (15-mer)
CBP3e: [202-220] QVLPEWVDVGLSGATG**AQ**R (19-mer)

After these synthetic ECA CBPs were synthesized, purified, and carefully characterized, they were tested both by SPR and CREDEX-MS. Unfortunately, the SPR analysis of these peptides yielded ambiguous results. Although a clear interaction between the CBPs and the immobilized lacNAc glycoprobe was observed by SPR, the reference channel also yielded response, sometimes with higher values than the specific channel. Therefore, SPR sensorgrams could not be used to calculate the kinetic constants. This anomalous response in the reference channel of the synthetic ECA CBPs, not the whole ECA lectin, could not be completely clarified but a possible explanation is that the SPR format includes a 3D dextran-based matrix that facilitates diffusion and non-specific interaction of the peptides, yielding an erroneous response. Therefore, an SPR approximation to epitope mapping is less straightforward than anticipated.

Back to the CREDEX-MS approach, peptides from each CBP family were combined in an equimolar mixture and subjected to a single extraction experiment with lacNAc-DVS-S. As shown in Figure 1.23, some specific CBPs were observed in the elution fraction of each family experiment despite slight differences with the rest of the members. Thus, for the CBP1 family (Figure 1.23A), three specific peptides (CBP1b, CBP1c, and CBP1e) showed better affinity for the sugar immobilized. In the CBP2 family (Figure 1.23B), only CBP2b seemed to bind the sugar better. Finally, in CBP3 family (Figure 1.23C), CBP3e was the only one appearing in the elution fraction.

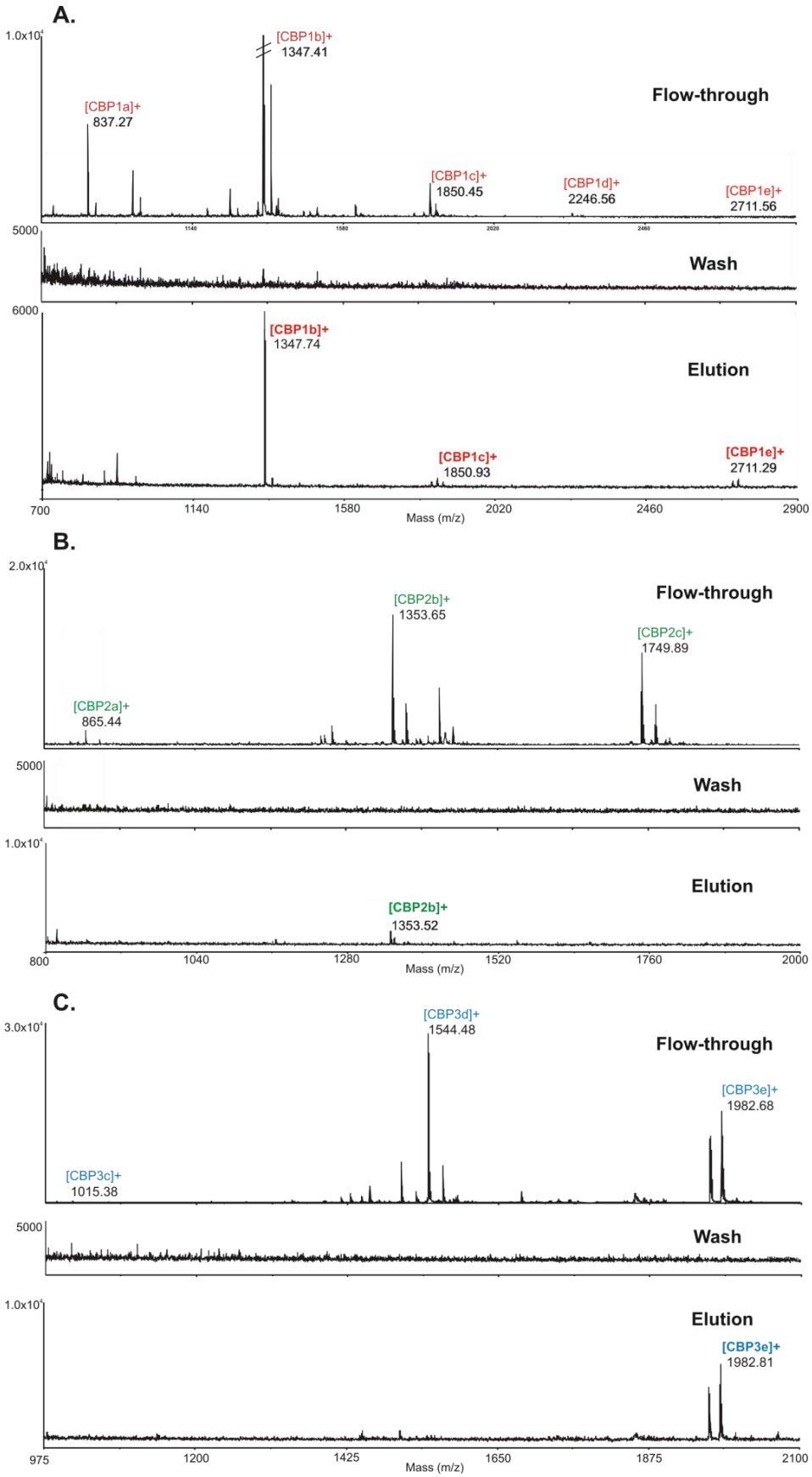


Figure 1.23. MALDI-TOF MS spectra corresponding to different fractions of an extraction experiment with a mixture of ECA CBPs of a same family: (**A**) CBPs of peptide candidate 1, (**B**) candidate 2 or (**C**) candidate 3.

Hence, results showed that, depending on each specific-binding peptide candidate (CBP1,2 or 3), different affinities exist between the corresponding CBPs (CBPa,b,c,…), though no general trend nor singular effects in the recognition of the carbohydrate could be observed regarding peptide length. Thus, specific peptide sequence and structure is what differentiates each CBP from the rest of family members and leads to recognition.

In summary, detailed molecular description of carbohydrate-protein interactions is feasible by the combination of SPR and CREDEX-MS, which use low amounts of both lectin and carbohydrate compatible with extraction from natural sources, in contrast with more sample-demanding techniques such as NMR or X-ray crystallography. The agglutinin ECA was chosen as a case study to test the applicability of these techniques. On the one hand, SPR-based experiments showed a higher affinity of ECA for lacNAc relative to other β-galactosides. Other applications of the SPR technology not yet exploited for carbohydrate-lectin interactions were evaluated during this thesis. First, additional structural information on the interaction was also provided by SPR, by comparing the differential binding responses between epitopes with subtle differences (*i.e.* glycosidic linkage or *N*-acetyl group at position C2). This analysis showed the hydroxyls at C3 and C6, as well as the *N*-acetyl at C2, being critical for interaction with ECA. Secondly, thermodynamic data on the interaction were also derived by SPR. While enthalpy values were equivalent to those obtained by ITC or NMR, the higher affinity constants determined by SPR translated into larger differences in entropy relative to ITC or NMR. Such discrepancies may be due to differences in the way the interaction studies are performed. Thus, whereas by ITC or NMR an interaction can be seen in a solution where all molecules are totally free, in SPR the surface immobilization of one reactant may restrain its kinetic and/or thermodynamic parameters. Thirdly, SPR-MS technology was successfully applied as a lectin capture platform for subsequent MS analysis. While this application of the SPR technology has been tested for other interactions such as antigen-antibody[103], in this thesis we have tested it for weaker interactions such as carbohydrate-lectin. Finally, SPR-based results provide an efficient combination with CREDEX-MS, which provides a molecular definition of the carbohydrate-binding site. Altogether, we have demonstrated that the SPR and CREDEX-MS approaches described here are perfectly complementary, constituting a valuable set of tools for decrypting carbohydrate-protein interaction details.

CHAPTER 2:

Bovine seminal protein – BSP-A1/A2, PDC-109

2 Bovine seminal protein – BSP-A1/A2, PDC-109

2.1 INTRODUCTION

In the male genital tract, secretions from the testes, epididymis, seminal vesicles, and other accessory glands contribute to the fluid portion of semen (*i.e.* seminal plasma), in which sperm cells (*i.e.* spermatozoa) are suspended. Seminal plasma of mammals is a complex fluid, which serves as a carrier for the spermatozoa in their journey from the testes to their target, the uterus. It contains both organic and inorganic molecules of low as well as high molecular weight. While the low molecular weight fraction contains a wide variety of chemical constituents such as metal ions, organic acids, sugars, lipids and amino acids, the only high molecular weight constituents found in seminal plasma are proteins; other biopolymers such as polysaccharides and nucleic acids are absent[154]. The panel of proteins identified in seminal plasma ranges from hormones, enzymes, proteinase inhibitors, growth factors and other components to proteins and glycoproteins of yet unknown nature or function. Since the beginning of the 20th century, researchers realized that the seminal plasma played more roles in sperm biology than a mere vehicle for spermatozoa. Growing evidence has pointed to this 'acellular' part of the semen as a fundamental player in mammalian fertilization and during recent years, intensive research has been performed regarding the characterization and function of seminal plasma proteins. However, the variability found in its protein composition among species, individual males and even fractions of the same ejaculate has made difficult to completely understand its effect in sperm function modulation.

The spermatozoa at the time of release from the testis are non-fertilizing. They acquire this capacity during epididymal transit. Apparently, the capacity is suppressed when sperm cells are accompanied by seminal plasma at the time of ejaculation, but fully recovered during residence in the female fallopian tube by a process called "capacitation" (*i.e.* a cascade of maturation events that are required for successful fertilization). Despite the fact that the process of capacitation was first described 60 years ago, its underlying molecular basis has only recently begun to be understood[155,156]. Studies on several mammalian species have provided evidence for a regulatory role of seminal plasma proteins in capacitation, in which a membrane

remodeling process involving changes in membrane fluidity and in composition, structure, and topography of integral and peripherial membrane proteins has been reported[157]. Since sperm cells have a limited biosynthetic ability[158], the interaction with the different seminal plasma proteins found along their journey to the site of fertilization will have a profound impact on sperm functionality. Proteins from the seminal fluid will be adsorbed into the sperm surface, remodeling the structure of the sperm membrane protein domains. These changes will endow the spermatozoa with the ability to fertilize the egg and will influence several essential steps of the fertilization process, such as capacitation, establishment of an oviductal sperm reservoir, modulation of the uterine immune environment and gamete interaction[159].

Therefore, studies on nature, structure, and functional properties of the protein constituents of seminal plasma, and their interactions with molecular ligands and spermatozoa, might provide clues regarding the role of seminal plasma in modulating the functional ability of spermatozoa. These insights will no doubt help advance aspects of reproductive biology such as diagnostics and treatment of dysfunction, since semen sample reflects the status of the male reproductive organs.

2.1.1 **Bovine seminal plasma proteins (BSP proteins)**

Development of high-throughput proteomics technologies has largely increased the number of proteins identified in sperm and seminal plasma. Although differences in the type and family of proteins are found among species, the current view is that their roles in the main events that lead to fertilization are essentially conserved[160]. Most of these seminal plasma proteins have been classified into three major families: (i) cysteine-rich secretory proteins (CRISP), (ii) proteins containing the fibronectin type II domain [Fn-2, also known as bovine seminal plasma (BSP) proteins] and (iii) spermadhesins[161].

The Fn-2 family represents the major protein fraction of bovine seminal plasma. It consists of four acidic bovine seminal plasma (BSP) proteins designated BSP-Al, BSP-A2, BSP-A3, and BSP-30kDa[162], which are secreted by the seminal vesicles and characterized by their conserved gelatin-binding fibronectin II type module. Similar proteins are also present in seminal plasma and/or seminal vesical secretions from humans[163], hamsters, mice, rats[164], stallions[165], and boars[166], suggesting that BSP analogues may be ubiquitous in mammals and may possible be involved in a common function.

These proteins constitute the major heparin-binding protein fraction of bovine fresh seminal fluid (30-50 mg/mL), and together comprise an average 47% of the total protein fraction[167]. BSP-Al, -A2, and -A3 have molecular mass of 15-16 kDa, whereas BSP-30kDa has a mass of 28-30 kDa[168]. Since BSP-A1 and BSP-A2 are present in seminal plasma in roughly equimolar concentrations and have identical amino acid sequence, differing only in the degree of glycosylation, they are considered to be a single chemical entity named BSP-A1/A2 (also called PDC-109)[169]. Additionally, all three proteins (BSP-A1/A2, -A3 and -30 kDa) share biochemical properties such as binding to gelatin, heparin, apolipoprotein A-I, glycosaminoglycans, choline phospholipids and low-density lipoproteins. Moreover, they play an important role in fertilization, including sperm capacitation and formation of oviductal sperm reservoirs, made possible by their binding to the choline group of phospholipids on the sperm membrane[170], and to high-density lipoproteins (HDL)[171] and heparin-like glycosaminoglycans (GAGs) present in the follicular and oviductal fluids[172,173]. Significantly, BSP proteins were also reported to have negative effects in the context of sperm storage[174].

2.1.2 **BSP-A1/A2 or PDC-109 protein**

PDC-109 is the most abundant of the heparin-binding BSP proteins, with a seminal plasma concentration of 15-20 mg/mL[175]. Thus, PDC-109 alone represents on average a 38% of the total protein fraction, whereas BSP-A3 and BSP-30kDa only a 3-4%[167]. At ejaculation, around 8 million PDC-109 molecules bind to the sperm surface, specifically to choline phospholipids[170] that comprise over 70% of total bovine sperm plasma membrane phospholipids.

Bovine PDC-109 has been purified and characterized by various researchers. Manjunath *et al.* isolated PDC-109, BSP-A3, and BSP-30kDa using gelatin-agarose affinity chromatography[162]. The absorbed fractions were resolved on Sephadex G-75 and G-200 and the BSP proteins isolated by this procedure were lipid-free. Calvete *et al*. described isolation of PDC-109 by RP-HPLC[175]. More recently, the same laboratory used a combination of size-exclusion, heparin affinity and DEAE ion exchange chromatography[176] for the stepwise isolation of bovine seminal plasma proteins, including aSFP, PDC-109, TIMP-2, BS RNase and a BSP-30K/PDC-109 complex. Major proteins were recovered in sufficient amount and purity for structural studies.

In view of its high abundance in bovine seminal plasma and its binding to spermatozoa, PDC-109 has been investigated extensively by a number of biochemical and biophysical studies that have outlined its binding properties and sperm capacitation potential.

2.1.2.1 *Biochemical properties*

PDC-109, "the peptide (P) having N-terminal aspartic acid (D) and C-terminal cysteine (C) with 109 amino acids", exists as previously mentioned in two forms, *i.e.* BSP-A1 and BSP-A2, the presence of a single O-linked oligosaccharide (Neu5Acα(2–6)-Galβ(1–3)-GalNAc-) attached to Thr11 of BSP-A1 being the only difference between these two forms[175,177]. BSP-A1/A2, collectively termed PDC-109, is an acidic protein of pI 5-5.5[168], composed of an *N*-terminal 23-residue *O*-glycosylated acidic stretch followed by two structurally similar domains (*a* and *b*) of 40 and 41 residues, respectively, each containing two disulfide bridges. Each of these domains displays the invariant residues and the disulfide bond topology of the fibronectin type II (Fn II) module[169] and is linked and clustered by a short polypeptide (Figure 2.1). The three-dimensional structure of PDC-109 complexed with *O*-phosphorylcholine, solved by single-crystal X-ray diffraction reveals that each Fn-II domain binds one choline-phospholipid molecule and that both binding sites are on the same face of the PDC-109 protein[178]. Ligand binding is mediated by a cation-π interaction between the quaternary ammonium group of the choline moiety and the indole ring of several core Trp residues (Trp47, Trp58, Trp93 and Trp106) plus hydrogen bonding between the phosphate group and exposed Tyr residues of the protein[178]. The solution structure of the C-terminal Fn2 domain, PDC-109 domain *b*, has also been determined by NMR spectroscopy and been shown to also bind collagen in a specific binding pocket[179].

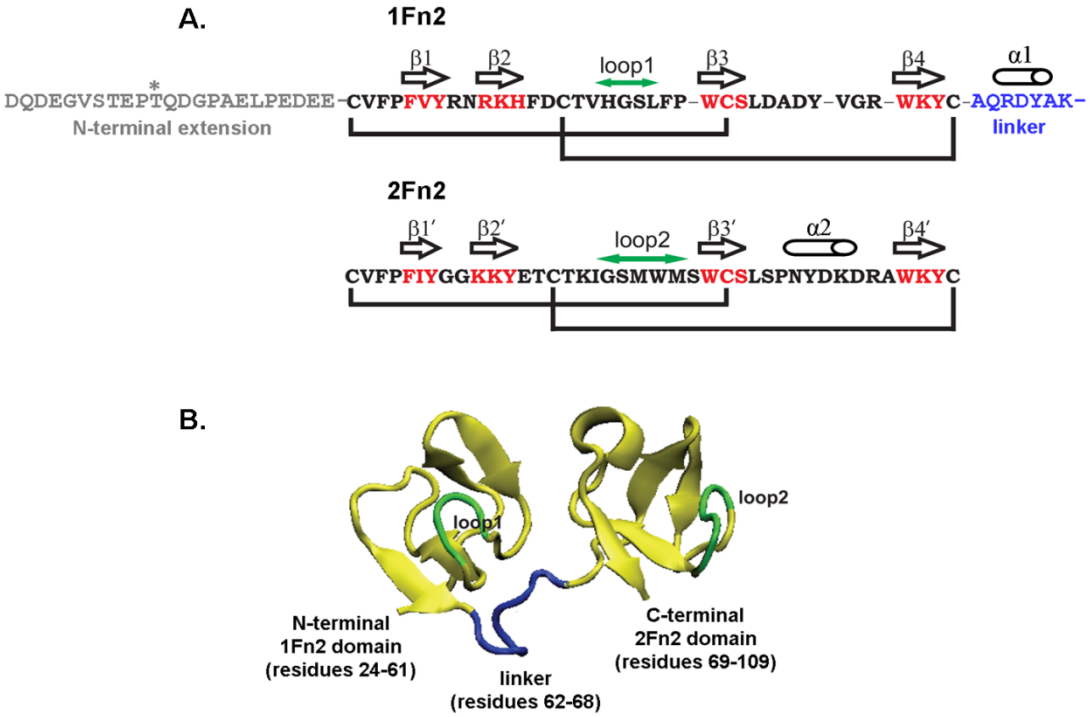


Figure 2.1. (**A**) Amino acid sequences of the N-terminal and C-terminal Fn2 modules of PDC-109, labeled 1Fn2 and 2Fn2, respectively, showing the location of secondary structure elements (arrows, β strand; cylinders, α helix). Disulfide bonds are indicated with black lines and an asterisk marks the position of the O-glycosylated Thr11 within the N-terminal extension (in grey), which is not visible in the crystallized protein. (**B**) X-ray crystal structure (PDB 1H8P) of PDC-109 monomer. The two tandem Fn2 domains are connected by a 7-residue linker peptide shown in blue. Loop 1 (residues H41-L44) between β2 and β3 strands and loop 2 (residues G87-M91) between the β2' and β3' strands are denoted by green arrows. Modified after [180].

Combination of size-exclusion chromatography, analytical ultracentrifugation, circular dichroism, Fourier-transform infrared spectroscopy, and differential scanning calorimetry has allowed the biophysical characterization of PDC-109 and its interaction with phosphorylcholine. Under native conditions and in aqueous solution, PDC-109 exists as a polydisperse multimeric self-associated molecule, its average oligomeric size changing with ionic strength and the presence of choline ligands, suggesting that both ionic and hydrophobic interactions are responsible for the aggregation tendency of PDC-109 monomers[181]. Ligand binding to PDC-109 results in disaggregation of the polydisperse protein, yielding a mixture of monomers and dimers; the latter being revealed in solution through the evaluation of high-resolution single crystals[182]. The disassembly may proceed by neutralization and partial structuration of exposed molecular surfaces on PDC-109 monomers involved in multimerization, *i.e.* modification in the environment of Tyr and Trp aromatic side chains, solvent exposure

of peptide bonds and a slight increase in the turn content at the expense of unordered segments. Another report on the oligomeric state of PDC-109 confirmed that the heparin binding region of PDC-109 is conformational and quaternary structure-dependent[183]. Likewise, the aggregation state of PDC- 109 has been shown to be important for its chaperone-like activity[184].

More recently, PDC-109 from bovine seminal plasma has been characterized by high-resolution Fourier transform ion cyclotron resonance (FTICR) MS. In addition to the glycosylated and non-glycosylated forms (*i.e.* BSP-A1 and BSP-A2, respectively), a number of previously unknown sequence variants were identified by top-down MS[185]. For example, a protein variant containing two point mutations (P10L and G14R) was identified along with another form having a 14-residue truncation in the N-terminal region. Two other minor truncated variants could also be identified from the affinity-purified PDC-109. Furthermore, native MS was used to probe the oligomeric state of PDC-109. At low protein concentrations, however, only the monomeric protein was observed, suggesting that the protein oligomers are weakly bound and dissociate easily into monomers, *i.e.* transient oligomers.

Altogether, these results demonstrate that PDC-109 is naturally produced as a mixture of several protein forms and its aggregation state seems to play an important role in modulating its interaction with other biomolecules.

2.1.2.2 *Functional properties*

The interaction of BSP proteins with spermatozoa is mediated by their interaction with specific phospholipids, particularly with zwitterionic phosphatidylcholine (PC)[170]. However, in addition to their heparin- and lipid-binding activities, BSP proteins interact with a variety of ligands including different collagens (types I, II, IV, V), fibrinogen, apolipoprotein Al, apolipoprotein Al / high-density lipoprotein (HDL) complexes, and calmodulin. This varied binding profile of BSP proteins, and PDC-109 in particular, suggests a multifunctionality, brought about by affinity towards various ligands (Table 2.1).

Table 2.1. Multifunctional properties of bovine seminal PDC-109 protein**.**

PDC-109 functional property	Evidence	References
Sperm binding	After ejaculation PDC-109 is located underneath the plasma membrane of the sperm middle piece.	[186]
Cholesterol efflux	At ejaculation, PDC-109 bind to choline lipids from sperm cell membranes, causing a conformational change after which the protein acquires the ability to recognize cholesterol and induce the first cholesterol efflux. The resulting sperm membrane destabilization and reorganization could increase the HDL- and heparin-docking sites on sperm surface.	[187]
	PDC-109-coated spermatozoa could interact with HDL in the latter part of the female genital tract, which would stimulate a second cholesterol efflux, resulting in further decrease in the cholesterol:phospholipids ratio, provoking fluidity changes and reorganization or destabilization of membrane.	[188]
Oviductal sperm reservoir formation	PDC-109 enables sperm to bind to oviductal epithelium and plays a major role in formation of the bovine oviductal sperm reservoir.	[189]
Sperm motility	Strong stimulatory effect of PDC-109 on sperm motility indicating activity of sperm plasma membrane-bound Ca^{2+}-ATPases.	[190]
Capacitation	PDC-109 binds to capacitation factors present in the oviduct such as HDLs and heparin at particularly high concentrations during the oestrus cycle modulating spermatozoa-induced capacitation.	[191, 192]
	Ability of heparin to potentiate PDC-109-induced sperm capacitation by modulation of PDC-109 aggregation state.	[193]
Acrosome* reaction	PDC-109 increases the proportion of acrosome-reacted spermatozoa, but only in the presence of heparin.	[173, 194]
Cryoinjury	Continuous sperm cholesterol efflux induced by prolonged exposure to free floating PDC-109 destabilizes the sperm membrane, making sperm prone to cryodamage.	[195]
	Sequestration of PDC-109 from ejaculates improves semen quality parameters in pre-freeze and post-thaw stages of cryopreservation.	[196]

* A membrane-bound cap-like organelle covering the tip of sperm head and containing enzymes that digest the outer surface of the egg, thus permitting sperm penetration.

In view of the data exposed above, it has been proposed that PDC-109 in particular, and BSP proteins in general, might be involved in the modification of the sperm membrane lipid composition at capacitation[197] following the subsequent *in vivo* mechanism of action during fertilization[198] (see Figure 2.2A). At ejaculation, sperm are exposed to PDC-109 proteins present in seminal plasma that remove a significant amount of

cholesterol accompanied by the release of some phospholipids. At the same time, sperm are coated with PDC-109 proteins through their interaction with membrane choline phospholipids. This PDC-109 binding, together with the first cholesterol efflux, may induce reorganization of the sperm membrane. Within the next 10–20 min sperm travel through the cervical mucus into the uterus leaving behind most of the seminal fluid; excess free floating PDC-109 protein is retained in the early part of genitalia, whereas that bound to spermatozoa travels further on, to participate in the fertilization process. Sperm coated with PDC-109 proteins may prevent free movement of phospholipids and thereby stabilize the membrane and prevent premature acrosome reaction of spermatozoa (arrested/decapacitated state). Thereafter, decapacitated spermatozoa (coated with PDC-109 proteins) travel through the female genital tract, where an oviductal sperm reservoir is formed by carbohydrate-mediated binding with the oviductal epithelium. While bound to epithelial cells, spermatozoa are quiescent and await ovulation. Endocrine signals associated with impending ovulation induce a change in sperm biochemistry characterized by an increase in reactive oxygen species, intracellular cyclic adenosine monophosphate levels and Tyr phosphorylation. Simultaneously, PDC-109-coated spermatozoa interact with capacitation factors now available at high concentrations in the oviduct such as HDL and heparin, which stimulate the second sperm cholesterol efflux; this causes a further decrease in the cholesterol:phospholipid ratio (membrane fluidity changes), modification of the composition of surface-adsorbed components, and redistribution of membrane proteins and lipids. In response to these signals, calcium is released from an intracellular store in a pulsatile manner inducing the expression of hyperactivated sperm motility (*i.e.* increase in flagellar bend amplitude and in beat asymmetry). The combined effects of these intracellular and extracellular events, mainly in the apical sperm-head plasma membrane and the flagellum, may complete capacitation, regulating the surface expression of sperm-egg receptors for subsequent binding and rendering spermatozoa responsive to inducers of the acrosome reaction[199]. Moreover, in this hyperactivated and capacitated state, spermatozoa are released from the oviductal epithelium, leave the reservoir and migrate up the Fallopian tube to fertilize the oocyte.

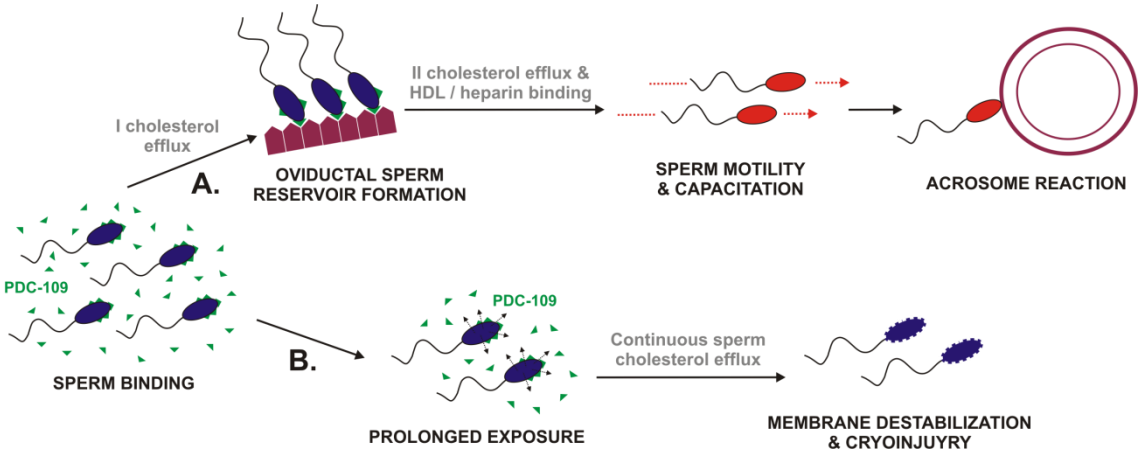


Figure 2.2. Functional roles of bovine seminal PDC-109 protein. (**A**) *In vivo* protein bound to spermatozoa stimulates cholesterol efflux from sperm plasma membrane, provoking fluidity changes, membrane destabilization and a reorganization of membrane-bound components. PDC-109 protein also promotes sperm reservoir formation by interacting with oviductal epithelium, enhances sperm motility, and induces sperm capacitation and acrosome reaction. (**B**) *In vitro* prolonged exposure of spermatozoa to excess free floating unbound PDC-109, as during processing for preservation, causes continuous efflux of sperm cholesterol leading to membrane destabilization and cryoinjury. Uncapacitated and capacitated sperm shown in blue and red, respectively. Binding to, and release from, the oviductal sperm reservoir appear to be involved in the regulation of capacitation.

In short, ultrastructural and biochemical studies suggest that sperm undergoing capacitation exhibit a membrane remodeling process in which PDC-109 is allegedly involved (Table 2.1). In fact, since incubation of sperm with a high concentration of PDC-109 accelerates capacitation[188], prolonged contact of sperm with seminal plasma could influence the amount of capacitated sperm. However, it was also reported that the cholesterol efflux induced by PDC-109 destabilizes sperm membranes, making them prone to cryodamage (Figure 2.2B). Therefore, PDC-109 in seminal plasma has the potential to act both as beneficial and detrimental for spermatozoa, depending on its concentration in seminal plasma and on exposure time, as during cryopreservation. For that reason, it is important to understand the interaction of PDC-109 with sperm cells, in an attempt to master the molecular events involved in the capacitation process. Most importantly, such an understanding can potentially lead to the development of novel fertility-related drugs.

2.1.2.3 *Lectin-like activity of PDC-109*

When they reach the isthmus, spermatozoa are bound to the ciliated epithelial cells, forming a sperm reservoir. This process seems to be mediated by carbohydrate residues present in the oviductal epithelial cells and lectin-like proteins on the sperm head[200]. The molecules involved in this process vary among species[201].

In cattle, fucose was detected on the surface of the oviductal epithelium by the Fuc-binding lectins from *Lotus tetragonolobus* (LTA) and *Ulex europaeus* (UEA-I). Pre-treatment of oviductal explants with fucosidase, fucoidan or by its Fuc constituent reduced sperm binding[202]. Furthermore, Fuc in an (α1–4) linkage to GlcNAc, as in the blood group trisaccharide Lewis^a (Le^a), was more effective at inhibiting sperm binding than any other linkage[203]. The Fuc-binding molecule was detected on bull sperm using fluorescent Fuc- and Le^a-neoglycoproteins, which labelled living sperm over the acrosomal region[203,204]. Thus, it was determined that Fuc-containing molecules are on the surface of the oviductal epithelium while Fuc-binding lectin-like molecules are on the surface of sperm (Figure 2.3). Concerning the latter, strong evidence supports the involvement of PDC-109 in sperm reservoir formation, acting as a lectin that recognizes Fuc residues in the epithelium. In fact, a Fuc-binding molecule purified from sperm extracts and identified as PDC-109[189,205] was shown to enable spermatozoa binding to the oviductal epithelium. This protein is either lost, modified or loses its affinity for fucose when sperm are capacitated, which may account for their release from the epithelium.

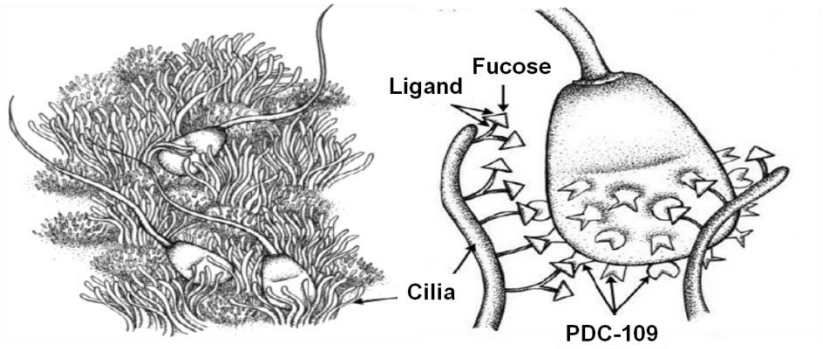


Figure 2.3. Sperm-oviduct binding in the bovine model is mediated by PDC-109, a lectin-like protein present in the sperm plasma membrane that recognizes Fuc in epithelial cell membranes.

2.2 OBJECTIVES

In line with the general scope of this thesis on studying carbohydrate-lectin interactions by SPR and CREDEX-MS, and once a proof of principle had been established and the interaction of the β-Gal-specific ECA lectin was investigated in detail (chapter 1), the next challenge was to extend the methodology to a biologically relevant target.

In this regard, the work in this chapter focuses on the lectin-like properties of PDC-109 (BSP-A1/-A2), the major protein of bovine seminal plasma, which putatively plays important roles in fertilization. To this end, highly purified bovine seminal plasma PDC-109 was obtained from Dr. J. J. Calvete (Instituto de Biomedicina-CSIC, Valencia) and its binding to several relevant oligosaccharides was evaluated by means of SPR and CREDEX-MS. Hence the goals of this chapter can be outlined as follows:

- Complete the lectin-binding profile of PDC-109 by evaluating specificity against hitherto unreported glycosidic epitopes.
- Establish the carbohydrate recognition domain of PDC-109.
- Provide insights into the functional properties of PDC-109 along the fertilization process.

2.3 RESULTS AND DISCUSSION

2.3.1 **Lectin profile of PDC-109 by SPR**

2.3.1.1 *Synthesis of neo-glycoprobes exposing relevant mammalian carbohydrate epitopes*

In order to establish the lectin-binding profile of PDC-109, initial work in this chapter was focused on the synthesis, purification, characterization and validation of new glycoprobes displaying some of the most relevant glycotopes in the mammalian system (Table 2.2 and Figure 2.4).

Table 2.2. Carbohydrate epitopes commonly encountered in mammalian specimens available for glycoprobe synthesis and corresponding lectins with well-defined binding properties.

	Carbohydrate	**Lectin**
Disaccharides with terminal β-galactose	Gal-β1,4-Glc	*Erythrina cristagalli* (ECA)
	Gal-β1,4-GlcNAc	
	Gal-β1,3-GlcNAc	
	Gal-β1,6-GlcNAc	
Mannobioses	Man-α1,2-Man	*Concanavalin A* (Con A)
	Man-α1,3-Man	
	Man-α1,6-Man	
Fucosylated disaccharides	Fuc-α1,3-GlcNAc	*Lotus tetragonolobus* (LTA) and *Ulex europeaus I* (UEA I)
	Fuc-α1,4-GlcNAc	
	Fuc-α1,6-GlcNAc	
Sialyl-containing oligosaccharides	Neu5Ac-α2,3-Gal-β1,4-GlcNAc	*Maackia amurensis* (MAA)
	Neu5Ac-α2,6-Gal-β1,4-GlcNAc	*Sambucus nigra* (SNA)
	Neu5Ac-α2,3/6-Gal-β1,4-Glc	MAA and SNA

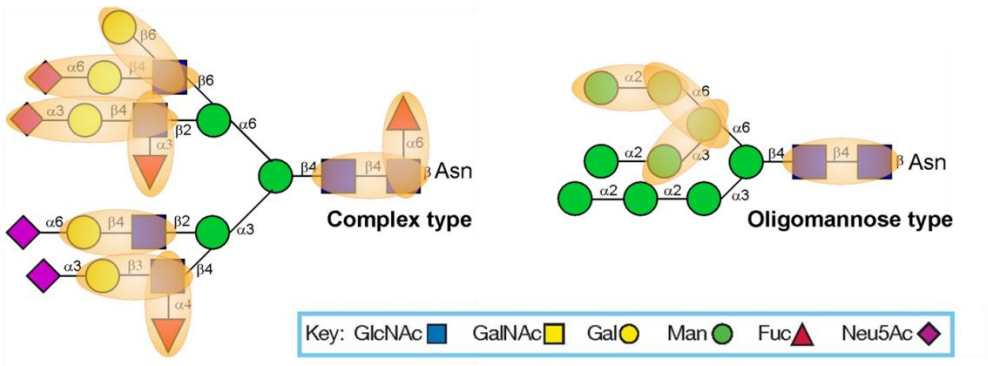


Figure 2.4. Typical complex (left) and oligomannose (right) type N-glycan structures found on mature glycoproteins, also present in the synthetic glycoprobes.

All the synthetic glycopeptides were prepared, as described before, by oxime ligation between the oligosaccharides in Table 2.2 and the optimized peptide module *N*[Me]-O-Aoa-GFKKG at 25 and 20 mM, respectively. Ligation reactions were performed at pH 3.5 (NAc-hexoses) or pH 4.6 (hexoses) and 37 ºC for 72 h, and the conjugates purified by HPLC and characterized by MS as detailed in Materials and Methods. Complementary to choosing an optimal pH for the every hexose; addition of aniline as a catalyst was evaluated with the aim of reducing incubation time, as some authors have reported significant yield improvements (from 62% to 100% in 3 h) for conjugations involving non-methylated Aoa-containing peptides and monosaccharides[206]. However, in our case and employing the *N*[Me]-O-Aoa-peptide, the yield did not increase substantially, except for a slight improvement observed after 72 h incubation, particularly in reactions with Fuc-containing disaccharides *i.e.* Fuc-α1,(3,4,6)-GlcNAc (*e.g.,* from 8 to 16% for Fuc-α1,4-GlcNAc).

A substantial problem during the synthesis of the glycopeptides was their HPLC purification, as the small change in hydrophobic properties between the glycopeptide product and unreacted *N*[Me]-O-Aoa-peptide precursor rendered separations difficult, particularly in conjugations with < 20% conversions. Attempts to increase the hydrophobic character of unreacted peptide, hence increase its retention time, by addition of carbonyl scavenger agents such as formaldehyde and acetone, or the alkylating agent *N*-ethylmaleimide[207], did not provide in our hands the desired improvement in resolution. An additional problem also related to purification, namely the presence of (trifluoroacetic) acid in typical HPLC eluents caused rapid degradation of the *N*[Me]-O-Aoa-peptides during lyophilization. Hence immediate adjustment to pH 5 of the glycoprobe solution was mandatory to preserve product integrity. Finally, the suitability of the established conjugation conditions was tested for the most acid-labile sugars, namely those containing terminal Sia units. These trisaccharides were successfully conjugated to the *N*[Me]-O-Aoa-peptide at mildly acidic pH, purified and characterized by MALDI-TOF MS, similar to other glycoconjugates. Aditionally, their long-term stability was tested and shown to be more than acceptable (> 95% of purity after two years) provided they are stored in lyophilized form.

2.3.1.2 *Glycoprobe evaluation by SPR studies with known lectins*

With neo-glycopeptides exposing the most relevant oligosaccharide epitopes of the mammalian system in hand, their binding to known lectins was evaluated by SPR interaction studies. Plant lectins in Table 2.2 had been selected based on both availability and selectivity for the indicated carbohydrate epitopes. For interaction experiments, lectin solutions were flown across the surface where the corresponding specific sugar unit had been immobilized. In this setting, Ac-GFKKG-amide, a non-glycosidic probe, was used as reference surface, given that the previously used one (*i.e.* *N*[Me]-O-Aoa-GFKKG-amide) gave an undesirable, non-specific binding response with PDC-109 protein.

Disaccharides with terminal β-galactose were first evaluated. As in chapter 1, a sensor chip with the same β-galactosides [*i.e.* Gal(β1-4)GlcNAc, Gal(β1-3)GlcNAc, Gal(β1-6)GlcNAc)] and the new Ac-GFKKG-amide reference surface was prepared. Interaction between β-galactose-specific lectin ECA and these β-galactosides yielded the same kinetic results than before, hence confirmed chip reliability.

Subsequently, glycoprobes displaying mannobioses only differing in their glycosidic linkage (Man-α1,2-Man, Man-α1,3-Man and Man-α1,6-Man) were immobilized on a sensor chip and their interaction with *Concanavalin A* (Con A) was studied. Con A is a relatively complex lectin, organized as a β-barrel-like tetramer at pH > 7, each dimer subunit (D_2 symmetry) with two CRDs situated on opposing faces of the protein. Sensorgrams of different Con A concentrations were recorded and kinetic constants determined by simultaneously fitting the experimental curves to a bivalent kinetic model (Figure 2.5). This bivalent model results in the first and second binding events being separately described by two sets of rate constants (Table 2.3).

The unconventional units (*i.e.* RU^{-1}) used to describe the second association event made comparison with constants derived from other methods (*e.g.* ITC) difficult, and only the affinity constants for the first event, K_{A1} ($K_{A1} = k_{a1}/k_{d1}$) could be compared. For this event, a higher affinity of Man-α1,3-Man over Man-α1,6-Man was found, in agreement with previous ITC data[208]; although the K_{A1} values obtained with the ITC method were one order of magnitude lower (~ 10^4) than with our SPR approach (~ 10^5). One possible explanation for this difference, keeping in mind the two-CRDs-per-dimer model of Con A, is that although the bivalent binding model used in SPR only allows to derive

standard-unit values for K_{A1}, the apparent k_{d1}, used for K_{A1} determination is also influenced by the second event. If this contribution is not factored in, a lower k_{d1}, and consequently a higher K_{A1}, results.

Table 2.3. Kinetic constants of the interaction of Con A with mannobiose-containing glycoprobes determined by SPR.

Ligand	k_{a1} (M^{-1} s^{-1})	k_{d1} (s^{-1})	k_{a2} (RU^{-1} s^{-1})	k_{d2} (s^{-1})	K_{A1} (M^{-1})
Man-α1,2-Man-*Aoa(m)-GFKKG*	1.68×10^{4}	1.33×10^{-1}	2.88×10^{-2}	7.06×10^{-2}	1.26×10^{5}
Man-α1,3-Man-*Aoa(m)-GFKKG*	2.28×10^{4}	6.61×10^{-2}	3.78×10^{-3}	6.29×10^{-2}	3.45×10^{5}
Man-α1,6-Man-*Aoa(m)-GFKKG*	8.91×10^{3}	2.89×10^{-2}	7.15×10^{-4}	1.90×10^{-2}	3.08×10^{5}

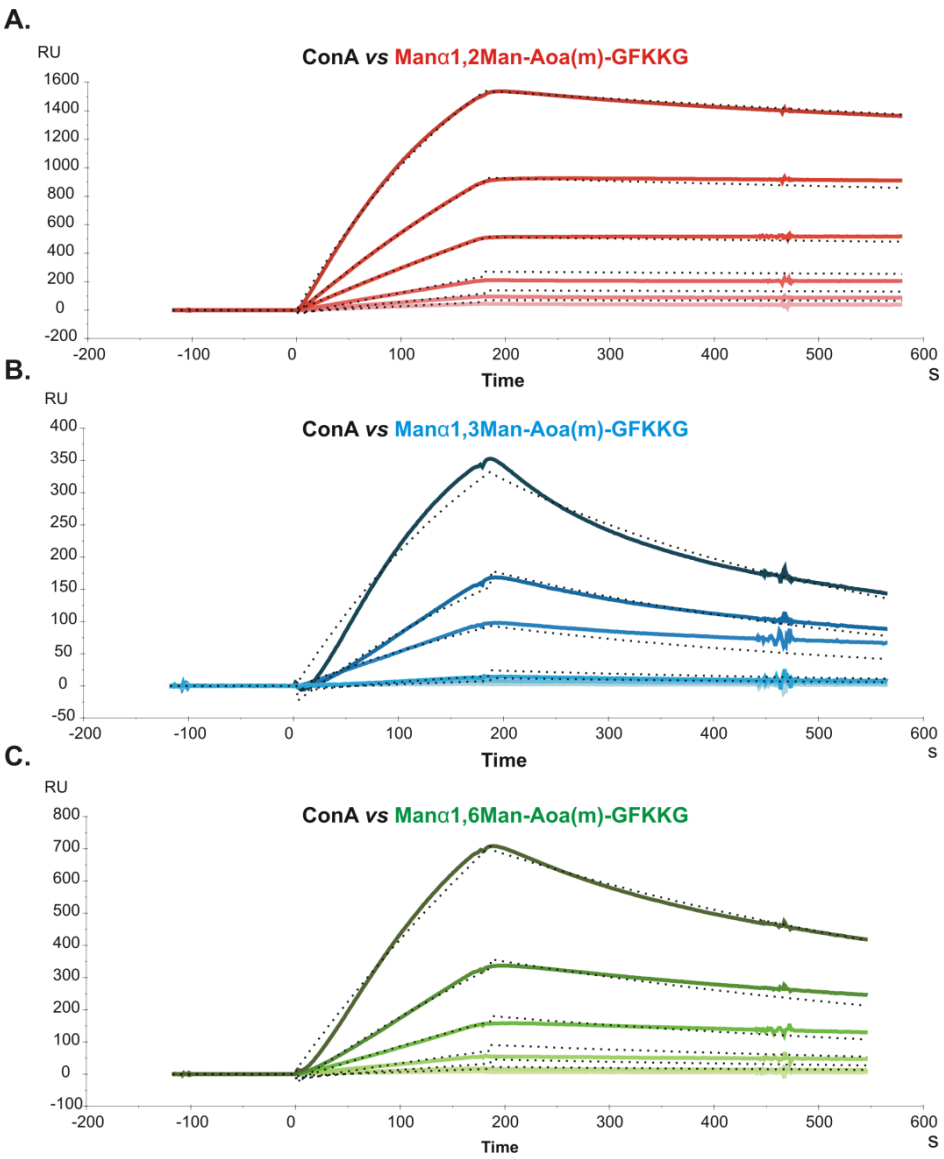


Figure 2.5. Binding of Con A to immobilized mannobioses ((**A**) Man-α1,2-Man; (**B**) Man-α1,3-Man; (**C**) Man-α1,6-Man) at six different concentrations (2.5, 5, 10, 20, 40 and 80 nM). Fitting curves using a bivalent model are shown in discontinuous black lines.

The last group of disaccharide glycoprobes to be evaluated included those with terminal Fuc units (Fuc-α1,3-GlcNAc, Fuc-α1,4-GlcNAc and Fuc-α1,6-GlcNAc). The latter structure is a potential core-fucosylation of N-glycans, and the former two are partial Lewis^{x} and Lewis^{a} epitopes, respectively. The glycoprobe surfaces were tested against two Fuc-specific lectins from *Lotus tetragonolobus* (LTA) and *Ulex europeaus* (UEA). For both lectins, sensorgrams clearly demonstrated affinity for Fuc epitopes (Figure 2.6; representative example for Fucα1,4GlcNAc), even though the carbohydrate affinity for both lectins has been described to be strongly specific for the blood group H determinant (Fuc-α1,2-Gal)[209,210], not considered in our chip.

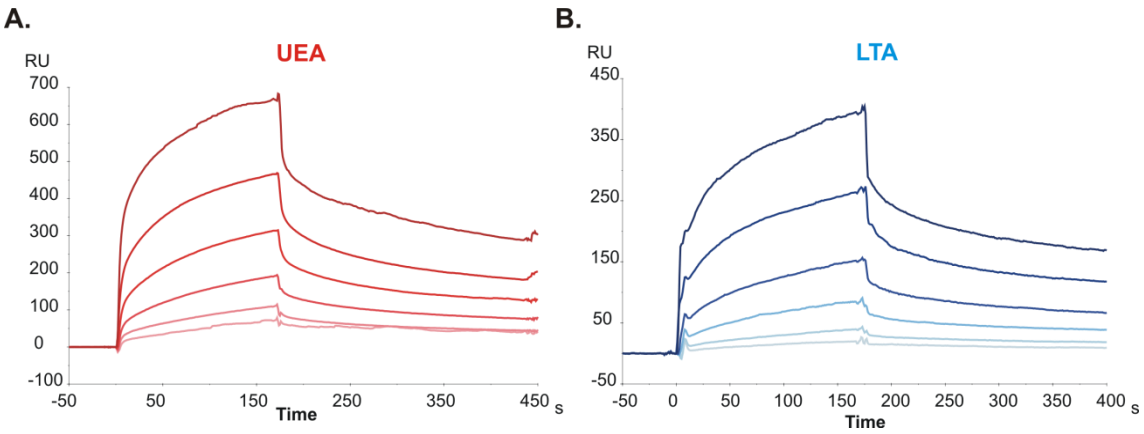


Figure 2.6. Binding sensorgrams of two Fuc-binding lectins ((**A**) UEA; (**B**) LTA) at six different concentrations from 31.25 nM to 1 µM, to Fuc-α1,4-GlcNAc–glycoprobe.

Finally, kinetic data for the interaction of Sia-containing trisaccharide glycoprobes with the *Maackia amurensis* (MAA) and *Sambucus nigra* (SNA) lectins were determined. A sensor chip with the Neu5Ac-α2,3-Gal-β1,4-GlcNAc-*N*[Me]-O-Aoa-GFKKG-amide, Neu5Ac-α2,6-Gal-β1,4-GlcNAc-*N*[Me]-O-Aoa-GFKKG-amide and Neu5Ac-α2,3/6-Gal-β1,4-Glc-*N*[Me]-O-Aoa-GFKKG-amide glycoprobes, representative of the two existing linkage types in adult mammalian glycoproteins, was prepared and the two Sia-specific lectins were flown over the surfaces. In fact, MAA is a lectin that reportedly requires three intact sugar units for binding[211]. The binding responses observed for the two lectins were in perfect agreement with their reported carbohydrate specificity. Thus, whereas SNA showed a marked preference for the Sia-α2,6-lacNAc isomer, MAA recognized only the Sia-α2,3-lacNAc containing glycoprobe (Figure 2.7).

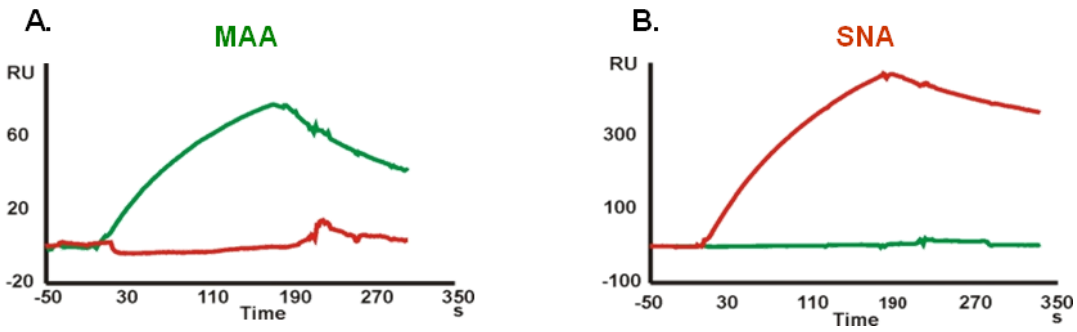


Figure 2.7. Binding sensorgrams showing the different carbohydrate specificity of both lectins ((**A**) MAA; (**B**) SNA) to their specific sugar epitopes (Sia-α2,3-lacNAc and Sia-α2,6-lacNAc in green and red trace, respectively).

For kinetic analysis, several concentrations in the 250 nM to 1.9 µM range for MAA and in the 74 to 563 nM range for SNA were analyzed. A 1:1 Langmuir binding model was used for sensorgram fitting (Figure 2.8, Table 2.4). As expected for interactions involving trisaccharides, the K_A equilibrium constants obtained were in the 10^6-10^7 M^{-1} range. For MAA, the affinity constant determined by our approach was 9.12×10^5 M^{-1}; of the same magnitude than reported previously[112], but one order lower than other constants determined using complex neoglycoproteins and immobilized lectins[212], in both cases by SPR. The higher affinity for neoglycoproteins can easily be explained by their multivalent nature. For SNA, a K_A of 6.27×10^6 M^{-1} was determined, consistent with the previously reported value[79] of 6.7×10^6 M^{-1}.

Table 2.4. Kinetic constants of MAA and SNA to their specific sialic-containing glycoprobes determined by SPR.

Lectin	Trisaccharide	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_A (M^{-1})
MAA	Neu5Ac-α2,3-lacNAc	5.55×10^3	6.08×10^{-3}	9.12×10^5
SNA	Neu5Ac-α2,6-lacNAc	1.02×10^4	1.63×10^{-3}	6.27×10^6

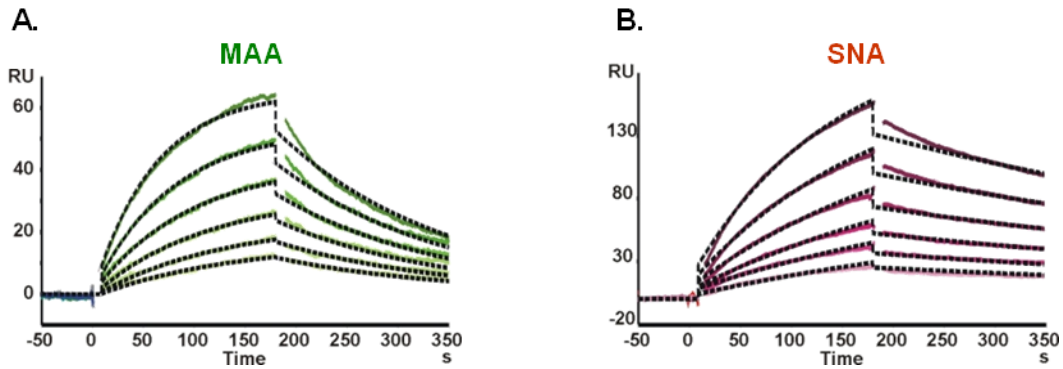


Figure 2.8. SPR sensorgrams of the two Sia-specific lectins at six different concentrations. (**A**) Binding of MAA to Sia-α2,3-lacNAc glycoprobe. (**B**) Binding of SNA to Sia-α2,6-lacNAc glycoprobe. Fitted curves are indicated as discontinuous black traces.

In conclusion, the kinetic and affinity constants of well-known plant lectins for their specific carbohydrate ligands determined by our SPR approach were consistent with previously reported data, corroborating that the glycoprobes attached to the sensor surface chips provide a reliable setting for further studies with the mammalian lectin, PDC-109.

2.3.1.3 *PDC-109 kinetic studies and binding features*

In view of the role of PDC-109 in the sperm reservoir formation, it is of particular interest to investigate its interaction with carbohydrates, especially those containing fucose residues. While the lipid selectivity and kinetics of this protein has been characterized in considerable detail[170,213], the kinetics of PDC-109 interaction with carbohydrate epitopes has not been investigated so far. Thus, in the present study SPR experiments were carried out with PDC-109 and the most relevant mammalian glycotopes. Once SPR sensor chips were functionalized with the different synthetic glycoprobes and evaluated with well-known specific lectins (section 2.3.1.2); they could be used to characterize both qualitatively and quantitatively the interaction of PDC-109 with the glycotopes. For kinetic analysis, several concentrations in the 6.25 to 100 μM range were analyzed and a 1:1 Langmuir binding model was chosen for sensorgram fitting.

The kinetic data, despite averaging 3-6 repeat experiments, exhibited considerable variability (CV values >50% in some cases). Data dispersion was attributable to the unusual sensorgrams observed, reflecting a complex binding event in the association phase of PDC-109. As shown in Figure 2.9A, the usual signal rise observed right after sample injection and due to bulk effects—refractive index differences between analyte solution and flow buffer—was followed by a rather abnormal association phase. In a conventional, straightforward binding event, the response—*i.e.*, the signal rise over time, reflecting protein binding to the sensor surface—reaches a constant value once equilibrium is established, and remains so until the protein solution passed over the sensor chip is replaced with buffer, giving way to the dissociation process and subsequent surface regeneration. However, in the present case visual inspection of all PDC-109 sensorgrams (represented by those in Figure 2.9A) clearly shows a biphasic phenomenon during the association curve displayed (enlarged area in Figure 2.9A).

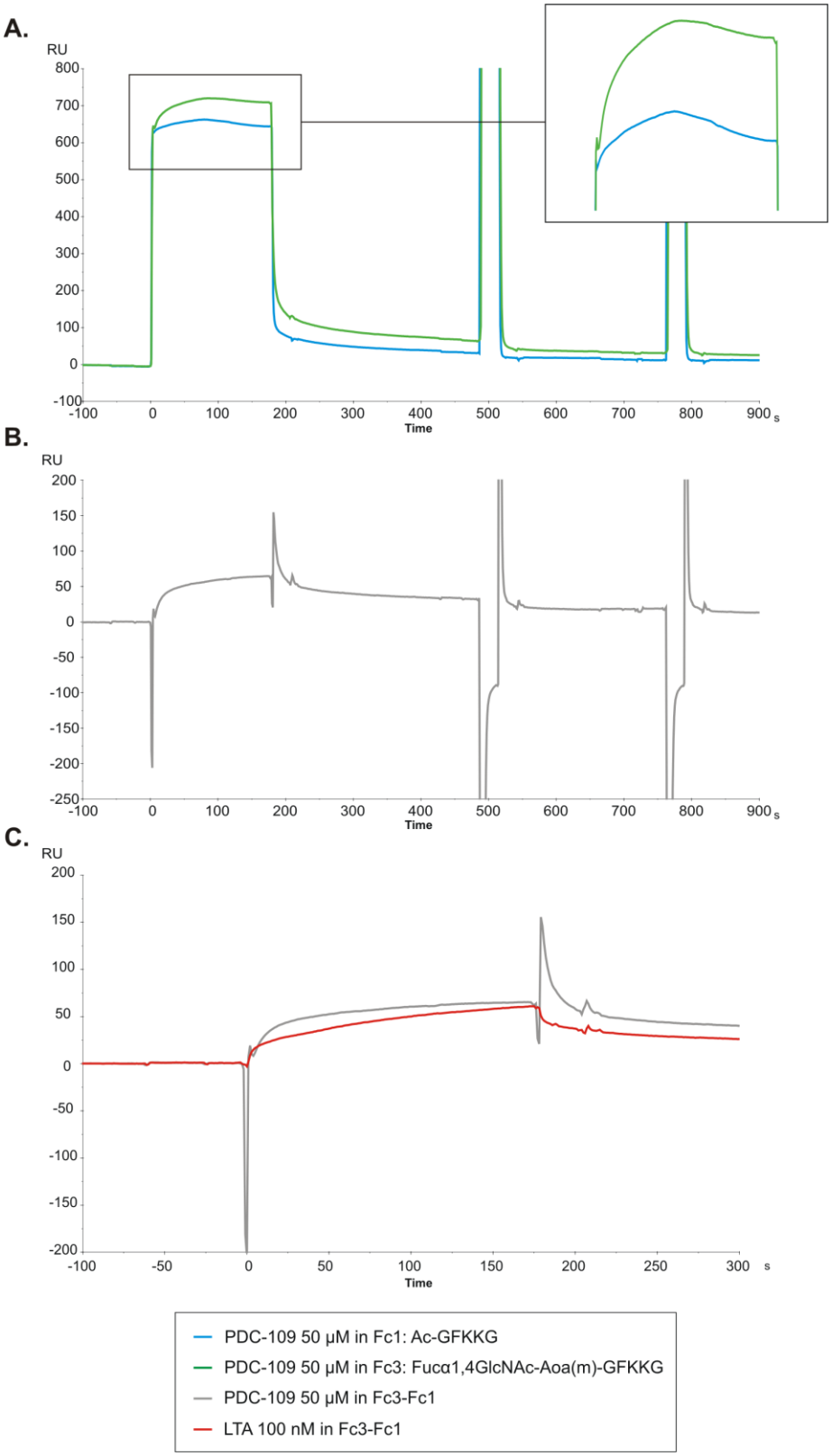


Figure 2.9. (**A**) SPR sensorgrams of the binding of PDC-109 at 50 µM to reference surface, *i.e.* Ac-GFKKG (blue trace), or to Fucα1,4GlcNAc-Aoa(m)-GFKKG glycoprobe immobilized (green trace). On the right, the unusual decreased signal in association phase is enlarged. (**B**) Differential curve for PDC-109 kinetic studies obtained after subtracting the reference channel. (**C**) Comparison of differential curves from PDC-109 (grey trace) *vs* the well-known lectin, LTA (red trace).

Thus, response increases until a high value (maximum binding) is reached, then slowly decreases even though PDC-109 is still being flown across the surface. Such biphasic behaviour during the association phase was observed not just with all immobilized glycoprobes but also with the reference surface (Figure 2.9A; green and blue trace sensorgrams, respectively), hence the differential curve resulting from reference subtraction (Figure 2.9B) displayed an irregular trace, with large spikes at the beginning and end of the injection that complicated data fitting, so that absolute values for the kinetic constants must not be judged entirely reliable, even if the established epitope preference is accurate.

This unusual biphasic profile, reflecting loss of mass from the sensor surface during association phase, was not observed with any of the commercial lectins studied earlier (Figure 2.9C), and suggests an uncharacteristic binding behaviour of PDC-109, confirmed after meticulous, painstaking evaluation of all possible variables (surface, ligand, analyte, buffer conditions, competitive reagents, solute stability, etc.) and with some precedents in other systems[213].

Plausible explanations for the unusual behavior of PDC-109, accounting for the biphasic shape of the SPR sensorgrams, would assume the existence of two different affinities, namely two competing analyte reactions, in the protein preparation. This can be due either to a single analyte with two different affinities, or to two different analytes, each with its own affinity.

Regarding the former situation, the possibility of non-specific binding between PDC-109 and the SPR surface was investigated, since the carboxymethyled dextran matrix on the sensor surface is structurally not too distant from the glycoprobe ligands, hence a weak interaction with the lectin could not be excluded. To this end, a solution of free carboxymethylated dextran (same as the CM5 chip coating) was added to the analyte solution; if the analyte had any affinity for either dextran or carboxymethyl, it would bind preferentially the free polymer over that on the sensor chip, thus reducing non-specific binding and avoiding competing reactions. This modification, however, did not have any effect on the curves, which remained biphasic in all cases. In a complementary attempt, a C1 chip lacking dextran matrix (only carboxymethyl groups directly bound to the gold surface) was tested to exclude any influence of the dextran matrix; again, the same biphasic profile could be observed during the association phase. From these results we concluded that no competing reactions with the SPR surface existed, hence

the biphasic behavior must proceed from analyte heterogeneity, whereby two analytes competed for the same ligand. If that was the case, the decreasing stretch in the association phase signal could result from a smaller molecule with lower affinity for the ligand but fast kinetics that competes with another, higher affinity interaction.

Different possibilities of analyte heterogeneity can be envisaged: i) PDC-109 may undergo conformational changes upon binding to epitope, ii) more than one isoform/glycoform is present in the PDC-109 sample, or iii) PDC-109 has different aggregation states; all these situations might produce characteristic sensorgrams with biphasic association curves, as schematized in Figure 2.10. Although previous reports suggest that PDC-109 undergoes a conformational change upon binding to PC membranes[214], the SPR technique is only fit to detect mass changes over time, therefore subtle conformational changes not resulting in mass changes will be undetectable. Hence, the former possibility could not be further pursued.

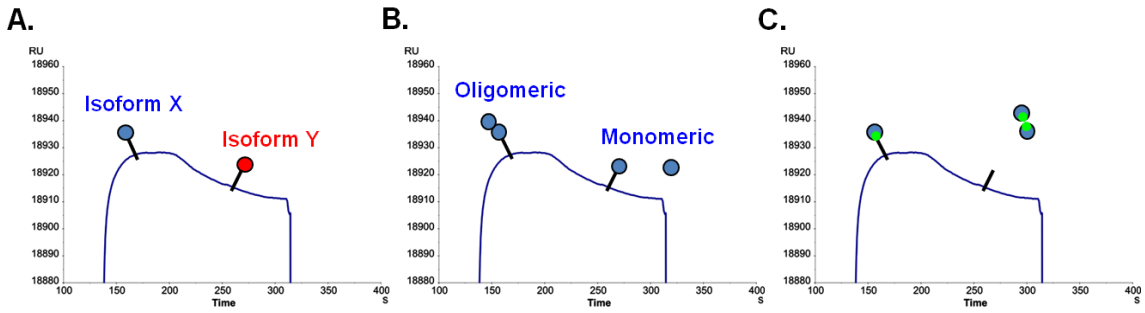


Figure 2.10. Schematic representation of different possible explanations for the biphasic behaviour of PDC-109 due to analyte heterogeneity: (**A**) Different isoforms; (**B**) Different aggregation states; (**C**) Competition from the same binding domain site for both lectin-ligand interaction and formation of protein aggregates. Ligand immobilized on the SPR surface is represented by a stick and analyte (PDC-109) as a circle.

To investigate the possibility of a heterogeneous analyte, different approaches were followed. Firstly, considering that PDC-109 exists in two major forms (*i.e.* glycosylated and deglycosylated)[175], enzymatic deglycosylation was performed in order to obtain a single analyte isoform. Given the relatively short supply of PDC-109, the amount of protein used for deglycosylation was too small to allow for a purification step; when the extent of the reaction was assessed by MALDI-TOF MS, it was shown that enzymatic glycan removal was not complete. Even so, the partially deglycosylated PDC-109 was tested on the same chips previously used with untreated protein. Although the resulting sensorgrams clearly showed a rising association phase without biphasic behaviour, the undetermined extent of glycosylation prevented formulating clear conclusions, as the

standard binding response could not be unequivocally attributed to a single isoform. To investigate the possibility of aggregation, PDC-109 was flown across a chip where the protein itself was immobilized. The resulting sensorgrams displayed the previously observed biphasic shape, supporting the possibility that interaction of PDC-109 with itself is the cause of the anomalous biphasic sensorgrams.

Evidence from the literature points out that PDC-109 is naturally produced as a mixture of several protein forms[185] and aggregation states[181] that play an important role in modulating its interaction with other biomolecules. This would allow to conclude that, whichever of the situations proposed in Figure 2.10 applies, including a combination of them, PDC-109 acts as a lectin-like molecule albeit with quite unique features that, though not abolishing its ability to recognize carbohydrates, give rise to unusual SPR sensorgrams.

Despite the above-discussed limitations, we found it worthwhile to determine the SPR binding parameters of PDC-109. Affinity constants (Table 2.5) revealed a fairly weak to moderate binding for carbohydrates, at least if compared to the considerably higher affinity ($K_A \sim 10^7$ M^{-1}) previously estimated for PC-containing membranes, also by SPR studies[213]. This seems to suggest that binding of PDC-109 to carbohydrate residues in the oviductal epithelial cells must probably involve multivalency factors in order to efficiently retain the sperm in the oviductal reservoir.

Table 2.5. Kinetic rate constants and the derived association (K_A) constant of PDC-109 to the different neo-glycoprobes prepared determined by SPR.

Ligand immobilized	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_A (M^{-1})
Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	8.62×10^1	8.60×10^{-3}	1.00×10^4
Gal-β1,3-GlcNAc-*Aoa(m)-GFKKG*	1.28×10^2	7.56×10^{-3}	1.70×10^4
Gal-β1,6-GlcNAc-*Aoa(m)-GFKKG*	4.72×10^2	2.08×10^{-2}	2.27×10^4
Man-α1,2-Man-*Aoa(m)-GFKKG*	1.02×10^3	1.30×10^{-2}	7.87×10^4
Man-α1,3-Man-*Aoa(m)-GFKKG*	3.86×10^2	4.40×10^{-3}	8.77×10^4
Man-α1,6-Man-*Aoa(m)-GFKKG*	4.10×10^2	1.73×10^{-2}	2.38×10^4
Fuc-α1,3-GlcNAc-*Aoa(m)-GFKKG*	7.63×10^2	1.65×10^{-3}	2.37×10^5
Fuc-α1,4-GlcNAc-*Aoa(m)-GFKKG*	2.91×10^3	8.83×10^{-5}	4.93×10^6
Fuc-α1,6-GlcNAc-*Aoa(m)-GFKKG*	6.45×10^2	3.45×10^{-3}	7.31×10^4
Neu5Ac-α2,3-Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	9.99×10^1	4.77×10^{-3}	1.24×10^4
Neu5Ac-α2,6-Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	1.88×10^3	2.53×10^{-3}	1.66×10^5
Neu5Ac-α2,3/6-Gal-β1,4-Glc-*Aoa(m)-GFKKG*	2.12×10^3	8.10×10^{-3}	2.77×10^4

From the above data it can also be seen that PDC-109 appears to have the highest binding affinity for Fuc-α1,4-GlcNAc among all the glycotopes in this study. This is in agreement with earlier findings that propose bull sperm binds to an oligosaccharide ligand on the oviductal epithelium that resembles Lewis^a trisaccharide (Galβ1-3[Fucα1-4]GlcNAc)[203]. For PDC-109 *vs* Fuc-α1,4-GlcNAc, the higher affinity was due to a reduction in the dissociation rate constant, consistent with a privileged interaction with this sugar.

2.3.2 **Carbohydrate recognition domain of PDC-109 by CREDEX-MS**

As seen in chapter 1, CREDEX-MS is a tool to map the interacting domains of sugar-lectin complexes, providing structural information by limited proteolysis (epitope excision/extraction) coupled with MS.

In order to test the complementarity of CREDEX-MS with the SPR-based glycoprobe approach for PDC-109, and to define the carbohydrate-binding site of this seminal plasma protein, glycotopes that in the SPR study (see section 2.3.1.3) had displayed higher affinities, *i.e.*, all fucosylated epitopes (Fucα1,3GlcNAc, Fucα1,4GlcNAc and Fucα1,6GlcNAc), were immobilized onto DVS-sepharose. Prior to the excision experiments with PDC-109, columns were validated for their Fuc-recognition ability by binding experiments with the Fuc-specific UEA-I lectin followed by SDS-PAGE detection. As done with ECA in the previous chapter, excision experiments were also carried out and the CRD of UEA-I was successfully delimited. The affinity-bound peptides eluted and identified by MS were in agreement with the previously reported structural basis of UEA-I carbohydrate specificity[210].

Once the functionality of the columns was confirmed, the excision experiments with PDC-109 were run. Briefly, PDC-109 was incubated with the Fuc-α1,(3,4,6)-GlcNAc-DVS-sepharose during 24h, columns were washed until no protein was observed by MALDI-TOF MS, then the lectin-sugar complexes were digested with trypsin at 37ºC overnight (top panel, Figure 2.11B; representative for Fucα1,4GlcNAc). Columns were again washed until no peptides were observed by MALDI-TOF MS (Figure 2.11B, middle panel), then the affinity-bound peptides were eluted with ACN:H_2O 2:1 TFA 0.1% and identified by MALDI-TOF MS (bottom panel). The mapped peptides are indicated in the primary structure of PDC-109 (Figure 2.11A).

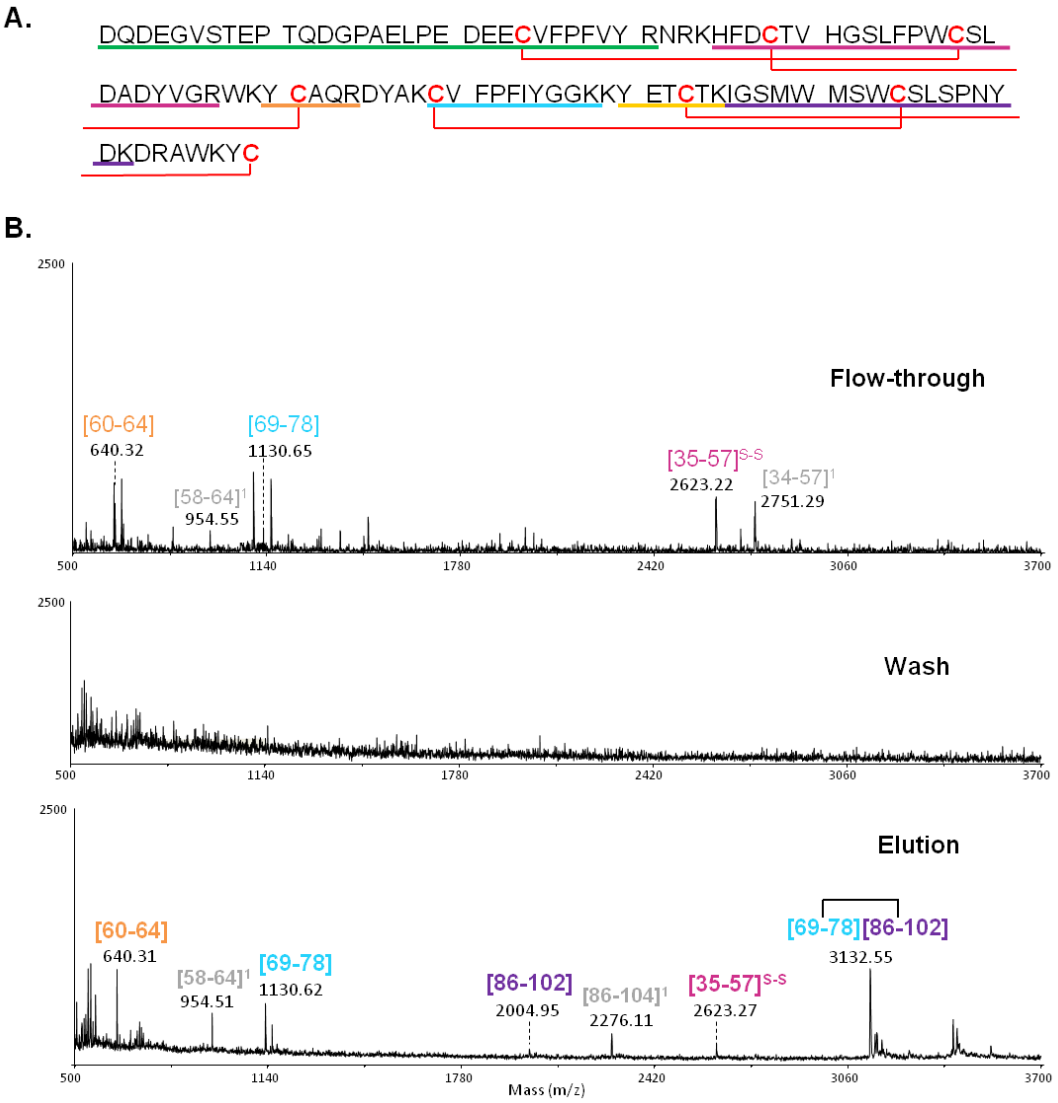


Figure 2.11. (**A**) Sequence of PDC-109; peptide fragments obtained by trypsin digestion in solution are shown in color. (**B**) MALDI-TOF MS spectra of the different fractions of PDC-109 *vs* Fucα1,4GlcNAc CREDEX-MS excision experiment.

In the Fucα1,4GlcNAc excision experiment (see Figure 2.11), five out of seven peptides (coloured) identified in the elution fraction correlated directly with the predicted ones. In addition, two other peptides (highlighted in gray) were observed in the elution fraction and corresponded to miscleaved fragments. The only peptide present in the elution fractions of excision experiments of all fucosylated glycoprobes tested was the disulfide-bridged fragment [69-78]-[86-102]. Its reduced components were also observed but this does not necessarily reflect the native binding situation. Hence it is very likely that these two peptides together, either disulfide-bound or individually, constitute the carbohydrate binding domain of PDC-109. This result with bovine seminal PDC-109 protein corroborates the viability of the CREDEX-MS approach for

elucidating carbohydrate-lectin interactions, not only at the demonstration level (*i.e.*, with "academic" lectins) but with "real" biological samples.

In summary, our SPR-based approach, using an extensive collection of well-defined glycopeptide probes exposing sugar epitopes in native-like form, has not only served to determine kinetic and affinity constants of these epitopes and specific plant lectins but also shown its applicability to PDC-109, a fertilization-relevant protein. While the binding affinity of PDC-109 for heparin[193] or PC[213,215], as well as its chaperone-like activity[216] have received considerable attention, our analysis of the lectin-like activity of PDC-109 is, to the best of our knowledge, the first quantitative study of the carbohydrate binding behavior of PDC-109.

Although several glycoconjugates have been synthesized by others laboratories using also *N*[Me]-Aoa- peptide versions as mimics of natural glycoproteins[217,218], only few of them have been used as probes for interaction studies, and not always with conclusive results. Thus, this work represents a step further to establish our SPR-approach as a serious alternative approach to study carbohydrate-lectin interactions. The limitations of the SPR methodology in this particular study of PDC-109 have nonetheless been satisfactorily bridged by the CREDEX-MS approach, again showing these techniques as a perfect duet for the detailed study of carbohydrate-lectin interactions.

CHAPTER 3:

Glycobiology of gametes and mammalian fertilization

3 Glycobiology of gametes and mammalian fertilization

3.1 INTRODUCTION

Fertilization is a fundamental event that follows a highly coordinated sequence of cellular interactions between gametes (*i.e.* the spermatozoon and the egg) in order to form a diploid zygote and, ultimately, a new individual. In mammals, fertilization occurs in the female reproductive tract. As shown in Figure 3.1, inside the vagina, spermatozoa travel through the cervix and uterus, and into the uterine horns. On reaching the oviduct, spermatozoa bind to the oviductal epithelium and are held back in the lower isthmus forming a sperm reservoir[219]. Following capacitation (term described in chapter 2), spermatozoa are released from the reservoir and travel into the oviduct. Upon encountering the oocyte, spermatozoa penetrate the cumulus complex and bind the extracellular glycoprotein matrix, known as the *zona pellucida* (ZP) that encapsulates the egg. This primary binding initiates acrosomal exocytosis from the spermatozoon resulting in the release of hydrolytic enzymes from the sperm acrosome vesicle and that, in consequence, leads to sperm penetration through the ZP and into the perivitelline space. Finally, secondary binding of the acrosome-reacted spermatozoon allows fusion with the egg plasma membrane and completes fertilization.

Upon ejaculation, millions of spermatozoa are deposited in the female reproductive tract, though only few thousands enter the oviduct, some of them reach the ampulla, and only one spermatozoon fertilizes the egg. To guarantee the encounter of the two highly specialized gametes at the right time and place, both the oviduct and oocyte coordinate sperm functions. In animals with internal fertilization, the oviduct provides the environment needed for gamete transport, final maturation, fertilization, and early embryonic development[220,221]. On the one hand, freshly ejaculated sperm cells cannot bind the ZP and need to be activated in the female genital tract via the capacitation process, in order to reach the site of fertilization and penetrate the oocyte envelopes[222,223]. On the other hand, oocytes also undergo modifications while travelling through the oviduct to the site of fertilization. Once in the oviductal ampulla and exposed to the oviductal fluid, oocytes undergo an oviductal ZP maturation that includes a pre-fertilization ZP hardening, hence contributing in particular to the control of polyspermy[224-226].

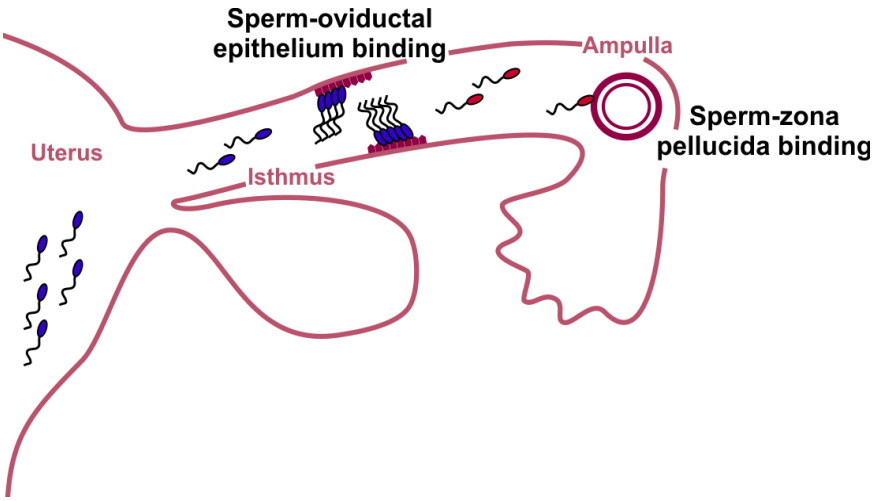


Figure 3.1. Sites of protein-carbohydrate interactions that regulate sperm function in the animal female reproductive tract. Uncapacitated sperm (blue) *vs* capacitated sperm (red).

Carbohydrates of the reproductive system have been neglected for a long time, mostly because they have been difficult to analyze. This notwithstanding, the last 10 years have witnessed major advances in the development of numerous methodologies to study the structure and function of glycans, hence our understanding of glycoproteins function (see general introduction). Today, overwhelming evidence exists that carbohydrate recognition plays a crucial role in fertilization from lower species to man[227,228] and sperm-surface carbohydrates have also been implicated in immune-mediated infertility[229]. In this context, it has long been accepted that the oviductal sperm reservoir and sperm-ZP binding are predominantly mediated by protein-carbohydrate interactions (Figure 3.1) involving carbohydrate moieties both on the oviductal epithelium and on the ZP, and lectins on the sperm cell surface. Due to the apparent evolutionary induced divergence in these binding events, the specific moieties must be identified independently in each species of interest. In fact, species specificity must reside in carbohydrate modifications on the egg surface, and in the coordinated assembly of a unique cohort of sperm proteins at capacitation[230]. Several candidate molecules possibly implicated in sperm-oviduct and sperm-egg interactions have been postulated for spermatozoa from different mammalian species[231], suggesting that distinct carbohydrate ligands and sperm lectins are involved in each one. However, there is still no consensus about the binding mechanisms and the molecules involved. Thus, although there is now a vast amount of information available on the events that occur when spermatozoa and oocyte meet, the linking pieces of the puzzle are still missing.

Identification and characterization of sperm proteins with ZP affinity are of major interest, since the precise mechanism of sperm-oocyte interaction is still unclear. Scientifically, this is of interest not only to gain understanding of the process of fertilization, which will be of utmost value in the field of assisted reproduction, but also for immuno-based contraception, where the sperm-ZP receptor may be a putative target for vaccination strategies[232].

3.1.1 Formation of the oviductal sperm reservoir is a carbohydrate-mediated event

In mammals the coitus results in the release of billions of sperm in the female reproductive tract, but only a small number of spermatozoa enter the uterine tube. Upon attachment to epithelial cells lining the caudal isthmus of the uterine tube (Figure 3.1), spermatozoa are arrested in a non-capacitated state forming a sperm reservoir[233], so that their fertile life and viability is extended until a combination of sperm maturational changes and peri-ovulatory signaling induces their detachment[234].

Sperm storage in oviductal reservoirs, a widespread reproductive strategy in eutherian mammals, has been described in mice[235], hamsters[236], sheep[237], pigs[238], cattle[239,240], horses[241], and also humans[242]. From an evolutionary perspective, pre-ovulatory binding of diverse populations of sperm cells to the oviductal epithelium forming these sperm reservoirs may have developed as a form of fine-tuning to assist in sperm selection, to synchronize completion of capacitation with the events of ovulation, hence ensuring that spermatozoa are present at the right time to fertilize the egg, and to promote monospermic fertilization through a controlled release of competent gametes.

Sperm binding to oviductal epithelium appears to involve carbohydrate recognition[243]. Several *in vivo* and *in vitro* studies have demonstrated that carbohydrates on the apical oviduct surface and lectin-like molecules on the sperm surface are involved in adhesion in a species-specific way. In hamsters and rats, sperm binding to oviductal epithelium is mediated by sialic acid[244,245] and by galactose in horses[246]. In pigs, both oligomannosyl structures[247] and 6-sialylated biantennary glycans[248] seem to be involved in sperm–oviduct binding. In cattle, strong evidence supports the involvement of fucose residues[202,203], and in llamas (camelid), N-acetylgalactosamine and galactose have been observed that inhibit the sperm binding to the oviductal cells[249]. On the other hand, less is known about molecules involved in sperm release. Around the time of ovulation, and

when sperm become capacitated, signaling from unknown factors associated with follicular fluid, oocytes and uterine tube secretion, promote the detachment of large numbers of capacitated spermatozoa with hyperactive motility that may contribute to the fertilizing pool[220]. However, the exact identity of either sperm and/or oviductal molecules that respond to these releasing signals is still unknown[201]. Knowledge of molecular mechanisms underlying sperm-oviduct interaction may advance our understanding of the behavior of sperm within the female reproductive tract and provide new tools for sperm selection, extension of fertile life, modulation of capacitation and increase reproductive efficiency in the field of artificial insemination and reproductive biotechnologies.

In the following subsections, and given that bull semen was the raw material used in this thesis to study carbohydrate-protein interactions in fertilization, specific information regarding formation of the oviductal sperm reservoir in bovine species will be further discussed. It should be noted that some of the following information is already introduced in chapter 2 (section 2.1.2.3).

3.1.1.1 *Sperm-oviduct adhesion in cattle*

Bovine sperm binding to explants of oviductal epithelium was determined to be specifically blocked by fucoidan as well as its non-sulfated component fucose[202]. Furthermore, the most effective inhibition was established when fucose was linked α1-4 to *N*-acetyl-glucosamine, as represented by the trisaccharide Lewis^{a} [203]. Thus, it was determined that molecules containing fucose, preferentially in an (α1-4) linkage to *N*-acetyl-glucosamine should be present at the surface of the oviductal epithelium and fucose-binding molecules on the surface of sperm. More specifically, employing lectin histochemistry with fucose-specific lectins UEA-I and LTA, it was demonstrated that the mucosal surface of the bovine oviduct is covered with molecules containing fucose. Pretreatment of bovine epithelium with fucosidase, but not galactosidase, reduced sperm binding[202]. In consonance, fucosylated bovine serum albumin, tagged with fluorescein as well as Lewis^{a} tagged with fluorescein-labelled polyacrylamide, reacted particularly with the rostral head regions of motile bovine sperm[203,204].

From the sperm side, the putative bovine sperm-surface lectin that recognizes fucose residues present in the oviductal epithelium has been putatively identified through Lewis^{a} affinity columns as bovine spermadhesin BSP1, also called PDC-109 or BSP-

A1/A2[189,205] (chapter 2). Moreover, BSP3 (BSP-A3) and BSP5 (BSP-30KDa) also enable sperm binding to oviductal epithelium, suggesting that maybe a redundancy of oviduct binding proteins is involved in reservoir formation to ensure reproductive success[250]. These three binder of sperm proteins, which are secreted by bovine seminal vesicles into seminal plasma and adsorbed onto sperm thus coating the sperm plasma membrane, respond in different ways to incubation of sperm under capacitating conditions[251]. Whereas BSP5 was lost from sperm and BSP1 loss was hardly detected, BSP3 was found to be cleaved on the sperm surface. Surprisingly, differences do not involve simply different rates of shedding of each type of BSP from the sperm surface. Thus, these alterations in the bull sperm surface proteins could play a role in releasing sperm from the storage reservoir by modifying sperm interactions with the oviductal epithelium, and suggest a regulated mechanism of sperm movement along the oviduct. Affinity purification of proteins extracted from oviductal apical plasma membranes using BSPs, identified the annexins ANXA1, ANXA2, ANXA4, and ANXA5 as strong candidates for the oviductal epithelium sperm receptors. Annexins contain fucose, bind with high affinity to heparin and related glycosaminoglycans, are present on the apical surface of oviductal epithelium, and antibodies to each of them block sperm-oviduct binding[252].

3.1.1.2 *Sperm-oviduct release in cattle*

Sperm must be released from the epithelium in order to ascend to the ampulla and fertilize oocytes. When bull sperm is capacitated *in vitro*, fewer bind to explants of oviductal epithelium[253] and to fucosylated bovine serum albumin[204]. Thus, changes in sperm upon capacitation, involving elimination or modification of carbohydrate-binding molecules on the sperm head, may be responsible for a loss in binding affinity for the oviductal ligand, hence explain sperm release. As bull sperm binding affinity for fucosylated molecules decreases during capacitation, suggesting that the fucose-binding molecule is either lost or modified, hyperactivated sperm motility could provide increased pulling force to aid in sperm detachment.

Although it appears that the oviduct epithelium does not release sperm by reducing the number of carbohydrate structures[240], it could cause release by secreting initiators of capacitation and/or hyperactivation. Soluble oviduct factors have been shown to enhance capacitation of bull sperm[254,255]. Thus, signals associated with impending

ovulation, such as rising hormone levels, could stimulate the secretion of capacitation and hyperactivation inducers that release sperm at the appropriate time for fertilization[220]. In fact, oviductal fluid molecules such as sulfated glycoconjugates[256] and disulfide-reductants[257,258] are able to release bovine sperm bound to the oviductal epithelium *in vitro* through the reduction of sperm surface protein disulfides to thiols[258], and trigger capacitation-related changes[257]. These results led to the hypothesis that conformational changes, driven by SH-SS exchange of specific redox-sensitive sperm surface proteins directly or indirectly involved in adhesion, might regulate the affinity of sperm for the oviductal epithelium. Therefore, sperm release from the oviductal epithelium is putatively mediated by a combination of factors secreted by the oviduct itself, that induce sperm capacitation during the estrus cycle and alteration of carbohydrate binding activities on the sperm surface.

3.1.2 **Carbohydrates are the signals for gamete recognition**

Following capacitation and release from the reservoir in the female reproductive tract, mammalian sperm travel a tortuous path and are eventually lured toward the egg[259,260]. After penetrating the mass of cumulus cells surrounding the egg[261] and gaining access to the egg surface, the remaining steps of fertilization can ensue. In mammals, sperm–egg interaction can be subdivided into six steps (Figure 3.2): acrosome-intact sperm first bind to the ZP (step 1), a thick extracellular matrix comprising the outer layer of the egg; interaction with the ZP signals sperm to undergo the acrosome reaction (step 2), a regulated exocytosis that is essential for subsequent steps; sperm penetrate the ZP (step 3) and, having done so, bind to (step 4) and fusion with (step 5) the egg plasma membrane; these latter events somehow "activate" the egg (step 6) to initiate zygotic development and to become refractory to additional sperm binding and penetration[262]. Thus, to deliver its haploid genome, a sperm must penetrate a first barrier imposed by a wall of cumulus cells, and then it must bind to and breach two barriers imposed by the proper egg, the ZP and the plasma membrane. Finally, the fusion of the two haploid cells (gametes) forms a diploid zygote, which is the functional definition of fertilization, resulting in the creation of new life.

These exquisite cell- and species-specific recognition events are among the strategically most important cellular interactions in biology. As already mentioned, carbohydrate recognition is also implicated in this primary binding between sperm and ZP.

Substantial evidence indicates that sperm express specific lectin-like proteins on their surface that bind to restricted oligosaccharide sequences present at glycoproteins of the egg outer surface. However, the carbohydrates involved in ZP binding must differ from those involved in binding to oviductal epithelium, because sperm lose binding affinity for the oviductal carbohydrate ligand when they are capacitated and thus ready to bind ZP. Despite the fundamental importance of gamete interaction, the precise molecular mechanisms that underpin and regulate this complex event remain to be fully elucidated. Such knowledge will contribute to the development of novel contraceptives for fertility control as well as novel diagnostic and therapeutic strategies for male infertility.

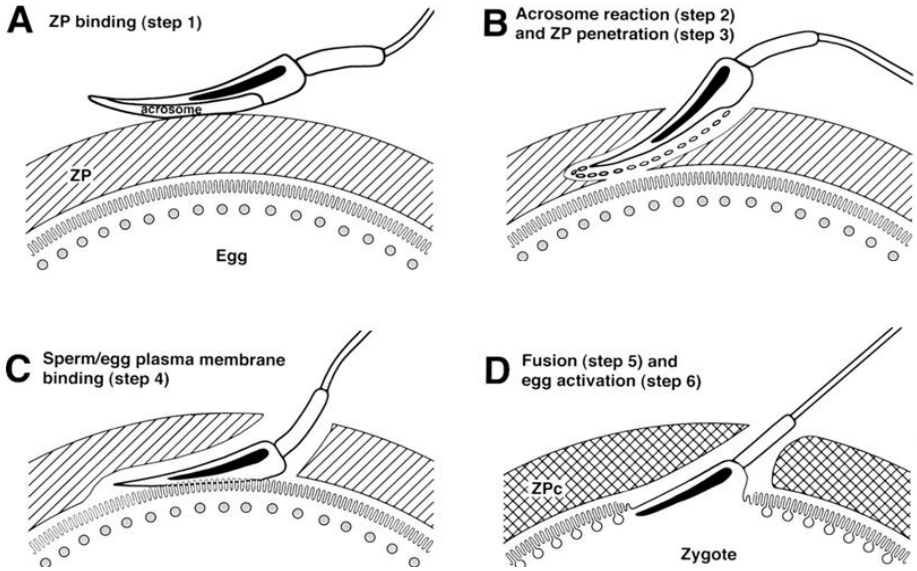


Figure 3.2. A basic scheme for the mammalian gamete interaction leading to fertilization: (**A**) Sperm bind to ZP. (**B**) Sperm undergo the acrosome reaction and penetrate ZP. (**C**) Sperm bind to the egg plasma membrane. (**D**) Sperm fuse with the egg plasma membrane and activate zygotic development. Reproduced from [263]. Copyright © 1996 by Cell Press with permission from Elsevier.

The first contact between the oocyte and the sperm cell is the binding of the latter to the ZP; ZP is therefore the target of intense investigation in all mammalian models. This specialized extracellular matrix not only plays pivotal roles in gamete recognition, fusion and blockage of polyspermy, but also serves to protect the early embryo during pre-implantation development and mediates immune recognition.

The ZP comprises three or four major glycoproteins (ZPGs). In many mammalian species, there are four ZPGs (ZP1 to ZP4)[264], although some species comprise three, *e.g.*, the mouse (ZP1 to ZP3), or swine and cattle (ZP2 to ZP4). In mice, ZP3 alone

binds acrosome-intact sperm[265], while in pigs and cattle a ZP3/ZP4 heterocomplex is required to bind acrosome-intact sperm[266,267], ZP4 being responsible for the binding activity in pigs[268]. A recent study showed that ZP2 alone can bind to acrosome-intact sperm in humans[269]. Thus, sperm–ZP binding mechanisms may be divergent between mammals, rendering it a highly selective and carefully regulated process that serves as an inter-species barrier to fertilization by preventing adherence of non-homologous sperm to eggs. Over the last decades, investigators seeking to understand ZP role in gamete recognition have proposed models based on biochemistry, cell biology and, more recently, mouse genetics. Depending on the vertebrate model, each component of the ZP matrix has been proposed as a ligand for sperm binding, which have been summarized in recent reviews[270,271].

Studies of murine sperm–ZP binding are arguably the most advanced among mammals. In mice, the carbohydrate chains of ZP3, specifically O-linked glycans attached to Ser332 and Ser334, were shown to be essential for sperm binding. Following fertilization, these glycan ligands would be modified by a putative glycosidase exocytosed from egg cortical granules to prevent sperm binding to 2-cell embryos[265,272,273]. Despite such compelling evidence, this classical glycan model has been increasingly questioned by studies using transgenic mice[269,274] which have proposed alternative models of gamete recognition including: (i) a supra-molecular model in which initial sperm-egg binding depends on the interaction of sperm with a complex of the three ZP glycoproteins [275], and following fertilization, ovastacin[276] is released from egg cortical granules and cleaves extracellular ZP2 effectively destroying the sperm-binding domain and preventing further sperm adhesion; (ii) a hybrid model that incorporates elements of both the glycan model and the supra-molecular model by proposing that sperm bind to an O-glycan that is conjugated to ZP3 at a site other than Ser332 or Ser334, and that sperm access to this glycan is regulated by the proteolytic cleavage state of ZP2[277]; and finally (iii) a domain specific model that envisages a dual adhesion system in which sperm protein(s) interact with the glycans and/or the protein backbone of ZP3 depending on its glycosylation state[278]. The evidence in support of each of these models of gamete interaction has been reviewed in depth recently[279-281].

What is clear from the current putative models of sperm-ZP binding is that the initiation of gamete interaction is not mediated by a simple lock-and-key mechanism involving a single receptor-ligand interaction. Rather it is likely that sperm engage in multiple

binding events with a variety of ligands within the ZP matrix, with both protein and carbohydrate moieties in the domain(s) of ZPGs involved in gamete recognition. An advantage of this complex adhesion system is that it would enhance the opportunities of sperm to bind to the oocyte and thus maximize the chance of fertilization. It may also account for the myriad of diverse candidate molecules that have been proposed and constantly being refined as putative sperm surface receptors in this initial egg recognition event[231,263,279,282]. Moreover, consistent with the notion that primary sperm-ZP interaction involves engagement with specific carbohydrate structures on ZP3, a number of the identified sperm receptors possess lectin-like affinity for specific sugar residues.

As done before, in the following subsection specific information concerning the amazing complexity of gamete interaction in bovine species will be also underscored.

3.1.2.1 *Gamete binding in the bovine model*

Cows are another useful mammalian model for investigating gamete binding. Definitive isolation of bovine ZP (bZP) was first reported in 1988, and subsequent analysis revealed that it is also composed of three major glycoproteins designated ZP2, ZP3, and ZP4[283]. Amino acid sequences of ZP2-ZP4 were determined by expression cloning[284] and bZP-associated N-glycans were also well characterized, with 23% remarkably consisting of a single neutral, high mannose type N-glycan ($Man_5GlcNAc_2$); and the remainder representing biantennary, triantennary, and tetraantennary sialylated N-glycan chains with some antennae extended with linear polylactosamine type sequences[285].

There is evidence that mannose Man residues associated to neutral N-glycans of bZPs could be involved in sperm-ZP binding[286]. However, 77% of bZP glycans are acidic chains containing Sia, and the number of such residues decreases significantly in fertilized oocytes, suggesting a role for Sia at some level of the fertilization process[285]. Furthermore, pretreatment of either gamete with GlcNAc significantly reduced bovine sperm-ZP attachment, also hinting at a GlcNAc role in a glycan sequence on the ZP surface crucial for gamete interaction[287]. Recently, the role of Sia in bovine sperm-ZP binding was confirmed through *in vitro* competition assays using different lectins, antibodies and glycosidases[288]. Specifically, Sia at the non-reducing ends of acidic N-linked and/or O-linked chains of bovine ZP glycans, with Neu5Ac(α2-3)Gal(β1-

4)GlcNAc as the minimal sequence, were involved in sperm-ZP interaction. Interestingly, this same sequence is contained in Sialyl Lewis^{x}, a well-known selectin ligand and the most abundant terminal sequence on N- and O-glycans of human ZP which represents the major carbohydrate ligand for human sperm-egg binding[289].

Evidences for a role of Sia in bovine sperm-ZP interaction, along with previous results implicating also Man or GlcNAc monosaccharides, would be in agreement with the observation that a ZP3/ZP4 heterocomplex is required for bull spermatozoa interaction[267] as well as with the above-mentioned putative models that indicate that gamete interaction could be mediated by a multiple complex involving both several sperm plasma membrane proteins and ZP carbohydrates.

The conserved structure and simple composition of mammalian ZP has facilitated identification of surface components that interact with spermatozoa regardless of the paucity of material available for research, especially in humans. In contrast, it has been postulated that the sperm receptors for the ZP are not conserved and that, in fact, such variation drives speciation. However, the complexity of protein expression patterns in sperm plasma membrane, the regional specificities of sperm protein distributions, and the difficulties in isolating plasma membrane fractions free of cytoplasmic contamination as well as purifying proteins biochemically without degradation, have made the identification of components binding to ZP glycoproteins a daunting task. Furthermore, research in this area has often suffered from biased approaches, focused on subjectively chosen proteins despite the fact that sperm-ZP interactions are known to be functionally complex and likely to involve different sperm proteins.

Several major approaches have been employed in the identification of sperm surface components exhibiting ZP binding activity, including: (i) immobilized, size-separated sperm proteins probed with radiolabelled or biotinylated ZP glycoproteins; (ii) sperm membrane components obtained by detergent solubilization, fractionated by a variety of chromatographic techniques and then tested for their ability to block sperm–ZP binding; (iii) affinity chromatography of detergent-extracted spermatozoa on ZP glycoprotein columns; (iv) polyclonal and monoclonal antisera inhibiting sperm–ZP binding; (v) cross-linking of sperm surface proteins to ZP ligands; and, in humans, (vi) correlation with the infertile state. As a result, a relatively large number of sperm surface proteins have been identified by these techniques and implicated in these initial egg recognition events, even though their characterization is still very incomplete. Significant data on

the bovine model is summarized in a list of the best characterized sperm receptor candidates (Table 3.1).

Table 3.1. Putative bovine sperm-*zona pellucida* receptor candidates.

Candidate sperm receptor	Evidence	References
Fertilin (α and β), a disintegrin and metalloproteinase (ADAM)	Fertilin α/β heterodimer may be present on the sperm surface as a higher-order oligomer and bovine fertilin β (ADAM2) disintegrin loop is involved in sperm-egg binding.	[290]
Fertization antigen 1 (FA-1)	Dose-dependent reduction effect of FA-1 mAb on fertilization rates of bovine oocytes.	[291]
β1,4-Galactosyltransferase (GalTase)	Selective localization of GalTase to the sperm plasma membrane suggests that it may serve as a generalized gamete receptor.	[292]
Sperm adhesion molecule 1 (SPAM1)	Widely conserved sperm surface protein. Significantly increases the ability of sperm to penetrate the cumulus of oocytes via hyaluronidase activity. Marker of sperm maturation and fertilizing ability.	[293, 294]
Spermadhesins (aSFP)	Major components of seminal plasma. Associate peripherally with the sperm surface and possess carbohydrate-binding domains.	[295]
Zonadhesin (ZAN)	Sperm membrane protein containing multiple cell adhesion molecule-like domains that binds in a species-specific manner to ZP matrix.	[296, 297]

While care has been taken to ensure that the major bovine candidates with strongest evidence for primary ZP binding are included, this table may still be incomplete. Please note that some candidates have also been characterized in other species.

Data indicating that both protein-carbohydrate and protein-protein interactions may play a role in gamete binding has highlighted the complexity of the situation and precluded the classical lock-and-key model that prevailed in this field for several decades. Although the biochemical basis of this multifaceted sperm-ZP adhesion process remains obscure, it is unlikely that it could be regulated by the activity of a single receptor, not to forget that, as stated previously, myriad molecules have been proposed as putative sperm-ZP binding mediators, underscoring the amazing complexity of gamete interactions and suggesting the coordinated action of several ZP receptor molecules assembled into functional multimeric protein complex(es)[298]. Multiple proteins recognizing different ZP ligands in defined spatial relationships to each provide a wider alphabet for species specificity than a single lectin. The proteomic profiling and functional characterization of these multiple sperm receptors and/or additional

multiprotein complexes promises to shed new light on the intricacies of sperm-egg interactions.

3.1.3 **Sperm cell proteomics**

3.1.3.1 *Proteomics technology and challenges*

Proteomics, *i.e.* the study of protein products expressed by the genome, has become one of the leading technologies employed by researchers in the postgenomic era due to the central role of proteins and protein–protein interactions in cellular function. After genomics and transcriptomics, proteomics is the next step in the study of biological systems. It is more complicated than genomics because an organism's genome is more or less constant, whereas the proteome differs from cell to cell and from time to time. Therefore it is critical to study proteins directly, and the comprehensive and systematic identification, quantification, and characterization of proteins expressed in cells are fundamental goals to gain new insights into cellular function.

Although by no means the only technique to be considered in proteomics studies, mass spectrometry (MS) is the major enabling technology for protein identification and localization of post-translational modification (PTM) sites[299]. In addition to sample preparation, consideration must be given to protein complexity, concentration and potential contamination in order to select the most appropriate MS instrumentation for downstream analysis. The different types of mass spectrometers use specific devices for ionization, mass analysis, fragmentation and detection, each with particular assets making it suitable for a given proteomic study. Over the past decade, MS instruments have undergone technical developments that now allow to achieve truly amazing levels of sensitivity, resolution and accuracy, as well as different sorts of experimental workflows.

Notwithstanding instrumental capabilities, the ability to draw meaningful biological conclusions from any proteomics experiment depends entirely upon an unbiased interpretation of the data. In this sense, the complexity of the proteomics process and the large number of spectra obtained in a given MS experiment are often used to justify complete confidence on software (bioinformatics) for sequence assignments, altough it occasionally may lead to unreliable conclusions. For example, the database searching tool Mascot[300], extensively used for the protein identification, works by extracting the

most abundant peptide *m/z* values from an experimental MS spectrum, comparing them with values generated from the *in silico* digestion of proteins contained within a user-selected database, and then assigning a protein or proteins to each spectrum on the basis of the matched peptide *m/z* values. Although the manner in which Mascot carries out these steps is proprietary, it uses the probability-based MOWSE algorithm[301] to associate a significance threshold with each protein assignment such that a sequence with a score above this threshold has a low probability of being assigned by chance. When the goal is not simply protein identification but localization of specific PTM sites, the primary amino acid sequence is required. Several other search engines and data mining tools based on similar principles are available; however, results often depend substantially on the search mechanisms employed.

Altogether, drawing meaningful conclusions from a proteomics experiment requires the successful integration of the diverse areas of sample sourcing/preparation, mass spectrometric analysis, data interpretation and determination of biological significance[302]. However, this is not a trivial task and there are limits to what can and cannot be achieved. Indeed, the major challenge of proteomics research is that it still really struggles to deal with the high degree of protein complexity and huge dynamic range of proteins (*i.e.* the difference in abundance between the most abundant detectable proteins and the least abundant ones) expressed in the complex biological mixtures[303], which exceeds six orders of magnitude in cells and ten orders of magnitude in body fluids. In MS technology, despite the high speed and mass accuracy of the most sophisticated machines, selection of peptide ions for fragmentation is in most cases still based on abundance, leaving a large proportion of ions unfragmented. This means that fast-changing, low-abundance proteins which may be most relevant will often remain unidentified. Fortunately, software developers and MS instrument developers are joining efforts to overcome such important limitations by way of bioinformatics.

Studying the mammalian sperm proteome is in some respects simpler than that of somatic cells, as many somatic cell features have been lost. The sperm cell is highly polarized and specialized, with a minimal amount of cytosol and organelles (no endoplasmic reticulum, Golgi apparatus, lysosomes, peroxisomes or ribosomes), and therefore mature sperm has lost the potential for gene expression and protein synthesis. Sperm are transcriptionally and translationally silent after leaving the testes, and their post-testicular development/maturation is controlled almost exclusively by the addition

of exogenous proteins in the female reproductive tract or by post-translational modification of their intrinsic protein complement. However, even with this reduction in complexity and probably dynamic range of protein abundance compared with other cell types, other factors compound to make the study of sperm proteins still quite challenging. Thus, the sperm cell has high membrane content and so relatively more membrane proteins than many cell types. Also, a single ejaculate can contain more than 100 million sperm cells at different stages of development and in various states of activity. In addition, these cells are transported in an active carrier and exchange extensively with both male and female reproductive tract. Hence the heterogeneity of sperm cells within a single ejaculate creates incredible variation. Furthermore, while the vast majority of cells present in a normal ejaculate are spermatozoa, a variable proportion of other cells such as immature spermatids, aberrant spermatozoa, somatic cells (leukocytes and male tract epithelial cells) and bacteria are also present. It is very important to take this fact into account and to implement appropriate methods to eliminate contaminating cells[304].

In recent years, knowledge on different aspects of sperm proteomics has been excellently reviewed[305-307], and studies have provided a better understanding of protein functions involved in processes such as sperm motility, capacitation, acrosome reaction, and fertilization. For instance, studies have shown how PTMs such as phosphorylation, glycosylation, proteolytic cleavage and mutations can bring about physiological changes in sperm function. Also, proteome analysis has allowed the study of spermatozoa in different functional states (immature *vs* mature, uncapacitated *vs* capacitated, normal *vs* defective), as well as provided insights into sperm function-dysfunction leading to the development of novel biomarkers for detection of disease states, genetic abnormalities, and risk factors in male infertility.

Sperm proteomics represents an emerging research area that synergizes the enormous technological advances in MS (and associated computational hardware/software) with the power of developmental genetics and functional genomics. The ultimate goal is to use all these tools to probe the entire functional landscape and evolutionary history of spermatozoa, eventually generating a systems-level understanding of sperm form and function. Moreover, the highly specific and constrained function of sperm make it an ideal cell type in which to identify proteins with pleiotropic functions in other tissues that may eventually provide insights on how proteins and protein functionality arose

during evolution. As we move forward, the greatest challenge in the field of sperm proteomics will be ensuring that non-proteomic specialists fully understand the complexity of the process and adequately validate and correctly interpret its results.

3.1.3.2 *Sperm surface proteins*

In order to understand initial gamete recognition events at fertilization (oviduct adhesion, capacitation, zona binding and acrosomal exocytosis), it is imperative to study specifically the sperm surface proteome[308]. Within this context, it is also important to understand the milieu of the sperm cell during transit from the testis to the oviduct, as proteins (or other entities) from the genital tract epithelia and fluids may also affect the composition and organization of proteins on the sperm surface. Therefore, it may be also very useful to consider the proteomes of the fluid or epithelium being investigated in order to assess the effects on the sperm surface proteome.

Historically, sperm surface proteins have been studied using labeling strategies with membrane impermeable tags to facilitate enrichment and identification. For instance, biotin tags can be covalently bound to surface proteins which are then run on a streptavidin-immobilized affinity column to isolate the biotinylated proteins which, after isolation, can have the tag enzymatically cleaved[309,310]. These approaches are not completely 'plasma membrane proof' as some intracellular proteins may also be labelled due to damaged cells or cells that deteriorate during preparation and experimentation. Moreover, sperm contains a small number of endogenously biotinylated proteins. Finally, non-labelled proteins may interact and thus co-purify with biotinylated ones.

As an alternative to surface labelling strategies, sperm cell plasma membrane fractions can be purified. To achieve this, specific disruption methods such as ultrasonication and nitrogen cavitation have been designed. Other techniques including detergent extraction, homogenization and hypotonic shock had previously been employed to isolate sperm membrane fragments. However, some techniques give low purifications and less defined membrane fractions, or involve various treatments that can denature the membrane proteins, inhibit enzyme activity, or affect the functional integrity of the sperm plasma membranes. After sperm disruption, differential centrifugation techniques need to be employed to isolate membrane protein fractions from insoluble cellular debris and soluble components. The researcher needs to consider whether the disruption

method as well as the isolation protocol is really delivering sperm plasma membrane or also intracellular membranes. This is especially relevant for proteins involved in ZP recognition. If the plasma membrane preparation also contains acrosomal contamination one can be sure that secondary (intra-acrosomal) ZP-binding proteins will be identified, which may overwhelm primary (plasma membrane) ZP-binding proteins[311].

An alternative to subcellular fractionation is enrichment in protein types from a whole cell lysate. For instance, sperm proteomic studies have investigated phosphorylation, known to be very relevant to various aspects of sperm function, including epididymal maturation[312] and capacitation[313]. Phosphoproteomic studies are generally performed using affinity-based approaches where enrichment in phosphorylated peptides is achieved on immobilized metal ion (or TiO_2) columns. Another PTM, S-nitrosylation, has been characterized in humans using a biotin switch assay for protein enrichment that provided novel insights on the role of nitric oxide in capacitation[314]. It is also possible to combine subcellular fractionation and protein enrichment. The best example in sperm is the use of nitrogen cavitation to produce a cytosolic fraction of bull sperm together with protein enrichment (affinity chromatography with poly-Glu: Tyr) to enable the isolation and identification of four tyrosine kinases that were specifically localized to the cell cytosol[315].

Use of immobilized lectins is another method to isolate and/or enrich surface proteins. Lectins can bind to specific sugar residues at the extracellular domain of integral membrane glycoproteins, with some marker lectins exclusively binding to sperm plasma membrane glycoconjugates. Therefore, affinity chromatography using immobilized lectins can be used to extract surface proteins[316] and this method can also be employed on nitrogen cavitated and solubilized sperm plasma membrane material. Finally, sperm head plasma membrane proteins with high primary binding affinity for the ZP were specifically isolated in bioaffinity assays that resemble as far as possible the physiological sperm-ZP interaction using ZP fragment columns[311,317].

3.1.3.3 *Bovine model considerations*

Fertilization is far more easily studied in aquatic organisms such as sea urchin and amphibians than in mammals. However, considerable progress has been made in understanding the molecular basis of fertilization in mammals, being the system involving sperm-egg interaction in the mouse, the most thoroughly studied.

In the human model there are some intrinsic limitations. A key issue is sperm cell heterogeneity. For instance, male semen displays a higher content of abnormal sperm, hence complicating further studies. Moreover, human oocytes in optimal circumstances are of limited access, compared with oocytes from species collected in slaughterhouses, for instance. Non-viable human oocytes are obtained under informed consent either after they have failed to fertilize during *in vitro* fertilization procedures or from ovaries removed from cadavers.

Other mammalian models, *e.g.*, porcine and bovine, have been extensively studied. They are attractive because of the relatively large amounts of both sperm and eggs that can be obtained compared to other mammalian models. Moreover, bovine sperm are an ideal model for male fertility in mammals because of advantages such as good breeding records, fertility data, and progeny records. Thus, proteins from bovine studies could be starting points for further research to understand male infertility in mammalian species, including humans. Moreover, inter-species comparison of proteins and domains may provide insights on essential conserved functions, variability, and evolution of sperm proteins. Also, in animals of agricultural importance, it has important industrial applications such as the development of novel biomarkers of sperm function, candidates for fertility control and methods of fertility diagnosis resulting in cheaper, less invasive reproductive technologies[318].

3.2 OBJECTIVES

In previous chapters we have discussed the preparation of well-defined glycoprobes, either attached to a peptide scaffold or to a carbohydrate polymer, for the study of carbohydrate-protein interactions by two complementary approaches (*i.e.* SPR and CREDEX) using purified lectins. After completing this preliminary phase with reasonable success, we are ready to apply these methodologies to analyze carbohydrate-driven interactions in more complex systems.

In this context, our group has been involved in a research project on molecular aspects of reproduction (in collaboration with Dr. M. Avilés, School of Veterinary of Murcia University) where the above mentioned methodologies are applied to identify carbohydrate-binding structures of spermatozoids during gamete interaction in bovines. After an initial finding that the first step in the interaction between mature sperm cells and oocytes is mediated by sialic acid residues in the glycans of ZP glycoproteins, discovery of the complementary protein/s was set as a main objective of this thesis.

Thus, work in this chapter is focused on two important regulatory steps, both mediated predominately by protein-carbohydrate interactions, in the journey of spermatozoa towards the egg in bovine female tract (see introduction; section 3.1), namely:

- Formation of the oviductal sperm reservoir in the oviductal epithelium.
- Gamete recognition [oocyte (ZP)-sperm interaction].

3.3 RESULTS AND DISCUSSION

3.3.1 **Sperm sample preparation**

Fresh bull semen was treated for selection of motile sperm and separation from seminal fluid by "swim-up" sperm washing, both in capacitating and non-capacitating media.

Sperm capacitation can be mimicked *in vitro* by using specific media that resemble the ionic and nutritional composition of the oviduct. Specifically, capacitation was induced by incubation of spermatozoa in Tyrode albumin lactated pyruvate (TALP) medium according to the procedure of Parrish *et al*.[172], but with minor changes in its composition. As the standard Tyrode medium contains 0.6% bovine serum albumin (BSA), which could represent a handicap in terms of subsequent MS analysis, an adjusted TALP medium with lower (0.06%) BSA was tested for sperm capacitation. The effects of this new medium in sperm appearance and motility were evaluated by electron microscopy, with similar outcomes for both BSA concentrations suggesting that a ten-fold decrease in BSA concentration is not a limiting factor in sperm capacitation[319]. Therefore, in our experiments with TALP capacitating medium (CM), 0.6 mg/mL BSA was chosen. On the other hand, the non-capacitating medium was essentially the complete medium mentioned above, but without albumin, sodium bicarbonate and ionic calcium (Ca^{2+}), as it has been shown that their removal prevents or reduces capacitation[156,320,321]. Glucose has been reported to significantly delay bull sperm capacitation, possibly by inhibiting a rise of intracellular pH[322]. Therefore, as a second source of non-capacitated material, sperm cells were incubated in the presence of glucose in a TALP medium devoid of calcium, bicarbonate and BSA (termed NCM)[323].

Once sperm was treated under both conditions, total sperm count was calculated using an improved Neubauer hemacytometer and only intact, viable and morphologically normal live sperm were used.

A biochemically defined sperm-binding assay that only detects primary sperm-ZP binding (at the level of the sperm plasma membrane) but efficiently eliminates putative secondary sperm-ZP (*i.e.* by acrosomal matrix proteins) is highly required for gamete interaction research. Although proteins have been traditionally solubilized from entire cells to produce whole cell lysates with no information on subcellular localization,

current practice strongly recommends subcellular fractionation strategies providing localization information[308]. However, obtaining intact high-purity, homogeneous, specific cellular fractions with no contamination with, *e.g.*, purified sperm apical plasma membranes is technically challenging, time-consuming, and difficult to reproduce[311]. Moreover, some techniques might not produce isolates that exactly mirror *in vivo* situations, because there is always the possibility of protein redistribution during the subcellular fractionation or they also can abolish the interaction of receptor and ligand due to treatments that interfere with the native state of the receptor and/or ligand. Protein solubilization is also a critical step in these strategies, since standard methods for extracting proteins from cells are usually non-specific and often fail to solubilize all proteins, or all proteins to the same extend. Therefore, in our novel approach entire sperm cells rather than solubilized sperm proteins, whole cell lysate, or subcellular fractions were used, to eliminate uncertainty related to the treatments as well as to preserve as much as possible the native 3D structure, conformation and aggregation state of surface sperm proteins. The importance of 3D structure and/or multimeric recognition complex assemblies for ZP binding[266,298] and subsequent induction of the acrosome reaction[324,325] has previously been recognized.

3.3.2 Isolation of carbohydrate binding proteins (CBPs) from bull spermatozoa by CREDEX-MS

As mentioned in the introduction, binding to both the oviductal sperm reservoir and to ZP is predominately mediated by carbohydrate moieties on the oviductal epithelium and ZP serving as receptors for complementary lectin-like proteins on the sperm cell surface. Specifically, it has been demonstrated that bovine sperm binding to oviductal epithelium involves fucose recognition[202,203] (Figure 3.3A) and that, following capacitation, spermatozoa are released from the reservoir and proceed further into the oviduct to meet the oocyte, where gamete interaction is mediated via sialic acid residues in the oocytes[288] (Figure 3.3B). However, further investigation is required to identify the specific glycoconjugates and lectins involved in these processes.

In order to study the two important carbohydrate-driven regulatory steps in fertilization (see 3.2 Objectives), CREDEX-MS binding experiments were performed, as SPR was not viable with intact sperm due to size restriction (sperm size of 55 µm *vs* SPR microfluidics system of 5 µm) and sample complexity.

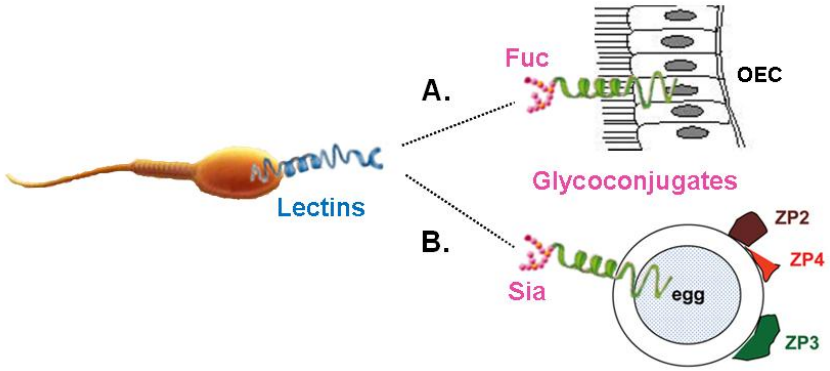


Figure 3.3. Fertilization events with protein-carbohydrate interactions involved. (**A**) Formation of the sperm reservoir in the oviductal epithelium. (**B**) Gamete recognition. The glycans (pink spheres) have been partially identified but complementary carbohydrate-binding agents in sperm (blue structure) are still unknown.

As a broad sampling of various carbohydrate structures, different affinity columns bearing specific epitopes were prepared and intact live bull sperm cells were loaded under both capacitating and non-capacitating conditions. With such an approach we aimed to mimick the *in vivo* recognition events between live sperm membrane-associated components and glycotopes present in ZP or oviductal epithelium glycoproteins. Our setup also reduced sample complexity by enriching for protein types of interest from a whole cell. Specifically, only sperm surface proteins with affinity for the homologous specific carbohydrate exposed were targeted, hence the challenges posed by the large intrinsic dynamic range of protein abundances could be faced.

Owing to sperm sample complexity, different issues had to be considered prior to CREDEX-MS experiments. Firstly, the mechanical aspects of the chromatographic system were addressed. Sperm size-related retention was tested for both filters and sepharose beads and an evaluation of sperm loss at each step of the protocol was conducted. On average, 20-30% of sperm loaded in CREDEX experiments was estimated to be retained in the chromatographic support. Although the estimation was by simple, relatively inaccurate, visual sperm count in a hemacytometer; even at this qualitative level, the loss was considered appropriate and accounted for in the different CREDEX-MS experiments.

The carbohydrate structures to be immobilized onto divinylsulfonyl-activated sepharose microcolumns were chosen according to literature and their role in some of the above mentioned carbohydrate-regulated steps[202,203,288,289] (Figure 3.4). For each carbohydrate-regulated event, two different glycoprobes were studied; one slightly more complex

than the other (di- *vs* trisaccharides), but both with the same functional monosaccharide (Fuc or Neu5Ac) allegedly involved in lectin recognition. With these presenting glycotopes, a comparison could be conducted on the length of the exposed carbohydrate and its effect on core sugar recognition by lectins.

Carbohydrate- immobilized	Carbohydrate-regulated event studied
Fucα1-4GlcNAc **(Fuc4N)**	Oviductal sperm reservoir formation
Galβ1-3GlcNAc (4 | Fucα1) **(Le^a)**	
Neu5Acα2-3Galβ1-4GlcNAc **(3'-SLN)**	Gamete recognition
Neu5Acα2-3Galβ1-4GlcNAc (3 | Fucα1) **(SLe^x)**	

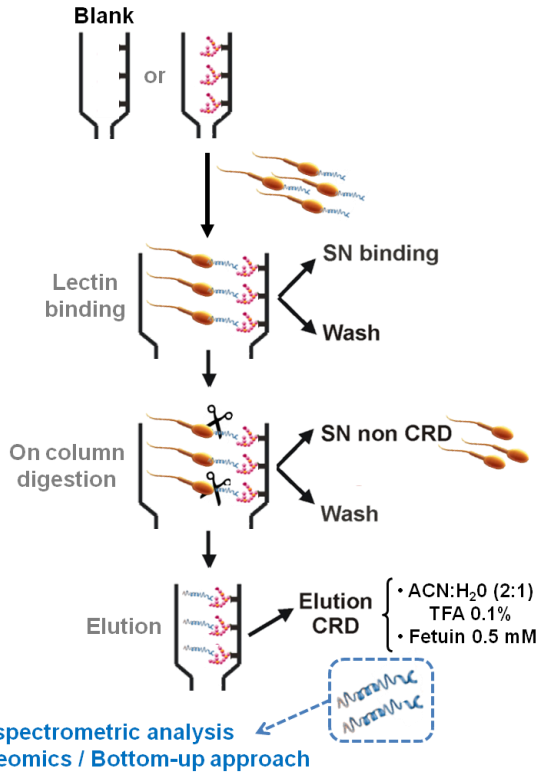


Figure 3.4. CREDEX-MS excision experiments with capacitated and non-capacitated bovine sperm using the epitopes in the table and a blank.

Once sugars were immobilized and their functionality confirmed through binding tests with specific pure lectins and SDS-PAGE detection, they were used in excision experiments with capacitated and non-capacitated entire sperm cells (Figure 3.4), as done with ECA or PDC-109 lectins in previous chapters. However, instead of MALDI-TOF MS, phase-contrast microscopy was used to check the different fraction compositions.

In addition to the carbohydrate-functionalized column, a monofunctionalized methylvinylsulfonyl-activated sepharose column devoid of glycan was included in each CREDEX-MS experiment as a blank to evaluate non-specific and solid-support binding.

In a typical experiment with entire sperm cells, capacitated or non-capacitated sperm were added onto the various sugar-DVS-sepharose or blank columns and incubated for 24 h at 37 ºC. Unbound sperm was removed in the flow through fraction (Figure 3.5A)

and the columns were washed extensively with CM or NCM until no sperm was observed by microscope (Figure 3.5B). Then, lectin-sugar complex/es were digested on the column with trypsin at 37 ºC overnight. This non-reductive on-column strategy, aside from allowing digestion after the native recognition event, hence emulating the *in vivo* situation; also allowed to remove indirectly interacting proteins or non-covalent aggregates. As some of these interactions involved amino acids in protein domains not implicated in carbohydrate recognition, they would be eliminated in the flow through or the supernatant of the digestion. Interestingly, the second flow through fraction obtained after digestion contained sperm heads and tails (Figure 3.5C) which confirmed that binding was indeed taking place on the column, as in previous washes no sperm was observed under the microscope (Figure 3.5B). Therefore, tryptic peptides belonging to the carbohydrate-recognition domain of sperm surface proteins with affinity for the immobilized sugars had been retained in the column; whereupon columns were further washed until no sperm fragments were observed (Figure 3.5D) and, finally, specifically bound peptides were eluted with ACN: H_2O 2:1 TFA 0.1% (Figure 3.5E) and analyzed by LC-MS/MS.

In excision experiments with Sia-containing columns (3'-SLN and SLe^x glycoprobes), a special strategy was introduced to ensure effective separation of bound peptides that might not elute under the standard ACN/H_2O/TFA conditions. To this end, an additional competitive elution was done with fetuin, a glycoprotein known to inhibit Sia-like moiety binding of hamster sperm to the oviductal mucosa. Since only fetuin, but not asialofetuin (*i.e.* fetuin lacking the terminal Sia residues), inhibits that interaction, it follows that the terminal Sia residues in fetuin are involved in the blockade[244], hence fetuin, due to its high Sia content (up to 12-13 carbohydrate chains with terminal Sia)[326] is a more effective eluting agent than Sia itself for those proteins that remain retained by Sia-containing glycoprobes and do not eluted with the standard ACN: H_2O 2:1 TFA 0.1%. Fetuin use in CREDEX-MS experiments with sperm samples was previously validated through binding tests with commercial specific lectins and SDS-PAGE detection. The structural basis of this interaction is not well understood, although it appears that glycoprotein conformation, hence specific steric relationships between oligosaccharide chains, plus proximity/orientation of the charged terminal residues, may bear on the ability of fetuin to interfere with sperm binding.

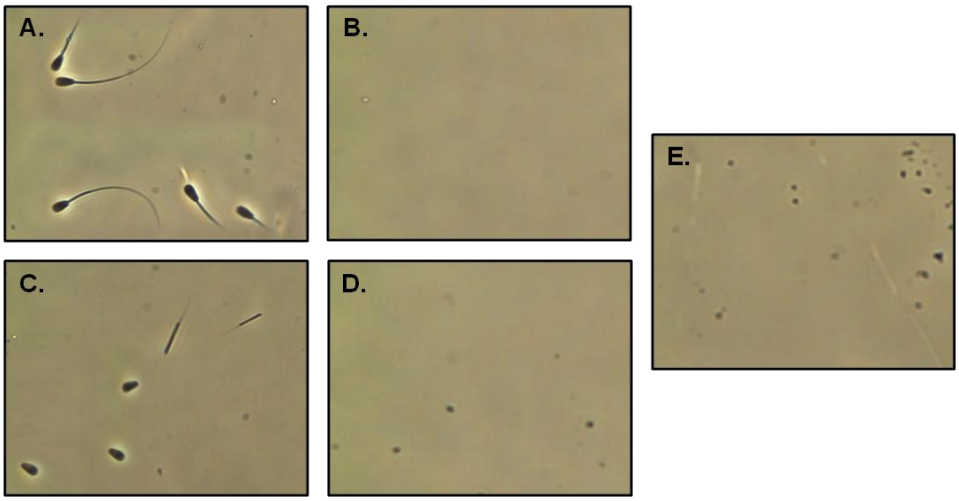


Figure 3.5. Phase contrast micrographs of the different fractions obtained during the Fuc4N CREDEX-MS experiment with sperm: (**A**) Flow through or supernatant fraction after binding. (**B**) Washing fraction after binding. (**C**) Flow through or supernatant fraction after digestion. (**D**) Washing fraction after digestion. (**E**) Elution fraction.

3.3.3 Trypsinization of sperm surface proteins

Trypsinization of bovine sperm was performed to identify the proteins localized on the sperm surface (*i.e.*, our total protein coverage). In order to compare these proteins with those obtained in CREDEX-MS experiments, sperm samples were treated analogously under both CM and NCM conditions as described in section 3.3.1 and subsequent non-reductive trypsinization was conducted in solution containing 150 µg/mL trypsin concentration. The ratio of sperm to trypsin, *i.e.*, number of tryptic peptides from sperm proteins *vs* trypsin autolysis peaks, was optimized to $5 \cdot 10^6$ spermatozoa : 500 µg/mL trypsin[327]. Trypsinized sperm samples were filtered to remove remaining trypsin, sperm heads and/or tails, and the filtered fraction was directly analyzed as described in the next section.

3.3.4 Proteomic analysis of proteins in the CREDEX elution fraction and in sperm trypsinization experiments

For reliability, all experiments with sperm samples were performed at least in triplicate, blank samples included. Specifically, 6 replicates from each different glycoprobe elution fraction (including fetuin-mediated elution of the two Sia-containing columns), 3 sperm trypsinization replicates, and 18 blank replicates were done for each CM and NCM condition. Moreover, to ensure functionality and stability, CREDEX columns were tested between replicates as well as at the end of each replicate set, using commercial specific lectins and SDS-PAGE detection.

Taking into account the sample complexity, and in order to improve the assignments in the subsequent MS/MS analysis, trypsin digests were reduced and alkylated before LC-MS/MS. Globally, MS/MS data resulted in the identification of 394 different proteins present in the elution fractions of all the different sperm CREDEX experiments under both CM and NCM conditions. Furthermore, a total of 158 and 171 proteins were identified in sperm trypsinization experiments with CM and NCM conditions, respectively.

Given the lack of information on which proteins might be involved in the carbohydrate-regulated fertilization events studied here, no targeted scanning and/or quantitative approach could be employed in our proteomic studies. Therefore, a post-processing analysis to determine presence/absence of protein identification under the different CREDEX parameters was chosen. In this analysis, exact measurements of proteomic data sets such as intensity of a protein or peptide were *a priori* less relevant than the change in value (presence/absence) from one sample to the other. To minimize the risk of false positives or negatives, changes in protein identifications were set to be relevant if occurring in two thirds (67%) of the total replicates. After this selection, a total of 94 different proteins were found to be altered in the combined CREDEX experiments: 36 in CM conditions, 13 in NCM, and 45 common to both conditions. These findings confirmed that enrichment indeed was taking place, since in the sperm trypsinization experiments (*i.e.*, total protein count) up to 188 proteins met the selection criteria, distributed as 54 proteins in CM conditions, 35 in NCM, and 99 common to both media.

Heat maps of the results, for capacitated (Figure 3.6) and non-capacitated sperm (Figure 3.7) allowed to establish patterns across proteins and biological samples simultaneously in the large data sets without the need for subjective summarizing[328]. Data were formatted appropriately with statistical programming language R and organized into an $n \times p$ matrix where n is the number of spectral features and p is the number of samples. Thus, a matrix was created so that each column contained data from a single sample (elution fraction of each different CREDEX column) and each row corresponded to a single feature (identified protein). The presence-absence criterion of a given sample was reflected by a 1 or a 0, respectively. The heat map software further reordered the matrix data so that rows and columns with similar profiles were brought close together, thus making patterns more visible. A subsequent procedure translated the numerical matrix into a color image, white for 0 (absence) and red for 1 (presence).

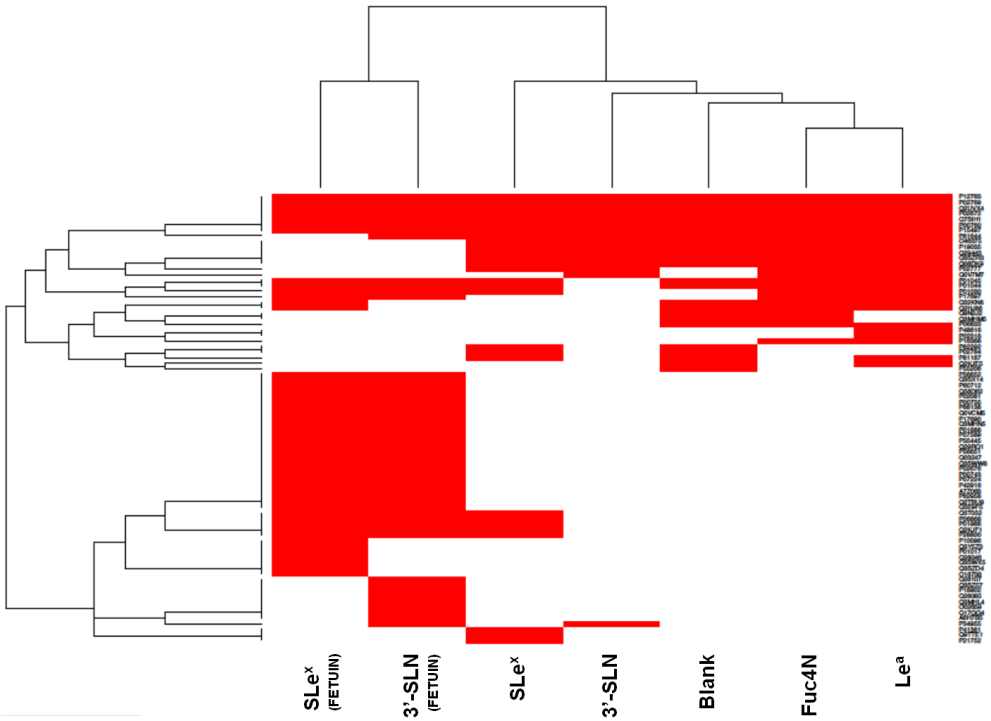


Figure 3.6. Heat map of the hierarchical clustering on the presence-absence protein identification profile matrix in two dimensions of the 7 different elution samples (x-axis) and 81 proteins identified (y-axis) for capacitated sperm.

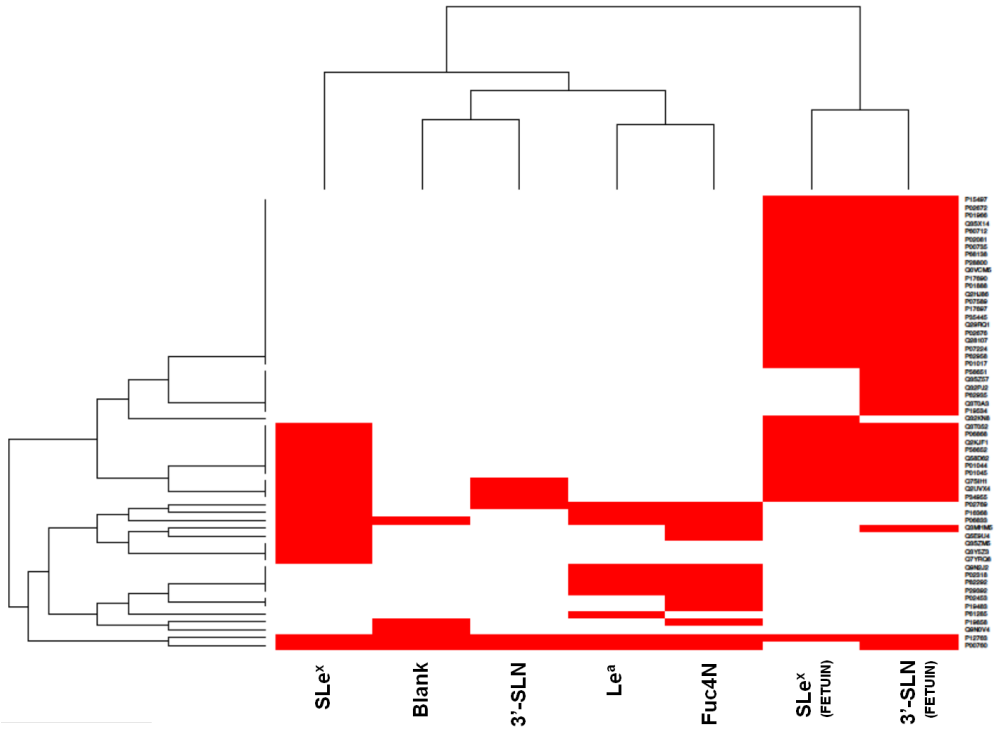


Figure 3.7. Heat map of the hierarchical clustering on the presence-absence protein identification profile matrix in two dimensions of the 7 different elution samples (x-axis) and 58 proteins identified (y-axis) for non-capacitated sperm.

Data reordering is critical to establish patterns in particular data, most heat maps using an agglomerative hierarchical clustering algorithm to that extent, and displaying this information as a dendrogram. When two clusters are merged, a line is drawn connecting the two clusters at a height corresponding to how similar the clusters are. The order of the objects is chosen to ensure that at the point where two clusters are merged, no other clusters are between them, but this ordering is not unique. Inherent to this procedure is the ability to measure the similarity between clusters, *i.e.* the distance between individual observations (distance) and distance between two clusters of observations (agglomeration). The most common measure of distance, *i.e.* Euclidean distance, calculates the difference in location with 0 indicating that the two objects are at the same location, and is the default for all heat map functions. On the other hand, agglomeration is the process by which clusters are merged into larger groups and more importantly, determining which clusters should be merged. The default metric used by the heat map function is called complete linkage. In other words, the distance between two clusters is calculated as the distance between the two most distant points in each cluster.

Taking into account all these considerations, significant differences between the elution samples were found from both heat maps. A general, particularly marked tendency could be observed in second fractions from fetuin-mediated elutions of Sia-containing glycotopes (3'-SLN and SLe^x), which clustered in a similar area on the left side of the heat map for CM (Figure 3.6) and on the right for NCM conditions (Figure 3.7). These clearly merged clusters stressed the similarities between fetuin elutions, the corresponding similar red-heated areas representing proteins with a specific affinity for Sia columns, as they were mainly eluted only under such binding conditions. Cluster connections could also be observed between Fuc-containing columns (Fuc4N and Le^a glycoprobes), some proteins being detected in the ACN standard elution of both columns types.

An alternative way to present data is by means of Venn diagrams, graphic organizers that show all possible logical relations between a finite collection of sets (Figure 3.8). In Venn diagrams, each circle contains the membership of the sample or condition being compared, and the intersection displays the members shared between two or more sets. In other words, non-overlapping spaces contain proteins unique to each condition.

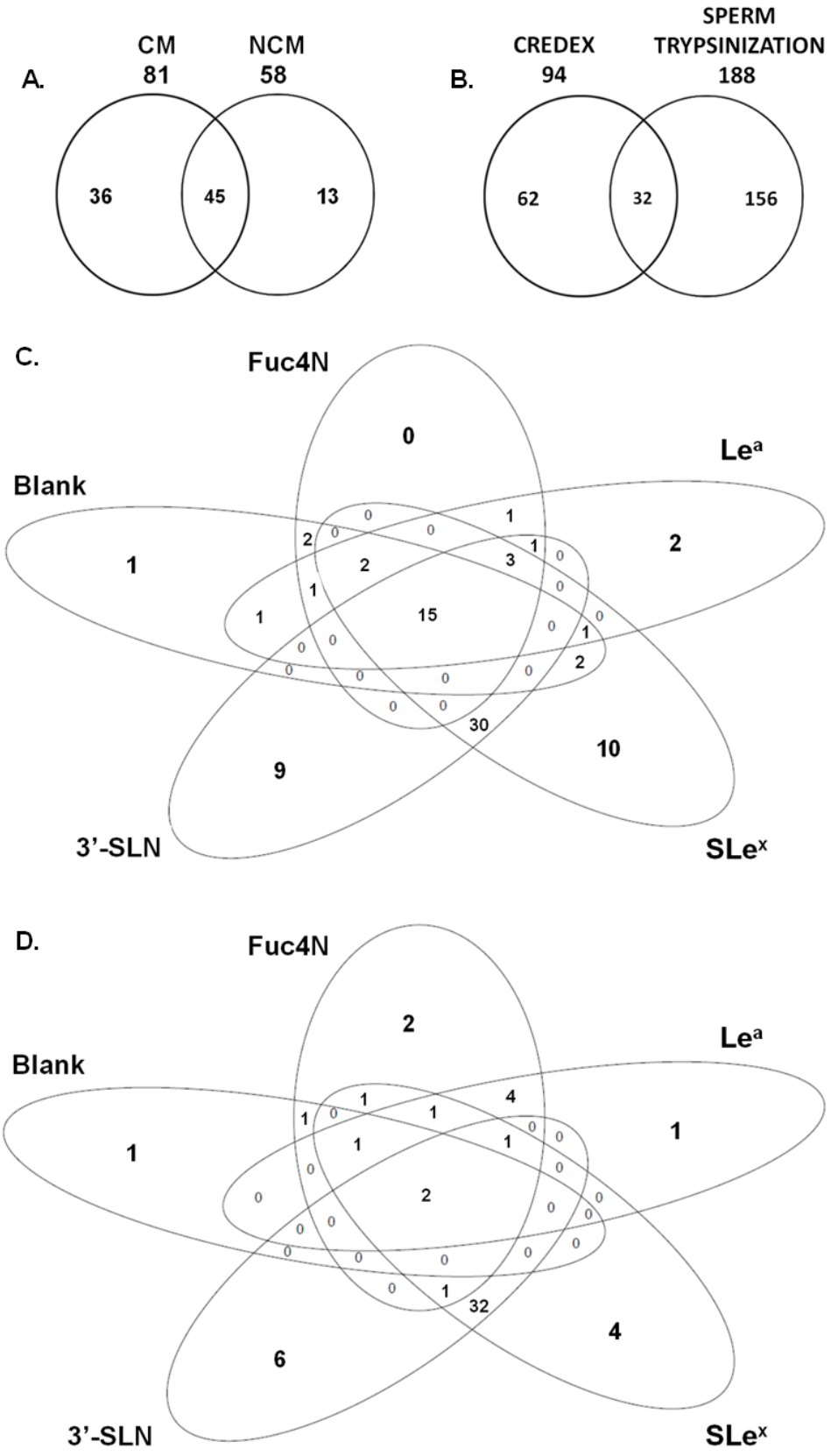


Figure 3.8. Venn diagrams comparing: (**A**) Proteins identified under each sperm condition. (**B**) Proteins identified in each experiment. (**C**) Capacitated sperm proteins identified by each glycoprobe. (**D**) Non-capacitated proteins identified by each glycoprobe.

In Figure 3.8A, the intersection between the 2 different sets of proteins obtained in all combined CREDEX experiments using different (CM and NCM) sperm treatments, contained almost 50% of the total number of different proteins in the collection. Hence, a large number of sperm proteins were conserved under both conditions and were essential for sperm form and/or function regardless of handling conditions. Some proteins, however, were only identified under CM conditions, demonstrating that capacitation involves subtle and intricate modifications of the sperm membrane, including alterations, removals or changes in organization and/or composition of the proteins and lipids at the sperm surface. Indeed, capacitation attempts to mimic *in vivo* conditions in the oviduct, where spermatozoa interact with their immediate environment, and this contact is likely to remodel sperm surface through, *e.g.*, adsorption, removal, aggregation and/or migration of specific proteins, preparing and enabling the spermatozoon to fertilize the egg[329].

As mentioned above, the number of proteins identified in total sperm surface digests was much larger than in the CREDEX experiments (Figure 3.8B). A plausible explanation could be that CREDEX, an affinity enrichment method, reduced sample complexity by yielding predominantly low-abundance, specific proteins[316]. In contrast, trypsinization is done upon whole cells, yielding more complex samples with more proteins identified. The same reasons could also explain the low number of coinciding proteins (32; 13% of total proteins) in the two sets, in comparison with the 62 (25%) exclusively found by CREDEX or the 156 (62%) in the sperm trypsinization experiments.

Venn diagrams in Figure 3.8C and Figure 3.8D compare CREDEX results obtained with different glycoprobes under CM and NCM conditions, respectively. For Sia-containing columns, results from the two different elution fractions (standard and fetuin) were taken together. Common proteins in multiple sets can be distinguished in the different intersections, representing primarily non-specific proteins. The same trends observed in heat map areas were also found with Venn diagrams. Specifically, up to 30 proteins in the intersection between the 3'-SLN and SLe^x glycoprobes sets under both conditions, *i.e.*, with specific affinity for Sia-containing core structures were found. By comparing these with those found with other glycoprobes, conclusions about specificity can be drawn. The complete list of proteins obtained in our proteomic study, organized from the Venn diagrams, is given in Table 3.2 - 3.4.

Table 3.2. Glycospecific proteins found in capacitated and/or non-capacitated bovine sperm.

	CAPACITATED SPERM PROTEINS		PROBE	NON-CAPACITATED SPERM PROTEINS		
			Fuc4N	*ATP synthase sub. α, mitochondrial**	*P19483*	≠
				Collagen alpha-1(I) chain	P02453	≠
=	P02318	Sperm protamine P1‡	Le^a	Dynein light chain 1, cytoplasmic	P61285	≠
≠	P48616	Vimentin*				
			Fuc4N & Le^a	Sperm protamine P1‡	P02318	=
				Spermadhesin-1	*P29392*	=
≠	Q3SZ57	Alpha-fetoprotein*	3'-SLN	Alpha-fetoprotein*	Q3SZ57	≠
≠	Q3MHL4	Adenosylhomocysteinase*		Apolipoprotein A-IV*	Q32PJ2	≠
=	P34955	Alpha-1-antiproteinase‡		Cadherin-2 (Fragment)	P19534	≠
≠	Q28107	Coagulation factor V‡		Complement factor D*	Q3T0A3	≠
=	*Q28085*	*Complement factor H*		ITI heavy chain H2 (Fragments)‡*	P56651	≠
≠	A6H7B5	Signalosome subunit 3*		Peptidyl-prolyl cis-trans isomeraseA*	P62935	≠
≠	O02659	Mannose-binding protein C				
≠	P18902	Retinol-binding protein 4*				
≠	Q17QQ4	Transcrip. initiation factor TFIID sub9*				
≠	Q3Y5Z3	Adiponectin*	SLe^x	Adiponectin*	Q3Y5Z3	≠
≠	Q28046	Adseverin*		CMP-NeuNAc synthase*	Q3SZM5	=
≠	P01017	Angiotensinogen (Fragment)‡*		*Tissue factor pathway inhibitor 2**	*Q7YRQ8*	=
≠	P41361	Antithrombin-III				
≠	Q3SWX5	Cadherin-6				
≠	O18738	Dystroglycan				
≠	*P10096*	*GAPDH**				
≠	Q9TTE1	Serpin A3-1				
=	P21752	Thymosin beta-10*				
≠	Q3SZD4	WD repeat-containing protein 18*				
≠	*P68138*	*Actin, alpha skeletal muscle*	3'-SLN & SLe^x	*Actin, alpha skeletal muscle*	*P68138*	≠
=	*P60712*	*Actin, cytoplasmic 1*		*Actin, cytoplasmic 1*	*P60712*	=
≠	*Q2KJF1*	*Alpha-1B-glycoprotein*		*Alpha-1B-glycoprotein*	*Q2KJF1*	≠
≠	*P28800*	*Alpha-2-antiplasmin*		*Alpha-2-antiplasmin*	*P28800*	≠
≠	P17690	Beta-2-glycoprotein 1*		Beta-2-glycoprotein 1*	P17690	≠
=	*P01888*	*Beta-2-microglobulin**		*Beta-2-microglobulin**	*P01888*	=
≠	P35445	Cartilage oligomeric matrix protein		Cartilage oligomeric matrix protein	P35445	≠
≠	Q29RQ1	Complement component C7		Complement component C7	Q29RQ1	≠
≠	*Q58D62*	*Fetuin-B**		*Fetuin-B**	*Q58D62*	≠
≠	P02676	Fibrinogen beta chain		Fibrinogen beta chain	P02676	≠
≠	P07589	Fibronectin		Fibronectin	P07589	≠
=	*Q3SX14*	*Gelsolin*		*Gelsolin*	*Q3SX14*	=
≠	P02081	Hemoglobin fetal subunit beta*		Hemoglobin fetal subunit beta*	P02081	≠
≠	P01966	Hemoglobin subunit alpha*		Hemoglobin subunit alpha*	P01966	≠
≠	P62958	Protein kinase C inhibitor 1*		Protein kinase C inhibitor 1*	P62958	≠
≠	Q0VCM5	ITI heavy chain H1*		ITI heavy chain H1*	Q0VCM5	≠
≠	P56652	ITI heavy chain H3*		ITI heavy chain H3*	P56652	≠
≠	Q3T052	ITI heavy chain H4*		ITI heavy chain H4*	Q3T052	≠
≠	*P06868*	*Plasminogen*		*Plasminogen*	*P06868*	≠
≠	P00735	Prothrombin		Prothrombin	P00735	≠
≠	P07224	Vitamin K-dependent protein S*		Vitamin K-dependent protein S*	P07224	≠
≠	Q03247	Apolipoprotein E*		Alpha-1-antiproteinase‡	P34955	=
≠	P00743	Coagulation factor X*		Angiotensinogen (Fragment)‡*	P01017	≠
≠	P42916	Collectin-43*		Coagulation factor V‡	Q28107	≠
≠	P56651	ITI heavy chain H2 (Fragments)‡*				
≠	*Q2TBU9*	*RuvB-like 2**				
≠	Q3SWW8	Thrombospondin-4*				
≠	Q5E9F5	Transgelin-2*				
≠	*Q3MHN5*	*Vitamin D-binding protein**				
≠	A7Z063	WAS protein family homolog 1*				

LEGEND FOR TABLES 3.2 – 3.4:

Each table displays the protein accession number and the description for each identified protein.

Glycoprobes with the same core structures, *i.e.* Fuc or Sia-containing core structures, are shaded in the same grey tone.

Ambiguous category contains proteins identified in multiple elution fractions of different columns, hence without defined glycan specificity.

Common proteins identified under both CM and NCM conditions and eluted from the same glycoprobe column are listed first and separated from singular proteins (exclusively detected under one condition) by broken lines.

Proteins found under both conditions but not exactly with the same glycoprobe (same row) though sharing the same functional monosaccharide (same grey tone) are differentiated by the character (‡).

= and ≠ signs denote proteins identified/unidentified in sperm trypsinization experiments, where the maximum possible proteins were identified.

Proteins previously found in other published bovine proteomic studies are shown in italics whereas newly identified proteins are in normal type.

Asterisk (*) refers to proteins with no reported specific function within the field of fertilization.

Table 3.3. Non-specific proteins found in capacitated and/or non-capacitated bovine sperm.

CAPACITATED SPERM PROTEINS				NON-CAPACITATED SPERM PROTEINS		
=	*P55206*	*C-type natriuretic peptide**	**Blank**	*Glutathione S-transferase Mu 1*	*Q9N0V4*	=
≠	*P12763*	*Alpha-2-HS-glycoprotein**	**Ambiguous**	*Alpha-2-HS-glycoprotein**	*P12763*	≠
=	*P06833*	*Caltrin*		*Caltrin*	*P06833*	=
=	P00760	Cationic trypsin*		Cationic trypsin*	P00760	=
=	*P02769*	*Serum albumin*		*Serum albumin*	*P02769*	=
=	*Q3MHM5*	*Tubulin beta-4B chain*		*Tubulin beta-4B chain*	*Q3MHM5*	=
=	*Q3SZR3*	*Alpha-1-acid glycoprotein**		*L-lactate dehydrogenaseA chain**	*P19858*	=
=	P81644	Apolipoprotein A-II*		tRNA-yW-synthesizing protein 3*	Q5E9U4	=
=	P19035	Apolipoprotein C-III*				
=	*Q2KJE5*	*GAPDH, testis-specific**				
=	Q08DK9	PWWP domain-containing MUM1*				
=	O46375	Transthyretin*				
≠	P01030	Complement C4 (Fragments)*				
≠	*P81187*	*Complement factor B*				
≠	P02777	Platelet factor 4*				
≠	*P02784*	*Seminal plasma protein PDC-109*				
≠	*Q29443*	*Serotransferrin*				
≠	Q0V7M7	Spindle and kinetochore protein 1*				

Table 3.4. Proteins found in both conditions but with different glycospecificity.

	ACC.NUM	DESCRIPTION	CM	NCM
=	*Q32KN8*	*Tubulin alpha-3 chain*	**Ambiguous**	**SLe^x**
=	*Q2UVX4*	*Complement C3*		**3'-SLN & SLe^x**
=	P02672	Fibrinogen alpha chain		
≠	*Q7SIH1*	*Alpha-2-macroglobulin*		
=	*P15497*	*Apolipoprotein A-I*		
≠	P01044	Kininogen-1		
≠	P01045	Kininogen-2*		
≠	*Q2HJ86*	*Tubulin alpha-1D chain*		
=	*P17697*	*Clusterin*		
=	*Q9N2J2*	*Glutathione peroxidase 4 (PHGPx), mitochondrial*		**Fuc4N & Le^a**
=	*P82292*	*Spermadhesin Z13*		
=	*P16368*	*Metalloproteinase inhibitor 2*	**Fuc4N & Le^a**	**Ambiguous**

Table 3.2 shows specific proteins identified by CREDEX under both sperm conditions (capacitated on the left, non-capacitated on the right columns); they are distributed in rows separated by solid lines, corresponding to elution fractions obtained with the various glycoprobes indicated in the central column. Table 3.3 presents non-specific proteins identified in blank columns, representing non-specific binding to CREDEX sepharose surfaces, as well as non-specific proteins identified in multiple elution fractions from different probes, hence without defined glycan specificity. Finally, Table 3.4 includes proteins identified under both conditions, but showing different glycan affinities depending on such conditions. Furthermore, in all tables concordance or discordance with total protein content (*i.e.*, from sperm surface trypsin digests) is shown by either = or ≠ signs, pointing out, as explained before, which low-abundance proteins were probably enriched by the CREDEX approach.

After presenting the proteomics experiments and results, a more detailed discussion of some particular proteins identified in this study and how their known functions may potentially impact sperm physiology, is given below by way of various comparative analyses. Proteins that could not be attributed a specific function within the field of fertilization (identified with an asterisk in Table 3.2 - 3.4) are not discussed.

3.3.4.1 *Comparison according to carbohydrate epitope*

a) Fuc-binding bovine sperm proteins

Spermadhesins form a family of low molecular weight (12–16 kDa) secretory proteins found in the male genital tract. They are main components of seminal plasma[295] that associate peripherally with the sperm surface and possess carbohydrate-binding domains. After ejaculation, spermadhesins form a protective coat around the apical portion of the sperm head plasma membrane overlaying the acrosome thus preventing premature acrosomal exocytosis. Moreover, they are believed to act as primary sperm-ZP binding receptor proteins (Table 3.1) owing to their ability to interact with both O- and N-linked oligosaccharides and, in addition, they possess heparin-binding and serine protease inhibitor-binding capabilities. This suggests they participate both in sperm-ZP binding and other events regulating sperm capacitation and acrosome stabilization. In the present study, two members of this family (spermadhesin-1 and spermadhesin Z13) were identified, both with a hitherto unreported specific affinity for Fuc residues (Table 3.2).

The identified acidic seminal fluid protein (aSFP or spermadhesin-1) has redox activity that may not only protect sperm from oxidative stress but also inhibit motility during storage in the cauda region[330], thus with great potential for predicting bull fertility[331]. Binding of aSFP to ejaculated sperm is lost after capacitation[332], suggesting that aSFP may act as a decapacitation factor on bull spermatozoa rather than as a ZP-binding molecule. This latter observation agrees with our results, since aSFP was exclusively detected in non-capacitated sperm. Therefore, considering that: i) sperm reservoir formation appears to be regulated by carbohydrate recognition between Fuc-containing glycans at the surface of the oviductal epithelium and Fuc-binding molecules on sperm surface, and that ii) non-capacitated spermatozoa bind to the oviductal epithelium and their release from the sperm reservoir is apparently coincident with the loss/modification of sperm surface proteins associated with the acquisition of capacitated status, our data would suggest spermadhesin-1 as a potential bovine sperm-surface lectin candidate recognizing Fuc residues in the oviductal epithelium and leading to sperm reservoir formation.

Interestingly, aSFP has 50% homology with spermadhesin Z13[333], the other member of the spermadhesin family identified in our proteomic analysis, for which an inverse relationship with fertility[334] has been proposed yet without a clear mechanism for its effect on bull sperm. In our study, spermadhesin Z13 was identified under both sperm conditions but only showed specificity for Fuc residues in NCM, in line with the above-mentioned role for aSFP.

A few other Fuc-glycospecific proteins were identified in the study (Table 3.2) such as sperm protamine P1 or dynein light chain 1; they are discussed later in section 3.3.4.3.

b) Sia-binding bovine sperm proteins

Many proteins identified in our study show specificity for Sia glycotopes (Table 3.2). This is not an unusual observation, given that Sias are typically attached to the outermost ends of glycoconjugate chains as well as on secreted glycoproteins. This high prevalence of Sias as glycan termini suggests that their predominant function is modulating interactions with the environment[335], ranging from primary to more complex interactions, including reportedly a role in gamete interaction[288]. However, one consequence of this ubiquity, also evident in our study, is that most proteins with this carbohydrate specificity are surface proteins involved in standard cellular functions, not necessarily with clearly established functions in sperm-ZP binding, as anticipated

initially. Even so, as discussed below, some of these common proteins also perform sperm-related functions.

In the oviductal environment, protease activity is modulated by protease inhibitors, and we have identified several of them in our analysis using Sia-containing glycoprobes, namely SERPINs (serine protease inhibitors), α-1-β-glycoprotein, α-1-antiproteinase, α-2-antiplasmin and antithrombin-III. All these components are likely to be important for maintaining protease activity equilibrium in essential processes that preserve sperm and tissue integrity[336-338].

Actin and gelsolin have been detected under both CM and NCM conditions and were specific of both 3'-SLN and SLe^x. Both proteins are functionally related, gelsolin being an actin-binding protein that acts as a key regulator of actin filament assembly and disassembly. While gelsolin can be found in soluble form, in addition to its intracellular form, actin is predominantly cytoplasmic and, in the reproductive context, is involved in the development of hyperactive motility necessary for successful fertilization[339] and, during sperm storage, forms a barrier preventing premature acrosome reaction[340]. Gelsolin, in turn, has been described as a major component of the actin scavenging system, regulating actin polymerization and consequent acrosome reaction[341], as well as protecting viable cells against toxic effects of actin[342]. Moreover, previous studies have shown that gelsolin is a substrate for metalloproteinase (MMP) cleavage[343], which results in considerable loss of actin filament-depolymerizing activity. This suggests that MMPs weaken the extracellular actin-scavenging system by cleaving plasma gelsolin and may, therefore, be involved in pathological conditions induced by extracellular actin. On the other hand, metalloproteinase inhibitors, also detected in this study (metalloproteinase inhibitor-2 (TIMP-2)[344]), complex with and irreversibly inactivate MMPs hence could overcome this limitation.

Plasminogen, the inactive enzyme precursor of plasmin, has been identified in the oviductal fluid and demonstrated to bind oocytes at ZP and oolemma level[345]. Upon sperm contact to the oocyte, plasminogen activators are released and increase the conversion of plasminogen into plasmin, such plasmin protease causes sperm detachment from the ZP, thus contributing to the regulation of sperm penetration in the oocyte[346]. Since plasminogen has been detected in our proteomic approach presenting specificity for Sia glycotopes, it may also bind sperm and further interact with ZP Sia

residues[285] during gamete interaction, thus contributing to plasminogen/plasmin system activation and regulation of the fertilization outcome.

Fibronectin has been postulated to be present on capacitated spermatozoa, since a cellular rather than an exogenous form of the protein has been unmasked on the sperm surface during capacitation. It has also been reported to participate in sperm oolemmal adhesion and gamete fusion[347]. The reportedly marked increase in the population of spermatozoa displaying fibronectin under capacitating conditions and its variation between different semen donors have raised the possibility that defects in fibronectin expression might play a role in sperm dysfunction leading to human infertility. In our study, this protein was identified in both CM and NCM, and shown to have specific affinity for Sia residues.

c) Non-glycospecific bovine sperm proteins

This classification includes a group of proteins (Table 3.3) successfully identified in our study but without defined glycan specificity. The routine observation of these proteins in multiple glycoprobe elution samples indicates that they do not have key functions regarding the sperm-ZP/oviduct binding and are probably involved in other processes vital to sperm physiology.

Albumin, present in all analyzed samples, is reported to modulate sperm lipid levels, protect sperm membrane against lipid peroxidation and maintain high intrasperm ATP levels[348]. Moreover, it mediates capacitation through its activity as a sterol receptor, can stimulate both sperm motility[349] and acrosome reaction[350], as well as participate in fertilization[351]. For that reason, it is an important component of the standard Tyrode's capacitating medium (CM).

Caltrin, also known as seminal plasmin (SPLN), acts as a calcium transport regulator in bovine sperm and also displays a potent antimicrobial activity against a broad spectrum of microorganisms[352]. It has been proposed as a regulator of sperm signal transduction pathways modulating acrosomal exocytosis in response to ZP binding[353] by control of Ca^{2+} influx in the acrosome reaction and activation of acrosin and other serine-proteases at the proper site and proper time to ensure successful fertilization[354]. Moreover, rat caltrin molecules bound to the sperm head during ejaculation prevent the occurrence of the spontaneous acrosomal exocytosis along the female reproductive tract[355]. Consequently, more competent spermatozoa with intact and functional acrosome would be available in the oviduct to participate in fertilization.

Iron is involved in oxidative damage and in the production of lipid peroxide radicals. As a consequence, the reproductive tract must have a scavenging strategy to deal with excessive iron, such as the expression of transferrin. Transferrin is likely important for normal reproduction since its seminal plasma concentration is correlated with some seminal parameters in male infertility[356].

PDC-109 (BSP A1/A2), the major protein of bovine seminal plasma studied in detail in the previous chapter, was evidently also detected with our proteomic approach, confirming earlier findings. However, in our settings no definite glycan specificity could be attributed to PDC-109, in clear agreement with the description in chapter 2 where the molecule displayed a promiscuous binding behavior in front of diverse glycotopes. As mentioned before, PDC-109 is an important mediator in sperm capacitation, probably through sperm membrane lipid modification events[197] and also appears to contribute to the formation of a sperm reservoir by binding to epithelial cells of the oviduct[189]. Additional experimental evidence documents how PDC-109 functions as a chaperone, possibly related to its role in sperm capacitation[184]. Though being initially a component of seminal plasma, upon ejaculation and like all BSP proteins, it binds to spermatozoa surface via choline phospholipids, hence becoming also a sperm surface protein. In addition, previous studies reported that PDC-109, physiological coating the surface in fresh bull sperm, remained on sperm membranes at the onset of capacitation in the sperm reservoir[251]. According these accounts, PDC-109 should have been detected under both CM and NCM conditions in our proteomic analysis, and also reportedly with Fuc affinity[205]. However, PDC-109 was only detected in CM and with ambiguous glycan specificity. The reason for this discrepancy could be the strict annotation criteria used in our proteomics approach. Thus, in our analysis of glycoprobe replicates (n=6), only proteins identified at least in 4 replicates were entered in the final list. Upon these criteria, PDC-109 was not included in the NCM list, although it was indeed identified in 5 Fuc replicates in total (2 Fuc4N replicates plus 3 Le^a replicates).

3.3.4.2 *Comparison according to medium composition*

Table 3.4 includes proteins identified under both conditions, but showing different glycan affinity depending on such conditions. These are probably important proteins by virtue of their presence under both conditions but with a definite role under only one of them, in terms of carbohydrate-specific recognition.

Reactive oxygen species are generated during normal maturation of sperm and can play a role in modulating signaling pathways required for sperm activation, but an excess of such components can be detrimental to sperm function and motility, and is positively correlated with infertility. To protect the integrity of sperm, the epididymis displays an arsenal of defensive proteins and antioxidant systems[357], some of which identified in the present study but presenting non-specificity or different glycan specificity depending on capacitating conditions: serotransferrin, albumin, glutathione-S-transferase, glutathione peroxidase and clusterin; the latter two identified as predictors of bull fertility along with metalloproteinase inhibitor 2 and spermadhesin Z13[334,344,358,359].

Clusterin, identified in both CM and NCM but with Sia specificity just in NCM, is secreted in response to cellular damage and involved in sperm protection, reabsorption of damaged sperm in the cauda epididymis and modulation of complement-induced cell lysis[360]. Protection of sperm from oxidative damage, agglutination or lysis could be important for the spermatozoa once in the female reproductive tract.

Another interesting protein detected, with lipophilic characteristics and glycan specificity for Sia but only in NCM, was apolipoprotein A1 (Apo-A1), a component of the high-density lipoprotein (HDL) complex that interacts with membrane components of ejaculated sperm. HDL in the female reproductive tract is involved in sperm capacitation[191] but whether Apo-A1 directly contributes to that event after ejaculation is unknown.

Glutathione peroxidase 4 (GPX4), a membrane glycoprotein also identified in both media but with glycan specificity only in NCM, has been recently shown to play a significant role in gamete interactions using *in vitro* fertilization assays[361]. Unlike previous proteins in this group, however, GPX4 presents specificity for Fuc instead of Sia residues.

Tubulins, also identified but in an ambiguous way, are structural proteins associated with sperm motility and major components of the flagella[362]. These proteins, even representing axonemal proteins localized mainly in sperm flagella were also present in the analyzed sperm surface proteome. Thus, their detection provided evidence of the limitations of working with entire sperm cells, where identifications of undesired abundant proteins without glycan specificity and no sperm head localization can't be avoided.

3.3.4.3 *Correlation with earlier bovine observations*

The 94 proteins identified with our CREDEX-MS approach were compared against a database generated, which included all proteins identified during a proteomic analysis of bovine spermatozoa. This database represented the work of 12 different proteomic investigations of either whole spermatozoa[363-367], or sperm fractions including isolated plasma membrane[368] and cytosolic matrix[315], or seminal plasma[369,370] including epididymal[371,372] and accessory sex gland[373] fluids. To the best of our knowledge, more than half of these proteins are described here for the first time (*i.e.* there are no previous descriptions of the expression of these proteins in bovine sperm; Figure 3.9). Specifically, out of the 94 proteins in our study, only 36 have been previously reported (Table 3.2 - 3.4; in italics) but 58 new proteins have been found in bovine sperm by our MS/MS approach.

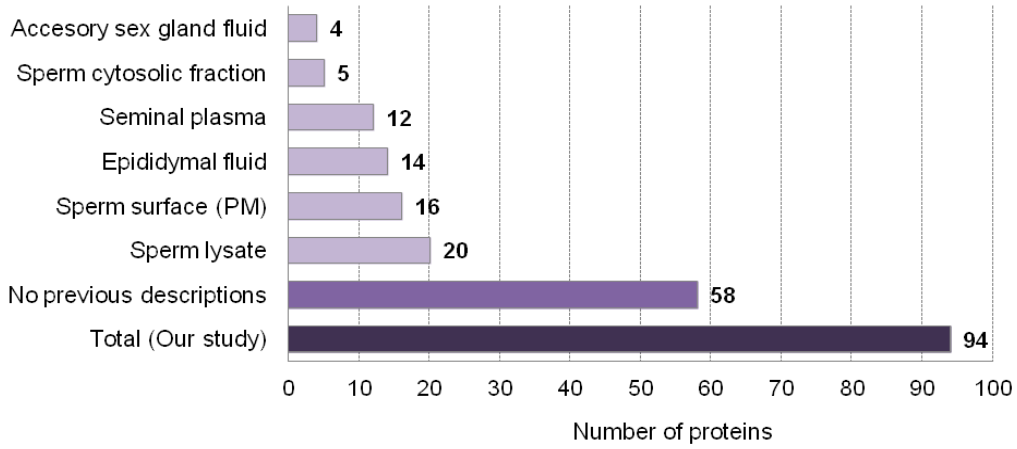


Figure 3.9. Bull sperm proteins in the present study that coincide with previously published descriptions[315,363-373]. Of the 94 proteins identified and considered here, 58 were not previously described in bovine sperm.

These newly identified proteins, most of them not found in the maximum possible protein set (sperm trypsinization experiments data), evidently represent low-abundance proteins identified only due to our enrichment methodology. To gain insight into the biological significance of these new identified proteins, gene ontology functional enrichment analysis was performed using agriGO[374], to group together proteins connected in various gene groups that possessed the same or similar biological function. Of these 58 proteins, 55 were found with GO annotation and classified according to molecular function, cellular component, and biological process (Figure 3.10). Moreover, the same protein categorization was performed with the total of the proteins

considered in the study (94 proteins), in order to compare and determine if significant differences exist. Nevertheless, the majority of proteins newly identified were similar to that identified in the totality (see Figure 3.10) and included binding proteins, transporters, enzyme modulators, hydrolases, cytoskeletal proteins, and others.

The UniProt Knowledgebase (UniProtKB/Swiss-Prot) web site was used to find potential functional information about the proteins identified. Concerning cellular localization (Figure 3.10C), most of the proteins were located in the extracellular region, as intended with our methodology. Other localizations included cytoplasmic vesicles, cytoskeletal microtubules, plasma membrane and nuclear part.

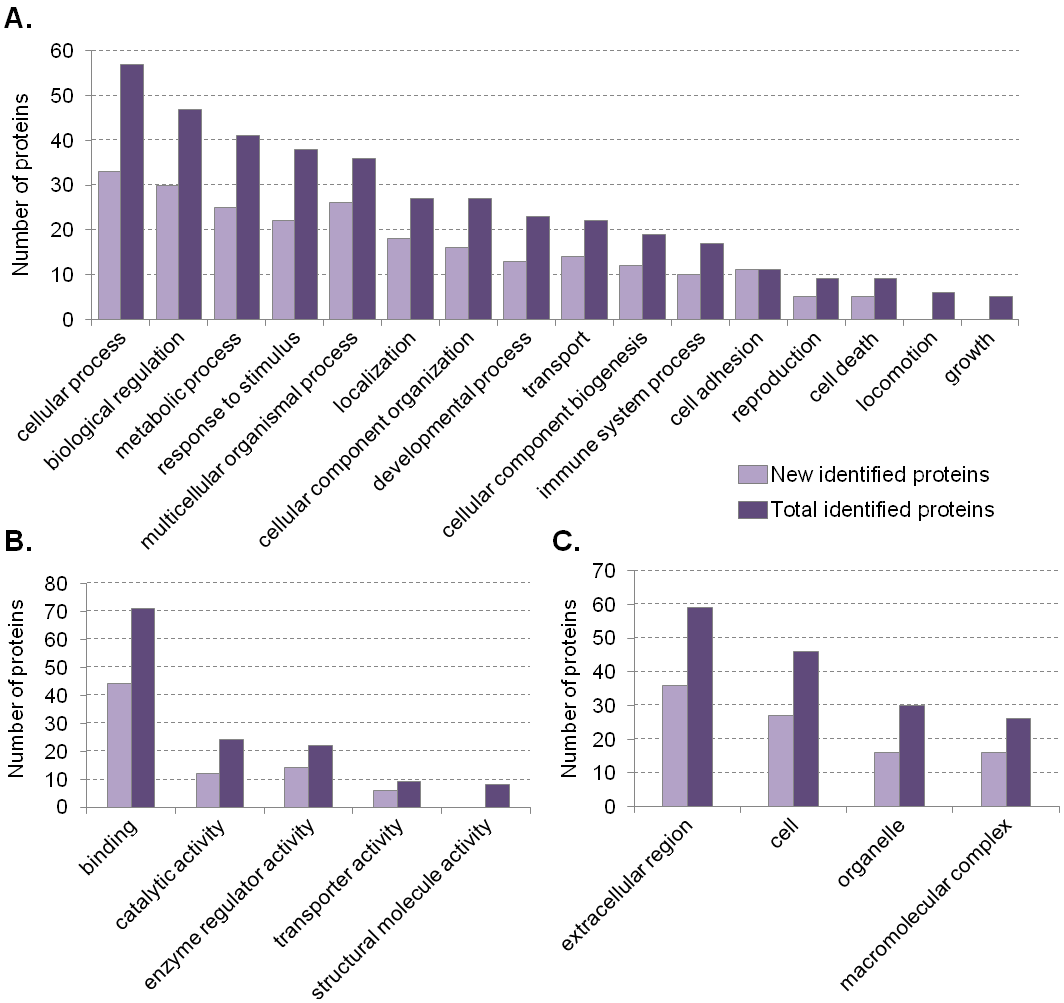


Figure 3.10. New *vs* total bovine sperm proteins identified in our study were grouped using agriGO database according to: (**A**) putative biological process, (**B**) molecular function, and (**C**) cellular component.

The categorization of proteins according to their main function/s (Figure 3.10B) or the biological process which are putatively implicated in (Figure 3.10A), suggested that bovine sperm surface contains proteins involved in standard cellular functions such as

cell metabolic regulation, stimuli and immune response, signaling, intracellular trafficking, *etc*. However, confirming the effectiveness of our approach, the list also includes proteins involved in different parts of the reproductive process such as spermatogenesis, sperm capacitation, fertilization, and acrosome reaction already commented in previous subsections.

An essential prerequisite for any systems-level understanding of cellular functions is to correctly identify and annotate all functional interactions among proteins in the cell. With this aim a STRING network analysis of protein-protein interactions was performed separately in newly identified proteins as well as for proteins previously identified (Figure 3.11) to establish functionally linked proteins and determine the potential underlying biological processes[375]. Networks are presented under confidence view, whereby stronger associations are represented by thicker lines or edges and vice versa. Proteins are represented as nodes with numerical codes that correspond to accession numbers of the proteins considered in our study (listed in Table 3.2 - 3.4) and present in the STRING database. Whenever there is a functional association between two proteins, an edge is set between the corresponding nodes in the graph and the weight of the edge will represent a confidence value on the evidence of such an association.

Figure 3.11A shows predicted interactions between our identified proteins that had been also previously described in published bovine proteomes. Of the 35 proteins present in STRING database, 28 proteins (80%) were found to be linked either directly or indirectly through one or more interacting proteins, suggesting the existence of known functional linkages between these highly conserved proteins. In order to further explore these connections, functional pathways were investigated using the KEGG PATHWAY database. Results revealed two overrepresented pathways, namely “complement and coagulation cascades” within the immune system (including proteins such as alpha-2-macroglobulin, complement C3, complement factor B, complement factor H, plasminogen and alpha-2-antiplasmin) and “Gap junction” cell communication (such as tubulin proteins). In contrast, the remaining 20% of the commonly observed proteins lacked functional interactions with other proteins. Interestingly, both spermadhesins found in our proteomic analysis as well as PDC-109 were assigned to this set of proteins, hence reinforcing the evidence that these proteins cover selective functions related to fertilization and rarely participate in other common cell processes.

Furthermore, the regular observation of these specific proteins in almost all bovine proteomic studies corroborates the important implication of these proteins in the reproductive process.

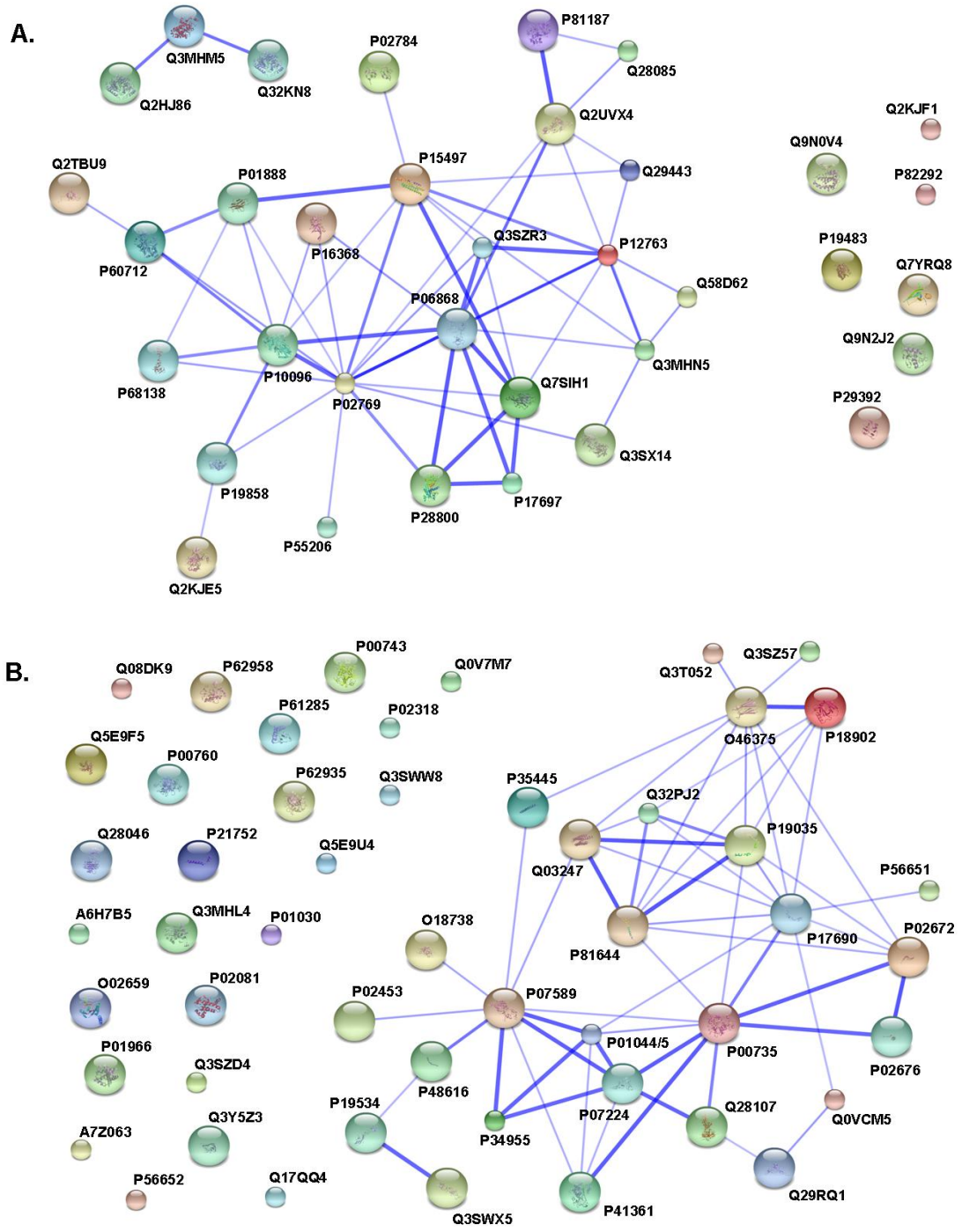


Figure 3.11. STRING connection networks showing the associations between: (**A**) proteins identified with our study that had been also previously described and (**B**) newly identified proteins not found in earlier bovine proteomic studies.

With regard to newly identified proteins, fewer associations between proteins were observed, as shown in Figure 3.11B. Nonetheless, around 50% of the new proteins were linked, and further analysis with the KEGG database revealed that these interconnected proteins were mainly common proteins that participate also in the immune system (with proteins such as prothrombin, coagulation factor V, complement C7, fibrinogen alpha and beta chain, kininogen 1, protein S, alpha-1 antiproteinase, antithrombin) and in extracellular matrix-receptor interactions (including for example cartilage oligomeric matrix protein, collagen alpha 1, dystroglycan, fibronectin). It could be that the carbohydrate-based enrichment process somehow includes or triggers these extracellular matrix-receptor interactions. The other half contains 24 bovine sperm proteins identified for the first time with our proteomic analysis and without known functional association neither extensive background. Of note are protamines (P02318), the most abundant nuclear proteins present in the sperm nucleus of many different species. Many different studies have demonstrated the presence of an altered expression of protamines in male infertility[376] and, recently, protamine 1 and 2 genes expressions levels were correlated to sperm quality parameters, especially to initial progressive motility[377]. Furthermore, dynein light chain 1 (P61285) was found to be involved in generation of murine sperm motility[378] as well as in male germ cell differentiation in *Drosophila*[379].

3.3.4.4 *Correlation with the human sperm proteome*

To date, some efforts have been made to identify the human sperm proteome, and in fact, a list of 1056 proteins has been reported[380]. Additionally, Johnston *et al.*[381] claimed identification of 1760 proteins in human sperm, and more recently with advanced MS and an optimize proteomics platform, 4675 human sperm proteins have been successfully identified[382]. With the available in-depth human sperm proteome, comparative proteomic analyses of sperm between human and other mammalian species such as bovine could be carried out. This will allow the establishment of parallel genetic models for the study of spermatogenesis and sperm function, and will allow for instance the identification of sperm factors responsible for male infertility in humans.

The 65 bovine genes (94 total proteins) from our datasets were mapped to 54 ortholog human genes. Comparison of these human orthologs with the complete list of around 4500 genes recently identified in a published human sperm proteome[382], revealed that

22 genes (41%) were present in the human sperm proteome whereas 32 genes (59%) were not coincident (Table 3.5).

Table 3.5. Orthologous relationship between genes from bovine and human sperm proteome. (a) Orthologous human genes identified in a published human sperm proteome. (b) Orthologous human genes not present in the study[382].

Acc. num *(a)*	Bovine Ensembl Gene ID *(a)*	Orthologous Human Ensembl Gene ID *(a)*	Acc. num *(b)*	Bovine Ensembl Gene ID *(b)*	Orthologous Human Ensembl Gene ID *(b)*
A6H7B5	ENSBTAG00000018973	ENSG00000141030	A7Z063	ENSBTAG00000011902	ENSG00000234769
P01888	ENSBTAG00000012330	ENSG00000166710	O02659	ENSBTAG00000007049	ENSG00000165471
P01966	ENSBTAG00000026418	ENSG00000188536	O46375	ENSBTAG00000010991	ENSG00000118271
P02081	ENSBTAG00000037644	ENSG00000244734	P00735	ENSBTAG00000007148	ENSG00000180210
P02784	ENSBTAG00000023434	ENSG00000188334	P00743	ENSBTAG00000016385	ENSG00000126218
P15497	ENSBTAG00000002258	ENSG00000118137	P01044	ENSBTAG00000005122	ENSG00000113889
P17697	ENSBTAG00000005574	ENSG00000120885	P02453	ENSBTAG00000013103	ENSG00000108821
P18902	ENSBTAG00000000442	ENSG00000138207	P02769	ENSBTAG00000017121	ENSG00000163631
P19858	ENSBTAG00000008683	ENSG00000134333	P07224	ENSBTAG00000023652	ENSG00000184500
P21752	ENSBTAG00000005654	ENSG00000034510	P12763	ENSBTAG00000000522	ENSG00000145192
P34955	ENSBTAG00000018843	ENSG00000197249	P17690	ENSBTAG00000001915	ENSG00000091583
P48616	ENSBTAG00000018463	ENSG00000026025	P19035	ENSBTAG00000012398	ENSG00000110245
P62935	ENSBTAG00000012003	ENSG00000196262	P19534	ENSBTAG00000021190	ENSG00000170558
P62958	ENSBTAG00000010959	ENSG00000169567	P28800	ENSBTAG00000020859	ENSG00000167711
Q03247	ENSBTAG00000010123	ENSG00000130203	P55206	ENSBTAG00000003253	ENSG00000163273
Q17QQ4	ENSBTAG00000027980	ENSG00000085231	P56652	ENSBTAG00000007846	ENSG00000162267
Q2HJ86	ENSBTAG00000030973	ENSG00000178462	P68138	ENSBTAG00000046332	ENSG00000143632
Q2KJE5	ENSBTAG00000015917	ENSG00000105679	P81187	ENSBTAG00000046158	ENSG00000243649
Q3MHL4	ENSBTAG00000018101	ENSG00000101444	P81644	ENSBTAG00000009212	ENSG00000158874
Q3MHM5	ENSBTAG00000025274	ENSG00000188229	Q08DK9	ENSBTAG00000030839	ENSG00000160953
Q5E9F5	ENSBTAG00000002068	ENSG00000158710	Q2KJF1	ENSBTAG00000009735	ENSG00000121410
Q9N0V4	ENSBTAG00000017765	ENSG00000134184	Q3SWW8	ENSBTAG00000012866	ENSG00000113296
			Q3SWX5	ENSBTAG00000012992	ENSG00000113361
			Q3SZ57	ENSBTAG00000017131	ENSG00000081051
			Q3SZD4	ENSBTAG00000002434	ENSG00000065268
			Q3T0A3	ENSBTAG00000048122	ENSG00000197766
			Q3Y5Z3	ENSBTAG00000019813	ENSG00000181092
			Q58D62	ENSBTAG00000017531	ENSG00000090512
			Q5E9U4	ENSBTAG00000005196	ENSG00000162623
			Q7SIH1	ENSBTAG00000018137	ENSG00000175899
			Q7YRQ8	ENSBTAG00000015844	ENSG00000105825
			Q9TTE1	ENSBTAG00000046540	ENSG00000196136

Therefore, cross-referencing our current bull sperm proteomic results with protein identifications published in recently human sperm proteomes revealed evolutionarily conserved sperm proteins (such as PDC-109; P02784), and some potentially new bovine

species-specific sperm surface protein identifications. Noteworthy, cadherins (P19534 and Q3SWX5) were not identified in the recent human study but other data indicate that they are expressed in human gametes and participate in sperm-oocyte interaction events[383]. Another interesting protein is mannose-binding protein C (O02659), which could be related to a human mannose receptor, a candidate protein proposed to interact with the ZP via carbohydrate-binding domains in a capacitated-dependent manner[384]. In fact, in our proteomic results, mannose-binding protein C was only detected in capacitated sperm and had no functional associations with other proteins, whereas cadherins were identified in both conditions and presented connections between them. Moreover, all these proteins were obtained exclusively in our bovine proteomic analysis. Since, in humans, these proteins present putative functions in fertilization processes, it will be interesting to study their bovine gene ortholog functions. Finally, this differential human-bovine protein/gene expression corroborates the importance of species-specific reproductive biology characterization, as some proteins may be present in both species but only present sperm functions in a specific one.

In summary, several categories of proteins in the bovine sperm surface have been detected with different glycan specificity, by means of a non-electrophoretic proteomics study (most of the previous bovine studies used 2D electrophoresis for isolation and identification of sperm proteins) using MS on affinity chromatography-enriched fractions and entire sperm cells. Based on results previously published by others and by ours, we suggest that proteins in the sperm surface are likely to play significant roles in sperm membrane remodeling, transport, protection, and function. The findings of this proteomic study, unveiling a new population of proteins previously not identified by MS/MS analysis in bovine spermatozoa, highlights the efficacy of our novel approach as well as the complex composition and function of sperm. However, one cannot exclude the possibility that some proteins in our list do not directly influence events occurring in the female reproductive tract associated with sperm transport, protection, capacitation and fertilization; but rather belong in an ultrastructural model of functional domains in sperm surface proteins. Moreover, given the highly restricted criteria employed in our post-processing proteomic analysis, sperm proteins with some important functions in fertilization may have been overlooked. Further analysis of these newly identified proteins along with those already known, will afford valuable insights into potential mechanisms of biogenesis, maturation, and function during fertilization.

Furthermore, although it is clear that glycosylation plays important roles in mammalian fertilization interactions, and that sugar binding affinities of the sperm plasma membrane proteins differ among species, little is known about the composition of these proteins. Accordingly, characterization of the gamete glycoproteome should effectively aid in clarifying the mechanisms of fertilization and provide a useful resource for future development of male contraceptives and for diagnosis of male infertility.

The number and diverse nature of ZP-binding candidates identified to date does not support the hypothesis that a single sperm protein be responsible for the various species-specific recognition events of all mammals. Indeed, most workers in the field recognize that multiple receptors or a hetero-molecular complex of proteins are probably involved, and that this is necessary to facilitate the many complex events that occur during sperm–zona binding. Initial sperm–ZP adhesion, specific tight binding, induction of a signal transduction cascade to initiate the acrosome reaction and secondary binding to facilitate ZP penetration need all to be co-ordinated, and there is probably a hierarchy of interactions involving different sperm proteins. The requirement for multiple sperm proteins, in complexed form or alone, is also supported by the temporal and spatial differences observed during sperm–ZP binding. Moreover, since fertilization is a fundamental process for propagation of the species, it might be expected that a number of regulatory processes would have evolved to ensure that initial contact between the spermatozoon and zona leads to fertilization. It is more likely that a combination of binding events takes place, some of which may not be essential, but nonetheless optimizing the chances of fertilization. There might also be a degree of redundancy involved in what is a crucial process to the individual.

Lectin activities on the sperm head have been demonstrated in several species and are commonly implicated in somatic cell–cell recognition, and in some species are clearly extracellular proteins. Thus, these appear to be likely candidates for the regulation of the initial sperm–egg recognition events, if only because no other function within such a complex could easily be attributed to them.

Cloning and site-specific mutagenesis experiments in the next few years will without doubt aid to identify the functional domains of the sperm surface proteins currently believed to be involved in early egg recognition events. It will be interesting to see if a common group of functionalities is shared by all mammalian, and if such functionalities are associated with analogous or distinctly different polypeptide chains in different

species. Moreover, correlations between saccharide moieties, glycan receptors and the outcome of mammalian IVF should be examined in order to help evaluate the precise carbohydrate binding criteria necessary for fertility. However, as fertilization *in vivo* occurs in the specialized milieu created in the ampulla of the Fallopian tube, we must be aware of the signals the environment can induce in spermatozoa, and the sperm surface remodeling it can cause as we draw conclusions as to the mechanisms of early stages in fertilization.

CONCLUSIONS

CONCLUSIONS

Within the context of this doctoral thesis chemical, structural and functional aspects of carbohydrate-driven interactions have been explored. One potent approach based on glycopeptide probes has been developed that, in conjunction with SPR monitoring, provides accurate qualitative and quantitative data on the interaction between carbohydrates and their cognate partners. Furthermore, this approach was complemented with a distinct strategy, denominated CREDEX-MS, to enable the analysis of complex biological specimens and map the carbohydrate recognition domain. Several, peer-reviewed, manuscripts have been published (see Appendixes) and are in-process of being published.

At a more detailed level the conclusions of this thesis are:

Technological:

1. The glycopeptide probe was optimized to ensure a native-like ring structure when using monosaccharides. This was achieved through the incorporation of *N*[Me]-Aoa- as a connector between the sugar and the peptide module.
2. The protocols established in this work are of general application for preparing a broad range of glycoprobes exposing the most relevant structural glycotopes encountered in the mammalian system that can efficiently and reproducibly be immobilized for carbohydrate binding studies with well-known lectins. In particular, the methodology allows differences in glycosidic linkage to be fully resolved employing lectin amounts in the nM to µM range.
3. For the CREDEX-MS approach the linkage between the carbohydrate and the DVS-functionalized support was evidenced by dedicated synthesis and meticulous NMR and MS analysis concluding that conjugation is exclusively through the anomeric carbon hydroxyl rendering a closed ring conformation.
4. Interactions based on carbohydrates are relatively weak but the peptides involved in carbohydrate binding are bound sufficiently strong and durable to withstand prolonged, sequential proteolysis experiments enabling the accurate isolation of all the interacted peptides and removal of the unspecific ones.

Methodological:

5. Glycopeptide probes displaying four β-galactoside containing epitopes were prepared and immobilized onto the SPR sensor surface and a meticulous study of the galactose-binding using *Erythrina cristagalli* agglutinin (ECA) as a lectin model was completed. Kinetic data were similar to those obtained by other reported studies, and therefore the proof of principle of our alternative approach to study carbohydrate-lectin interactions was established.
6. The application of our SPR-based approach for thermodynamic analysis resulted in enthalpy values comparable to those obtained by ITC. However, significant differences were observed in entropic values, probably due to mass transport limitations.
7. The SPR-based approach is fully compatible with affinity capture of lectins and subsequent characterization by MALDI-TOF MS, hence making our methodology applicable for pull-down assays.

Biological:

8. The applicability, convenience and complementarity of SPR and CREDEX-MS approaches were successfully tested with the simplified biological sample, PDC-109. Its lectin-binding profile against the most relevant mammalian glycotopes, characteristic binding due to polydisperse nature and carbohydrate-binding site were evaluated by means of both techniques; results were consistent with previous observations.
9. CREDEX-MS technique was used to examine a complex biological sample such as seminal plasma, providing a bioaffinity assay that resembles as far as possible the physiological situation. Carbohydrate-binding sperm surface proteins differentially expressed depending on different glycotopes were enriched, identified by state-of-the-art proteomics, and mapped with the fertilization events in bovine species.
10. Further analysis of the identified sperm surface proteins along with those already known and those present in other cell types of the male and female genital tracts, will pave the way to a greater understanding of the molecular mechanisms that control the process of mammalian reproduction, which will be of utmost value in the field of assisted reproduction.

MATERIALS AND METHODS

MATERIALS AND METHODS

i. Materials and biological samples

i.i. Chemicals

Fmoc (N^{α}-(9-fluorenylmethyloxycarbonyl)) protected amino acids were purchased from Senn Chemichals (Dielsdorf, Switzerland). Dicyclohexylammonium (DCHA) salt of Boc (tertbutyloxycarbonyl)-methylaminooxyacetic acid was from NeoMPS (Strasbourg, France). Rink amide MBHA resin was from Novabiochem (Läufelfingen, Switzerland). 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) was obtained from Iris Biotech (Marktredwitz, Germany). *N*,*N*-diisopropylethylamine (DIEA) was from Merck Biosciences (Darmstadt, Germany), and triisopropylsilane was from Sigma-Aldrich (Madrid, Spain). HPLC-grade acetonitrile (ACN), *N*,*N*-dimethylformamide (DMF), trifluoroacetic acid (TFA), diethyl ether, and pyridine were from SDS (Peypin, France).

Monosaccharides (Glc, Fuc, GlcNAc, αMe-Man and Sia) and lactose (Gal-β1,4-Glc) were purchased from Sigma-Aldrich (Madrid, Spain). Other carbohydrates (Gal-β1,(3,4,6)-GlcNAc, Man-α1,(2,3,6)-Man, Sia-α2,3-Gal-β1,4-GlcNAc, Sia-α2,6-Gal-β1,4-GlcNAc, Sia-α2,3/6-Gal-β1,4-Glc, Sialyl Lewis^{x} and Lewis^{a}) employed in this work for glycopeptide synthesis were from Dextra (Reading, United Kingdom). Disaccharides with terminal Fuc (Fuc-α1,(3,4,6)-GlcNAc) were obtained from Toronto Research Chemicals (Toronto, Canada) and αMe-Fuc was from Iris Biotech GmbH (Marktredwitz, Germany).

Lectins from *Erythrina cristagalli* (ECA), *Maackia amurensis* (MAA), *Sambucus nigra* (SNA), *Canavalia ensiformis* (Con A), *Lotus tetragonolobus* (LTA) and *Ulex europaeus* (UEA) were purchased from Sigma-Aldrich (Madrid, Spain). Lectin from *Aleuria aurantia* (AAL) was from Vector laboratories (Burlingame, USA).

Aniline, 2,5-dihydrobenzoic acid (DHB), and α-cyano-4-hydroxycinnamic acid (CHCA) were obtained from Sigma-Aldrich (Madrid, Spain). Sinapinic acid (SA) was from Fluka (Madrid, Spain). Poros R2 was obtained from Applied BioSystems (Foster City, USA) and ZipTip® Pipette Tips were from Merck Millipore (Merck KGaA, Darmstadt, Germany).

Sequencing-grade modified porcine trypsin was from Promega (Madison USA). Sequencing grade chymotrypsin, Glu-C, and PNGase F enzyme were obtained from Roche Diagnostics GmbH (Penzberg, Germany).

CM5 and C1 sensor chips, 1-ethyl-3-(3-diethylaminopropyl)-carbodiimide (EDC), *N*-hydroxysuccinimide (NHS), ethanolamine hydrochloride pH 8.5, and HBS-P (0.01 M HEPES pH 7.4; 0.15 M NaCl; 0.005% v/v surfactant P20) buffer were from Biacore (GE Healthcare, Uppsala, Sweden).

Sepharose-4B, divinylsulfone and methyl-vinylsulfone were purchased from Sigma-Aldrich (Madrid, Spain). Microcolumns and 35-µm pore size filters were from MoBiTec (Goettingen, Germany).

NuPAGE® Novex® 4-12% Bis-Tris precast polyacrylamide gels, NuPAGE® MES and MOPS SDS running buffers, NuPAGE® sample reducing agent and antioxidant, Colloidal Blue Staining Kit, BenchMark™ Protein Ladder, and Novex® Sharp Unstained Protein Standard were from Invitrogen (Life Technologies, Carlsbad, CA, USA). Silver staining kit and Coomassie Brilliant Blue R-250 were from Bio-Rad Laboratories (Hercules, CA, USA).

Components of sperm culture media (NaCl, KCl, $NaH_2PO_4 \cdot H_2O$, $CaCl_2 \cdot 2H_2O$, $MgC1_2 \cdot 6H_2O$, HEPES, $NaHCO_3$, BSA, sodium pyruvate, sodium lactate, gentamicin) were purchased from Sigma-Aldrich (Madrid, Spain).

i.ii. PDC-109

PDC-109, purified from bull seminal plasma, was kindly donated by Dr. JJ Calvete. The purity of PDC-109 was assessed by SDS-PAGE, where the protein moved as two closely spaced bands of Mr ~13 kDa, corresponding to the glycosylated and non-glycosylated forms. Acrylamide gels of 15% or 18% were used. Moreover, MALDI-TOF MS characterization showed the two forms and several oligomers of PDC-109 in the linear mode and positive polarity with SA matrix. The concentration of purified PDC-109 was estimated from its extinction coefficient at 280 nm of 2.5 for a 1-mg/mL sample concentration[176].

Putative N-deglycosylation of PDC-109 was done using PNGase F enzyme (1 µL; 250 units / 0.25 mL) in phosphate buffer, 50 mM pH 7.3, 16h, 37ºC. After filtration (10 KDa

filter) to elimate buffer salts, the efficiency of the procedure and the resulting compound/s were checked by SDS-PAGE and MALDI-TOF MS.

i.iii. Semen preparation

Freshly ejaculated bull semen was obtained after electroejaculation of bulls (Asturiana de los Valles breed, *Bos taurus*) at the Cenero (Asturias) artificial insemination facility and kindly donated by "Asociación Española de criadores de ganado vacuno selecto de la raza Asturiana de los Valles (ASEAVA)". Semen aliquots were stored in liquid nitrogen immediately after collection and kept at -176 °C during transportation and storage. Prior to use, aliquots were thawed for 10 s at room temperature and immediately afterwards placed in a water-bath at 37 °C for 40 s. The liquefied semen was next subjected to the washing swim-up technique, as follows.

Briefly, semen samples (250 µL aliquots) were layered in cryotube vials under 1 mL of either capacitating or non-capacitating medium (previously conditioned at 37 °C), and incubated for 1 h at 37 °C. During this time sperm are allowed to swim up in the medium, with the purpose of collecting the most motile, active and normal spermatozoa, free of debris and seminal plasma. The supernatant (~700 µL) was collected and centrifuged at 200 g for 10 min. The top layer was discarded and the final pellet (~500 µL) kept at 37 °C till further use in subsequent affinity chromatography experiments. Samples prepared in this way were examined under light microscope before use to evaluate sperm motility and morphology. Total sperm count was assessed using an improved Neubauer hemacytometer and ~1.5×10^6 sperm cells employed in each experiment.

Capacitating medium (CM)[172] was composed of 114 mM NaCl, 3.2 mM KCl, 0.3 mM $NaH_2PO_4\cdot H_2O$, 10 mM sodium lactate, 2 mM $CaCl_2\cdot 2H_2O$, 0.5 mM $MgC1_2\cdot 6H_2O$, 10 mM HEPES, 25 mM $NaHCO_3$, 0.06 % BSA, 1 mM sodium pyruvate, 50 µg/mL gentamicin, pH 7.3. Non-capacitating medium (NCM) was 100 mM NaCl, 0.36 mM $NaH_2PO_4\cdot 2H_2O$, 8.6 mM KCl, 23 mM HEPES, 0.5 mM $MgCl_2\cdot 6H_2O$, 11 mM glucose, pH 7.6. Both media were filtered (0.22 µm) and stored at 4 °C (maximum 1 week) before use.

For trypsinization of sperm surface proteins, an aliquot was thawed and separated in two fractions with three replicates per fraction that were washed following the swim-up procedure explained above in CM and NCM, respectively. Final pellets obtained under

both conditions were resuspended in 85 µL of 25 mM NH_4HCO_3 (pH 8.5) and 15 µL of 1 g/l trypsin in NH_4HCO_3 were added. The resulting solutions were incubated overnight at 37 °C, then trypsin digests were filtered (10 kDa centrifugal filter) to remove remaining trypsin and sperm heads and/or tails. Finally, samples were lyophilized before LC-MS/MS analysis.

ii. Peptide and glycopeptide synthesis

ii.i. Carbohydrate Binding Peptides (CBPs) synthesis

The linear peptides were synthesized in automated peptide synthesizer: a Prelude (Protein Technologies, Tucson, AZ) or ABI433 (Applied Biosystems, Foster City, CA) running standard Fmoc chemistry protocols at 0.1 mmol scale on a Rink amide MBHA resin (0.70 mmol/g substitution). The side chain functionalities were protected with ^{t}Bu (Asp, Glu, Ser, Thr, Tyr), Boc (Lys, Trp), Pbf (Arg), and Trt (Asn, Gln, His) groups. Couplings were done with an eight-fold molar excess (0.8 mmol) of Fmoc-L-amino acids, HBTU, and HOBt; in the presence of double molar that amount of DIEA (0.16 mmol), in DMF as solvent. The sequences of the distinct peptides are depicted in Table 1.

Table 1. Synthetic carbohydrate-binding peptides (CBPs) of ECA lectin.

Code	Peptide from ECA	Sequence	$[M+H]^+$ calc	$[M+H]^+$ exp
CBP1a	[84-91]	RPLPADGL	836.49	836.24
CBP1b	[82-93]	YTRPLPADGLVF	1346.73	1346.91
CBP1c	[80-95]	QPYTRPLPADGLVFFM	1849.95	1849.35
CBP1d	[78-97]	IEQPYTRPLPADGLVFFMGP	2246.16	2246.11
CBP1e	[76-99]	SIEQPYTRPLPADGLVFFMGPTK	2709.40	2709.14
CBP2a	[129-135]	DTFSNPW	864.38	864.91
CBP2b	[127-137]	EFDTFSNPWDP	1352.57	1352.23
CBP2c	[125-139]	AVEFDTFSNPWDPPQ	1747.78	1747.99
CBP3a	[214-220]	GATGAQR	658.35	658.48
CBP3b	[212-220]	LSGATGAQR	858.47	858.32
CBP3c	[210-220]	VGLSGATGAQR	1014.56	1014.43
CBP3d	[206-220]	EWVDVGLSGATGAQR	1543.77	1543.94
CBP3e	[202-220]	QVLPEWVDVGLSGATGAQR	1981.04	1981.31

Peptide cleavage and work-up

Peptides were fully deprotected and cleaved with TFA-water-triisopropylsilane (95:2.5:2.5 v/v, 90 min, r.t.) or TFA-water-1,2-ethanedithiol-triisopropylsilane (94:2.5:2.5:1 v/v, 90 min, r.t.), precipitated by addition of chilled diethyl ether, taken-up in aqueous AcOH (10% v/v), and lyophilized. RP-HPLC purification gave homogeneous materials with the expected mass by MALDI-TOF MS (Table 1).

ii.ii. Peptide module synthesis

Peptide substrate GFKKG-amide was synthesized by Fmoc-based solid-phase synthesis on a Rink amide MBHA resin (0.70 mmol/g) using Fmoc chemistry at a 0.1 mmol scale in an Applied Biosystems 430A automated synthesizer.

Boc-N[Me]-Aoa-OH coupling

Prior to manual coupling with the peptide substrate, Boc-methylaminooxyacetic acid·DCHA salt (500 mg Boc-*N*[Me]-Aoa-OH/DCHA) was converted to the free carboxyl form by acid extraction with 0.1 M HCl and ethyl acetate (50 mL each). The organic phase containing Boc-*N*[Me]-Aoa-OH was recovered from three consecutive extractions. The ethyl acetate was evaporated and the residue weighed. Manual couplings with 3 equivalents each of Boc-amino acid and HBTU, and 6 equivalents of DIEA were used for 1 h, r.t., in DMF.

Peptide acetylation

A fraction of the peptide substrate GFKKG-amide resin was acetylated for SPR reference channel purposes with acetic anhydride (1 mmol; 94.5 µL) and DIEA (348.4 µL) in DMF for 1h. Then, the resin was cleaned and prepared for subsequent cleavage.

Peptide cleavage and work-up

Resin cleavage and full deprotection were done with TFA-water-triisopropylsilane (95:2.5:2.5, v/v, 90 min, r.t.). Peptides were isolated by precipitation with cold diethyl ether and centrifugation, then solubilized in water and lyophilized. The synthetic products (*N*[Me]-O-Aoa-GFKKG-amide and Ac-GFKKG-amide) were >95% pure by analytical RP-HPLC and had the correct mass by MALDI-TOF MS ($[M+H]^+_{calc}$ ~ $[M+H]^+_{exp}$; 622.4 and 577.7 respectively).

ii.iii. Chemoselective ligation of carbohydrates

Conjugation between *N*[Me]-O-Aoa-GFKKG-amide peptide and oligosaccharides was done at 20 and 25 mM, respectively in 0.1 M NaOAc for 72 h at 37 ºC and pH 3.5 for NAc-hexosamines or pH 4.6 for hexoses. Particularly, conjugation reactions with Fuc-containing disaccharides (Fuc-α1,(3,4,6)-GlcNAc) were performed both in the presence and absence of aniline (100 mM) as a nucleophilic catalysts for improvements in conjugation yields.

All glycoconjugates were purified by semi-preparative RP-HPLC on SphereClone C18 (Phenomenex, 250 x 10 mm; 5 μm) using a 10-20% linear gradient of acetonitrile into water (both eluents with 0.1% TFA). Immediately after purification, glycopeptide-containing fractions were neutralized with 10 mM NH_4HCO_3 (up to pH~5) to prevent acid degradation, and lyophylized. All synthetic di/trisaccharide-*N*[Me]-O-Aoa-GFKKG-amide glycopeptides were >90% pure by analytical RP-HPLC and had the expected mass by MALDI-TOF MS as shown Table 2.

Table 2. Synthetic neo-glycoprobes displaying different oligosaccharides.

	Glycopeptide	Conversion* (%)	t_R ♯ (min)	$[M+H]^+_{calc}$	$[M+H]^+_{exp}$
Gal sugars	Gal-β1,4-Glc-*Aoa(m)-GFKKG*	70	5.91	946.48	946.78
	Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	69	7.07	987.50	987.86
	Gal-β1,3-GlcNAc-*Aoa(m)-GFKKG*	51	7.34	987.50	987.94
	Gal-β1,6-GlcNAc-*Aoa(m)-GFKKG*	33	6.58	987.50	987.85
Man sugars	Man-α1,2-Man-*Aoa(m)-GFKKG*	12	6.72	946.48	946.41
	Man-α1,3-Man-*Aoa(m)-GFKKG*	42	6.56	946.48	946.39
	Man-α1,6-Man-*Aoa(m)-GFKKG*	43	7.23	946.48	946.62
Fuc sugars	Fuc-α1,3-GlcNAc-*Aoa(m)-GFKKG*	9	7.79	971.51	971.35
	Fuc-α1,4-GlcNAc-*Aoa(m)-GFKKG*	16	7.80	971.51	971.36
	Fuc-α1,6-GlcNAc-*Aoa(m)-GFKKG*	55	8.30	971.51	971.58
Sia sugars	Neu5Ac-α2,3-Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	54	6.26	1279.60	1279.48
	Neu5Ac-α2,6-Gal-β1,4-GlcNAc-*Aoa(m)-GFKKG*	46	6.16	1279.60	1279.63
	Neu5Ac-α2,3/6-Gal-β1,4-Glc-*Aoa(m)-GFKKG*	47	5.54	1223.55	1223.79

*Conversion of conjugation reactions were estimated by integration of HPLC peak areas at 220 nm.

♯ Retention times in an analytical RP-HPLC with linear gradients started at 10% buffer B with a flow rate of 1 mL/min, and gradually increased to 20% buffer B in 20 min (buffer B: 0.036% TFA in ACN and buffer A: 0.045% TFA in H_2O).

iii. High performance liquid chromatography (HPLC)

RP-HPLC analysis and purifications were performed using the following systems:

Compact Shimadzu LC-2010A for analytical purposes with a Phenomenex Luna C_8 (3 µm, 50 x 4.6 mm) for peptides and a Sphereclone C_{18} (5 µm, 250 x 10 mm) for glycopeptides. Linear gradients of buffer B (0.036% TFA in ACN) into buffer A (0.045% TFA in H_2O) were used over 15 min for Phenomenex Luna column and over 20 min for Sphereclone column at a flow rate of 1 mL/min.

Shimadzu SCL-10A for semi-preparative purifications. Phenomenex Luna C_8 (10 µm, 250 x 10 mm) for peptides and SphereClone C_{18} (10 µm, 250 x 10 mm) for glycopeptides. Linear gradients of buffer B (0.1% TFA in ACN) into buffer A (0.1% TFA in H_2O) over 30 min for Phenomenex Luna column and over 20 min for Sphereclone column at a flow rate of 5 mL/min.

Under each condition the yield was estimated by integration of analytical HPLC peaks at 220 nm.

iv. Mass spectrometry (MS)

iv.i. MALDI-TOF

Peptides and glycopeptides were dissolved in water and mixed with the corresponding matrix solution (1:1 v/v) and 1 µL of the mixture was applied to the MALDI target and allowed to dry at room temperature. For both synthetic peptide and conjugate analysis, a solution of DHB (10 mg/mL) in ACN:water:TFA (70:30:0.1 v/v/v) was chosen. For peptide mixtures generated after digestion, a saturated solution of CHCA in ACN:water:TFA (70:30:0.1 v/v/v) was used. For protein analysis, a solution of SA (10 mg/mL) in ACN:water:TFA (70:30:0.1 v/v/v) was used.

Experiments were carried out on a Voyager-DE™ STR Biospectrometry workstation (Applied Biosystems, Foster City, USA) equipped with a N_2 laser (337 nm). Peptides and glycopeptides were measured in reflectron mode and positive polarity, except for sialic acid containing probes that were measured both in the positive and negative mode. Proteins were measured in the linear mode and positive polarity. External calibration of the spectrometer was performed using Sequazyme™ Peptide Mass

Standard Kit (PerSeptive Biosystems) of the desired range. Recorded data were processed with Data Explorer™ Software (Applied Biosystems, Foster City, USA).

v. NMR spectroscopy

NMR experiments were performed on a Varian Inova VXR-500 spectrometer (Parc Científic de Barcelona, Barcelona, Spain). For NMR experiments, Gal-β1,4-Glc, Gal-β1,4-GlcNAc, Gal-β1,4-Glc-MVS and Gal-β1,4-GlcNAc-MVS were repeatedly exchange in D_2O with intermediate lyophilizations, and finally dissolved in 500 µL D_2O. ^1H 1D and 2D NMR spectra (z-TOCSY) were recorded at 500 MHz at 25 ºC. Chemical shifts (δ) were quoted in parts per million (ppm) and referenced to internal trimethylsilyl propanoic acid (TMSP) (δ 0.00 for ^1H NMR in D_2O); coupling constants (*J*) are quoted in Hz. Spectra were processed using MestReNova (Mnova) sofware (version 6.0.2, MestreLab Research, Santiago de Compostela, Spain).

vi. Quantification methods

vi.i. UV quantification

Pure glycoprobes were quantified by UV-spectroscopy by measuring the absorbance at 258 nm (λ_{max} of a Phe residue) using a Nanodrop device (Nanodrop Technologies, Inc., Wilmington, USA). An operational extinction coefficient for the peptide *N*[Me]-O-Aoa-GFKKG-amide was experimentally obtained (ε 0.1438 mM^{-1} cm^{-1}; l = 1 cm).

vi.ii. Amino acid analysis (AAA)

AAA was used to quantify peptides (*i.e. N*[Me]-O-Aoa-GFKKG-amide and Ac-GFKKG-amide) and glycopeptides (Table 2), as an alternative to UV quantification. Prior to AAA, peptide-containing compounds were hydrolyzed. Briefly, around 10 nmol (10 µg) of peptide or glycopeptides (exact concentration will be quantified) were dissolved in 200 µL of HCl_{conc} (12 M), and 200 µL Nle 0.1 M (20 nmol NorLeu as internal standard) were added in a Pyrex glass tube. Thus, sample hydrolysis was performed with a final concentration of 6 M HCl for 24 h at 110 °C in a heating block. Once hydrolysis was complete, the mixture was evaporated to dryness using a rotary evaporator and samples were stored liophilized until analyzed by GC-MS.

Prior to analysis, amino acids (from samples and calibration standards) were derivatized by resuspending in 50 µL of *N*-tert-butyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA) and 50 µL of acetonitrile. MTBSTFA, forms tert-butyl dimethylsilyl (TBDMS) derivatives when reacted with polar functional groups containing an active hydrogen. Mixtures were heated at 100 °C for 4 hours.

Finally, samples were subjected to GC-MS analysis using an Agilent 6890N Network GC system with an Agilent 5973N mass selective detector. Each sample was injected four times. For GC separation a Phenomenex Zebron ZB-5 cross-linked 5% phenyl polymethyl siloxane capillary column (15 m × 0.25 mm i.d., 0.25 µm film thickness) was used. The oven temperature was initially held at 100 °C and then raised to 310 °C at 25 °C/min. The interface and ion source temperatures were 150 °C and 230 °C, respectively. Column gas flow was 20 mL/min. The MS was operated in full scan mode (50-650 amu; scan rate 2.46 scans/s) with at least two characteristic mass fragments being recorded for each amino acid. Quantification was performed from interpolation of calibration standard curves.

vii. SDS-PAGE

SDS-PAGE protein detection was carried out as previously described by Laemmli[385]. Briefly, 12%, 15% or 18% acrylamide handcasted gels were prepared, and protein samples were separated using the Bio-Rad Mini-PROTEAN 3 Electrophoresis Cell (Bio-Rad Laboratories, Hercules, CA, USA). Gels were run using 1X running buffer (diluted from 10X running buffer stock containing 30.0 g of Tris base, 144.0 g of glycine, and 10.0 g of SDS in 1000 mL of H_2O) and stained using the commercial silver staining kit or with Coomassie Brilliant Blue R-250 from Bio-Rad.

NuPAGE® Novex® 4-12% Bis-Tris precast polyacrylamide gels (Invitrogen Corporation, Carlsbad, CA) were also employed for optimal separation and resolution of small- to medium- sized proteins (1–200 kDa) under denaturing gel electrophoresis conditions using the XCell SureLock® Mini-Cell gel running tank. Gels were run using either NuPAGE® MES (2-200 KDa proteins) or MOPS (14-200 KDa proteins) SDS running buffer to obtain different separation ranges and visualized by staining with Colloidal Blue Staining Kit from Invitrogen.

Proteins were prepared in non-reduced (intact) and reduced forms. For the preparation of non-reduced protein samples, Laemmli sample buffer was added to the protein sample and then vortex for 2 minutes. For the preparation of reduced protein samples, NuPAGE® Sample Reducing Agent (500 mM dithiothreitol) was also added to the protein sample buffer. The mixture was vortexed and then heated at 70 ºC for 10 minutes. Moreover, NuPAGE® Antioxidant was added to the running buffer in the upper (cathode) buffer chamber when performing protein gel electrophoresis under reducing conditions to maintain proteins in a reduced state. Electrophoresis was performed using vertical electrophoresis slab gel apparatus at a constant voltage of 60 volts when samples were in the stacking gel. When the dye front reached the resolving gel, voltage was increased to 120 volts. The run was stopped when the dye front was 2 to 3 mm away from the bottom edge of the gel. BenchMark™ Protein Ladder or Novex® Sharp Unstained Protein Standard both from Invitrogen were used as molecular-weight size markers standards for SDS-PAGE.

viii. Surface Plasmon Resonance (SPR)

All SPR experiments were performed in a BIAcore 3000 instrument (GE Healthcare's Biacore SA, Uppsala, Sweden) by using carboxymethyl-functionalized CM5 or C1 sensor chips and HBS-P (0.01 M HEPES pH 7.4; 0.15 M NaCl; 0.005% v/v surfactant P20) as running buffer, supplemented when necessary with $CaCl_2$ (5 mM) and $MnCl_2$ (1 mM) or other competitive reagents/adjuvants such as bovine serum albumin (BSA), carboxymethyldextran (CMD) or polivinilpirrolidone (PVP). SPR sensorgrams analyses and curve fitting was done with the BIAevaluation 4.1.1 sofware package (GE Healthcare, Uppsala, Sweden).

In all SPR experiments, the specific binding response were obtained by subtracting form each channel the reference channel response. Moreover, in most cases a double referencing was applied (*i.e.* reference Fc1 channel plus an internal reference standard, namely buffer injection subtraction).

viii.i. Immobilization of glycoprobes

Glycoprobes immobilized on each sensor chip (plus their immobilization level) are indicated on Table 3. Moreover, on each chip, a reference surface (preferentially, first flow cell) was created with a non-specific probe immobilized (*N*[Me]-O-Aoa-GFKKG-

amide or Ac-GFKKG-amide peptides) to subtract the non-specific lectin binding to the dextran surface.

Table 3. Sensor chips prepared during this work.

Sensor chip	Fc channel	Ligand immobilized	Amount immobilized	
			RU	pmol/mm^2 (♯)
1	1	Aoa(m)-GFKKG	500	0.82
	2	Gal-β1,4-GlcNAc-Aoa(m)-GFKKG	120	0.12
	3	Gal-β1,3-GlcNAc-Aoa(m)-GFKKG	124	0.13
	4	Gal-β1,6-GlcNAc-Aoa(m)-GFKKG	146	0.15
2	1	Aoa(m)-GFKKG	819	1.34
	2	Gal-β1,4-Glc-Aoa(m)-GFKKG	126	0.14
3	1	Ac-GFKKG	73	0.13
	2	Galβ1,4GlcNAc-Aoa(m)-GFKKG	28	0.03
	3	Galβ1,3GlcNAc-Aoa(m)-GFKKG	149	0.15
	4	Galβ1,6GlcNAc-Aoa(m)-GFKKG	191	0.20
4	1	Ac-GFKKG	20	0.04
	2	Manα1,2Man-Aoa(m)-GFKKG	300	0.32
	3	Manα1,3Man-Aoa(m)-GFKKG	805	0.87
	4	Manα1,6Man-Aoa(m)-GFKKG	718	0.77
5	1	Ac-GFKKG	195	0.34
	2	Fucα1,3GlcNAc-Aoa(m)-GFKKG	81	0.09
	3	Fucα1,4GlcNAc-Aoa(m)-GFKKG	31	0.03
	4	Fucα1,6GlcNAc-Aoa(m)-GFKKG	296	0.31
6	1	Ac-GFKKG	178	0.31
	2	Siaα2,3lacNAc-Aoa(m)-GFKKG	323	0.26
	3	Siaα2,6lacNAc-Aoa(m)-GFKKG	444	0.35
	4	Siaα2,3/6lac-Aoa(m)-GFKKG	91	0.08
7(⁎)	1	Ac-GFKKG	-	-
	2	EDC/NHS – $EtNH_2$ (‡)	-	-
8	1	EDC/NHS – $EtNH_2$ (‡)	-	-
	2	PDC-109	1645	0.09

Sia= Neu5Ac; lac=Gal-β1,4-Glc; lacNAc= Gal-β1,4-GlcNAc

(♯) We questioned if the manufacturer's suggested equivalence of 1 RU = 1 pg of immobilized analyte, devised for globular proteins, was indeed applicable to our present case, where the ligands are short glycopeptide. To clarify this point, specific refractive index for the peptide was calculated from amino acid compositions by the Lorentz–Lorenz equation[386,387]. Thus, 0.981 RU = 1 pg equivalences were calculated for GFKKG.

(⁎) C1 chip (unique): provides the same functionality as sensor chips CM5 (the rest of used chips) but has no dextran matrix (the carboxyl groups are attached directly to the surface layer).

(‡) A simple activation and deactivation surface control.

For all immobilizations, the carboxyl functionalities of the CM5 sensor chips (or C1 unique chip) surfaces were activated by injecting a mixture of 0.05 M NHS and 0.2 M EDC (1:1 v/v; 50 µL) at 5 µL/min. Then, glycopeptide solutions were passed at 1-2 mg/mL in 10 mM NaOAc for approximately 12 min over the activated surface. Two different pH values were used conveniently (pH 4 *vs* pH 6) for the immobilization process. In the particular case of PC-109 immobilization, protein solution was passed at 0.05 µg/µL in 10 mM NaOAc pH 4 over the activated surface. Afterwards, unreacted groups on all chips surface were blocked by injection of ethanolamine-HCl (1 M; pH 8.5; 60 µL) or H_2O (60 µL). The difference between the resonance units after the surface activation and the final response corresponded to the amount of immobilized glycopeptides (0.981 RU ~ 1 pg/mm^2 for glycopeptides, whereas 1 RU = 1 pg/mm^2 for proteins; see legend Table 3).

viii.ii. Binding and kinetic experiments

All kinetic measurements were carried out by running a manually programmed sequence of binding experiments using several protein range concentrations and two flow rates.

Regeneration of the sensor surface was accomplished in a specific-manner by injecting the complementary carbohydrate (10 mM lac, 100-500 mM αMe-Man, 0.5 M αMe-Fuc or L-Fuc, 0.5 M GlcNAc, 0.5 M Glc, and 10 mM Sia). Alternatively, other procedures such as glycoprotein solutions (*e.g.* fetuin at 2-5 g/L), low pH-buffers (*e.g.* 10 mM glycine pH 1.5-2.5), high salt (*e.g.* 1 M NaCl), high pH or specific chemicals were also used to break the interaction.

Experiments carried out on sensor chips with ECA (chapter 1)

For determination of kinetic parameters, several concentrations in the 66 nM-2.5 µM range were prepared by a set of 2/3-fold dilutions of the most concentrated sample in a running buffer with lower NaCl concentration than standard HBS-P (10 mM HEPES, 25 mM sodium chloride, 5 mM calcium chloride, 1 mM manganese (II) chloride, pH 7.4). Binding experiments were performed at 25ºC at two flow rates (10 and 50 µL/min). After the lectin injection (3-min pulse), sample solution was replaced by running buffer and the dissociation phase was monitored for 6 min. Sensor surface was regenerated

with a 50 µL-injection of 10 mM lactose. Two replicates were performed for each injection.

Experiments carried out on sensor chips with PDC-109 (chapter 2)

For the evaluation of the sensor chips, binding experiments and kinetic studies were carried out with the distinct glycoprobes and specific well-characterized lectins (ECA, Con A, LTA, UEA-I, MAA, and SNA). Several concentrations of each lectin were prepared following a set of 2/3-fold dilutions of the most concentrated sample in running buffer supplemented with $CaCl_2$ (5 mM) and $MnCl_2$ (1 mM). Thus, concentrations varied from 66 nM to 2.5 µM for ECA; from 2.5 to 80 nM for Con A; from 31.25 nM to 1 µM for LTA & UEA-I; from 250 nM to 1.9 µM for MAA and from 74 to 563 nM for SNA. Analyses were performed at 20 µL/min whereas kinetic studies at two flow rates (10 and 40 µL/min). After lectin injection, sample solution was replaced by running buffer and the dissociation phase was monitored for 3-6 min. Sensor surfaces were regenerated with a 10-50 µL-injection of 10 mM lac for ECA; 500 mM αMe-Man for Con A; 0.5 M L-Fuc for LTA & UEA-I and 5 g/L fetuin for MAA & SNA.

For determination of PDC-109 kinetic parameters, several concentrations in the 6.25-200 µM range were prepared by a set of 2/3-fold dilutions of the most concentrated sample in HBS-P running buffer as a standard condition. Other buffer conditions (*i.e.* HBS-P buffer supplemented with $CaCl_2$ (5 mM) and $MnCl_2$ (1 mM), or adding other reagents such as BSA 0.5% carrier protein, CMD 0.5% and PVP 0.5% competitors) were tested to improve SPR sensorgrams and subsequent kinetic fittings. Binding experiments were performed at 25 ºC and 30 µL/min flow. After PDC-109 injection (3-min pulse), sample solution was replaced by running buffer and the dissociation phase was monitored for 6 min. Sensor surface was regenerated with a 60 µL-injection of specific carbohydrate (10 mM lac, 500 mM αMe-Man, 0.5 M αMe-Fuc, and 10 mM Sia) followed by a 60 µL-injection of NaCl 1 M.

For PDC-109 binding behavior studies, CM5 surface *vs* C1 surface without dextran matrix were tested to assess the influence of this structural element. Binding experiments were performed by injecting both PDC-109 untreated and after deglycosylation at 50 µM, 25 ºC, and 20 µL/min flow. After protein injection, the sample solution was replaced by running buffer and the dissociation phase was

monitored for 4 min. Sensor surfaces were regenerated with a 30 µL-injection of NaCl 1 M. Alternatively, PDC-109 aggregation tests were monitored by SPR. Briefly, protein solution of 50 µM PDC-109 was passed over immobilized PDC-109 and reference flow cells at 20 µL/min. After dissociation phase, surface regeneration was accomplished by injection of 30 µL NaCl 1M. Two replicates were performed for each injection in all binding behavior studies.

viii.iii. Thermodynamic experiments

For thermodynamic experiments, the Gal-containing surface and five ECA concentrations (100, 250, 400, 550 and 700 nM) in running buffer were explored in the 10 to 25 ºC interval, with 2.5 ºC increments controlled by a Peltier device. Replicates of each solution were injected over the lacNAc-funcionalized and the reference surface at 50 µL/min. As in the kinetic experiments, after lectin injection sample injection was replaced by running buffer and the dissociation phase was monitored for 6 min. After each cycle, the sensor surface was regenerated with 50-µL injections of 25 mM lactose.

viii.iv. Lectin capture experiments

For recovery experiments, a 1 µM solution of ECA was injected for 3 min at 10 µL/min over the lacNAc-bearing surface at 25 ºC. By means of the MS recover function, captured material was eluted in a 2-µL volume of 10 mM lacNAc (Gal-β1,4-GlcNAc), concentrated by vacuum centrifugation and analyzed by MALDI-TOF MS.

ix. CREDEX-MS

ix.i. CREDEX-MS with purified proteins

Prior the CREDEX-MS experiment, ECA (5 µg), UEA-I (20 µg) and PDC-109 (20 µg) were digested in solution with different proteolytic enzymes. The following incubation conditions were used:

- Trypsin: 0.25 or 1 µg trypsin in 25 mM NH_4HCO_3, pH 8.5 (1:20 enzyme:substrate ratio), overnight at 37 ºC for ECA or UEA-I, respectively. Otherwise, 0.4 µg trypsin in 25 mM NH_4HCO_3, pH 8.5 (1:50 enzyme:substrate ratio), overnight at 37 ºC for PDC-109.

-Chymotrypsin: 0.25 µg chymotrypsin in 100 mM Tris-HCl, pH 7.8, 10 mM $CaCl_2$ (1:20 enzyme:substrate ratio), 24 h at 35 ºC for ECA.

-Glu-C: 0.25 µg Glu-C in PBS pH 7.8 (1:20 enzyme:substrate ratio), 24 h at 35 ºC for ECA.

Resulting peptide mixtures were desalted by reverse phase micro-column (Poros R2) or by ZipTip® Pipette Tips, and analyzed by MALDI-TOF MS.

For disaccharide immobilization, 5 mg of sugar (Gal-β1,4-GlcNAc for ECA; Fuc-α1,(3,4,6)-GlcNAc for UEA-I / PDC-109) was dissolved in 50 µL of 0.5 M K_2CO_3 (pH 11), and the solution was incubated in a microcolumn with 50 µg of dry divinylsulfonyl-activated Sepharose, overnight at r.t. under continuous shaking (800 rpm). Then, the microcolumn was washed sequentially with 50 mM NH_4OAc (pH 4) and 50 mM NH_4HCO_3 (pH 8). Finally, it was equilibrated with binding buffer (10 mM HEPES, 25 mM NaCl, 5 mM $CaCl_2$ and 1 mM $MnCl_2$) and stored at 4 ºC.

Before performing excision experiments, columns were tested for correct binding. Briefly, 20 µg of ECA or UEA-I / PDC-109 were added onto the Gal-β1,4-GlcNAc-sepharose or Fuc-α1,(3,4,6)-GlcNAc-sepharose columns respectively, and were incubated in binding buffer for 24 h at 37 ºC. Unbound materials were removed by extensive washing with running buffer. Bound lectins were eluted with 60% ACN (0.1% TFA), and the protein content of each fractions (flow through; wash and elution) was analyzed by 1D-SDS-PAGE electrophoresis.

For excision experiments, 20 µg of ECA or UEA-I / PDC-109 in running buffer (75 µL) were loaded onto the corresponding microcolumn (Gal-β1,4-GlcNAc-Sepharose or Fucα1,4GlcNAc-Sepharose, respectively) and incubated for 24 h at 37 ºC. Unbound lectin was removed by washing with running buffer until no signal could be detected by MS. Then, the sugar-lectin complex was digested overnight with trypsin (ratios commented above) in 25 mM NH_4HCO_3, pH 8.5 at 37 ºC. After 15 h, the flow through containing digestion products was removed and the column was washed with binding buffer. For ECA, a second chymotrypsin digestion was performed with 1 µg enzyme in 75 µL of 100 mM Tris·HCl, 10 mM $CaCl_2$, pH 7.8 added to the microcolumn and incubated for 24 h at 35 ºC. After washes with running buffer, specific-bound peptides were eluted with 60 µL of 60% ACN (0.1% TFA), concentrated and lyophilized. Prior

to MALDI-TOF analysis, each sample was desalted by means of micro-column purification (Poros R2) or by ZipTip® Pipette Tips.

ix.ii. CREDEX-MS with bovine sperm

Preparation of affinity chromatography columns

For carbohydrate immobilization, 5 mg of each glycotope (Fuc4N, Le^a, 3'-SLN and SLe^x) were dissolved in 50 µL of 0.5 M K_2CO_3 (pH 11) and the solution was incubated with 50 µg of prepared divinylsulfonyl-activated sepharose. The coupling reaction was carried out overnight at room temperature under continuous shaking (800 rpm). The microcolumn was washed sequentially with 50 mM NH_4OAc (pH 4) and 0.1 M Tris (pH 8), and reequilibrated with either CM or NCM depending on the experiment to conduct. For replicate performance, 2 microcolumns for each four different glycotope were prepared (8 columns in total).

Additional microcolumns serving as blanks, without glycan immobilization, were prepared with monofunctionalized methyl vinyl sulfone (MVS) as activating agent instead of DVS [50 µg Sepharose beads, 5 µL MVS in 50 µL of 0.5 M K_2CO_3 (pH 11), 70 min at room temperature under stirring]. Blank microcolumns were also washed and equilibrated with either culture media as glycan microcolumns.

CREDEX-MS excision experiments

For the excision experiments, ~1.5×10^6 sperm cells were loaded immediately after swim-up treatment (with either CM or NCM) on the microcolumns and incubated for 24 h at 37 ºC under continuous stirring. The flow-through from each column, containing unbound sperm, was collected and the column washed with the corresponding medium until only residual spermatozoa were observed by microscopy. Sugar-lectin complexes were then digested overnight with trypsin (150 µg/mL) in 25 mM NH_4HCO_3, pH 7.8 at 37 ºC under stirring. After digestion, each column flow-through, containing non-specific digestion products, was removed and columns washed again with culture media. After gentle washing until no spermatozoa were observed, specific-bound peptides were eluted (2 × 300 µL ACN: H_2O (2:1) 0.1% TFA, 15 min, 37 ºC, stirring). In excision experiments with Sia-containing columns (3'-SLN and SLe^x glycoprobes), second competitive elutions were done with 400 µL of 0.5 mM fetuin for 15 min, 37 ºC,

with stirring. Fetuin was removed by filtration (30 KDa centrifugal filters) and all elution samples were lyophilized prior to LC-MS/MS analysis.

In total, 3 replicates were carried out with each 2 columns prepared for each four different glycotope (6 replicates per each glycoprobe) in the two different capacitating conditions successively. Moreover, 6 experimental blank replicates were done for each condition (3 replicates with each 2 blank columns). Thus, 30 CREDEX-MS excision experiments (24 with carbohydrate columns and 6 with blank columns) were carried out for each CM and NCM condition.

Column functionality verification

Before performing the excision experiments, between replicates and at the end of each replication set, the different glyco-columns were tested to confirm their functionality through binding tests with specific, pure, lectins. Briefly, 20 µg of specific lectin (UEA-I for Fuc4N; LTA for Le^a; MAA for 3'-SLN and AAL for SLe^x) was added onto the corresponding carbohydrate-sepharose column and incubated in 100 µL HEPES running buffer (10 mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$ and 1 mM $MnCl_2$, pH 7.4) for 24 h at 37 ºC. Unbound material was removed by extensive washing with running buffer. Bound lectin was eluted with ACN:H_2O (2:1) 0.1% TFA, but exceptionally for Sia-containing columns (3'-SLN and SLe^x glycoprobes), a second competitive elution was performed with 1 mM fetuin. The protein content of each fraction (flow through; wash and elution) were analyzed by 1D-SDS-PAGE electrophoresis, and column functionality preservation was confirmed by the gel band detection of the specific lectin in the elution fraction.

Besides, after each excision experiment, a hydration (HEPES buffer) - dehydration (ACN:H_2O (2:1) 0.1% TFA) washing cycle was carried out to eliminate BSA excess from the CM or to eliminate residual fetuin used as a competitor in elution fractions with sialyl-containing columns, thus leaving the columns suitable for their reuse.

x. Shotgun proteomics

Proteomic analyses were carried out with all the elution fractions from all CREDEX experiments and with sperm trypsinization samples (including all the replicates).

In order to evaluate instrumental reproducibility for each blank biological replicate, 3 analyses (3 analytical replicates) were performed reaching a total of 18 blank replicates per each capacitating condition.

In total, 54 sample injections (12 fetuin glycoprobe elutions, 24 standard glycoprobe elutions and 18 blank elutions) per each sperm condition (CM and NCM) were performed for MS/MS analysis of CREDEX experiments. In order to minimize instrumental variability, a defined batch file was programmed for sample injection. Specifically, samples were injected in groups of 9 including the six replicates of the same glycotope alternated with 3 sample blanks. Additionally, 6 sample injections of the triplicates of sperm trypsinization experiments per condition were also injected separately.

Prior to injection, lyophilized samples containing tryptic peptides were resuspended in 200 mM NH_4HCO_3, reduced with dithiothreitol (60 nmol, 1 h, 37 °C), alkylated in the dark with iodoacetamide (120 nmol, 30 min, 25 ºC) and purified using UltraMicroSpin C18 columns (The Nest Group, Inc, Southborough, MA , USA). Finally, desalted and purified peptides were dried in a vacuum centrifuge and redissolved in H_2O (0.1% HCOOH) for subsequent MS analysis.

x.i. Nano-LC-MS/MS

Samples were analyzed using a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EasyLC (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded directly onto the analytical column at 1.5-2 µL/min using a wash-volume of 4 to 5 times the injection volume and were separated by reversed-phase chromatography using a 12 cm column with an inner diameter of 75 µm, packed with 3 µm C18 particles (Nikkyo Technos Co., Ltd. Japan). Chromatographic gradients started at 97% buffer A / 3% buffer B with a flow rate of 300 nL/min, and gradually increased to 93% buffer A / 7% buffer B in 1 min, and to 65% buffer A / 35% buffer B in 40 min. After each analysis, the column was washed for

10 min with 10% buffer A / 90% buffer B (buffer A: 0.1% formic acid in water; buffer B: 0.1% formic acid in acetonitrile).

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.2 kV and source temperature at 325 °C. Ultramark 1621 for the FT mass analyzer was used for external calibration prior the analyses. Moreover, an internal calibration was also performed using the background polysiloxane ion signal at m/z 445.1200. The instrument was operated in data dependent acquisition (DDA) mode and full MS scans at resolution of 60.000 were used over a mass range of m/z 30-2000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1×10^6, dynamic exclusion (60 seconds) and charge state filtering disqualifying singly charged peptides was activated. In each cycle of DDA analysis, following each survey scan the top ten most intense ions with multiple charged ions above a threshold ion count of 10000 were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via high-energy collision dissociation (HCD) were acquired in the Orbitrap mass analyzer with a resolution setting of 7500, AGC was set to 5×10^4, isolation window of 2.0 *m/z*, activation time of 0.1 ms and maximum injection time of 100 ms was used. All data were acquired with Xcalibur software v2.2.

x.ii. Database searching and dataset composition

Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific) and the Mascot search engine (v2.3.1, Matrix Science[300]) were used for peptide identification and quantification. The data were searched against an in-house generated database containing all *Bos taurus (Bovine)* entries from UniProtKB/Swiss-Prot database (released March 2013; 6,121 protein entries) and a list of common contaminants (around 599 entries in total). A precursor ion mass tolerance of 7 ppm at the MS1 level was used, and up to three missed cleavages for trypsin were allowed. The fragment ion mass tolerance was set to 20 mmu. Met oxidation, N-terminal acetylation, and Ser, Thr and Tyr phosphorylation were defined as variable modifications whereas Cys carbamidomethylation was set as a fix modification. False discovery rate (FDR) in peptide identification was evaluated by using a decoy database and was set to a maximum of 5%.

x.iii. Data mining and bioinformatic analyses

Script tasks and comparisons of protein lists were performed using the R software (http://www.R-project.org) and for plotting data in a graphical matrix, the package 'gplots' 2.11.3 version was used. To generate the 0/1 matrix for the presence-absence protein analysis, identified proteins were sorted according to sample replicates, following the criteria:

BLANK REPLICATES (n = 18): If peptide/protein peak areas $\neq 0$ in $n \geq 12$ replicates, the identified protein was considered (matrix value = 1).

GLYCOPROBE REPLICATES (n = 6): If peptide/protein peak areas $\neq 0$ in $n \geq 4$ replicates, the identified protein is considered (matrix value = 1).

SPERM TRYPSINIZATION REPLICATES (n = 3): If peptide/protein peak areas $\neq 0$ in $n \geq 2$ replicates, the identified protein is considered (matrix value = 1).

Subsequent hierarchical clustering was performed following a Euclidean distance metric and maximum linkage criteria, and final heat maps/dendrograms were represented displaying X-axis (clustering of samples) *vs* Y-axis (clustering of identified proteins). Proportioned Venn diagrams were drawn using Venn Diagram Plotter (http://omics.pnl.gov/software/VennDiagramPlotter.php).

Proteins identified were compared against a database generated with recently published studies on bovine sperm proteomics. PubMed (http://www.ncbi.nlm.nih.gov/pubmed/; National Center for Biotechnology Information, U.S. National Library of Medicine, National Institutes of Health, Bethesda, MD) and UniProt Knowledgebase (UniProtKB/Swiss-Prot) were also used, whenever needed, especially to check whether each of the proteins had been previously described in bovine sperm and to find information about the proteins identified. Comparisons were done using either the Swiss-Prot accession number (when available in the analyzed literature) or the names of the proteins (in this case, all the alternative names of a single protein were verified).

Gene Ontology (GO) resources and tools available at agriGO (Gene Ontology analysis toolkit and database for agricultural community) were used to obtain all existing GO annotations available for known proteins in our datasets as well as to identify the molecular functions, biological processes and cellular locations represented in differentially expressed proteins in our datasets. DAVID informatics was also employed to provide a comprehensive set of functional annotations of our list of proteins/genes.

Moreover, the KEGG PATHWAY database was used to map our datasets for biological interpretation of higher-level systemic functions. Thus, an overrepresentation analysis was performed in order to recognize functional biological pathways and to wire diagrams of molecular interactions, reactions, and relations.

Protein-interaction properties of the identified proteins were predicted using the Search Tool for Retrieval of Interacting Genes/Proteins (STRING v 9.05) database of physical and functional interactions. Network analysis was set at medium stringency (STRING score = 0.4). Proteins were linked based on seven criteria; neighbourhood, gene fusion, cooccurrence, co-expression, experimental evidences, existing databases and text mining.

Finally, human orthologs of bovine genes were batch downloaded from Ensembl 65 using BioMart (http://www.ensembl.org/biomart/martview/).

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Spiro, R.G. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002**, 12, 43R-56R.

2. Cummings, R.D. The repertoire of glycan determinants in the human glycome. *Mol Biosyst.* **2009**, 5, 1087-1104.

3. Paulson, J.C., Blixt, O., and Collins, B.E. Sweet spots in functional glycomics. *Nat Chem Biol* **2006**, 2, 238-248.

4. Moremen, K.W., Tiemeyer, M., and Nairn, A.V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **2012**, 13, 448-462.

5. Joao, H.C. and Dwek, R.A. Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its individual glycoforms. *Eur. J. Biochem.* **1993**, 218, 239-244.

6. Rudd, P.M., Morgan, B.P., Wormald, M.R., Harvey, D.J., van den Berg, C.W., Davis, S.J., Ferguson, M.A., and Dwek, R.A. The glycosylation of the complement regulatory protein, human erythrocyte CD59. *J. Biol. Chem.* **1997**, 272, 7229-7244.

7. Varki, A. Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb. Perspect. Biol.* **2011**, 3.

8. Hart, G.W. and Copeland, R.J. Glycomics hits the big time. *Cell* **2010**, 143, 672-676.

9. Haltiwanger, R.S. and Lowe, J.B. Role of glycosylation in development. *Annu. Rev. Biochem.* **2004**, 73, 491-537.

10. Ohtsubo, K. and Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **2006**, 126, 855-867.

11. Ruhaak, L.R., Uh, H.W., Beekman, M., Koeleman, C.A., Hokke, C.H., Westendorp, R.G., Wuhrer, M., Houwing-Duistermaat, J.J., Slagboom, P.E., and Deelder, A.M. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. *PLoS One.* **2010**, 5, e12566.

12. Pucic, M., Muzinic, A., Novokmet, M., Skledar, M., Pivac, N., Lauc, G., and Gornik, O. Changes in plasma and IgG N-glycome during childhood and adolescence. *Glycobiology* **2012**, 22, 975-982.

13. Wang, J., Balog, C.I., Stavenhagen, K., Koeleman, C.A., Scherer, H.U., Selman, M.H., Deelder, A.M., Huizinga, T.W., Toes, R.E., and Wuhrer, M. Fc-glycosylation of IgG1 is modulated by B-cell stimuli. *Mol Cell Proteomics.* **2011**, 10, M110-

14. Springer, S.A. and Gagneux, P. Glycan evolution in response to collaboration, conflict, and constraint. *J Biol Chem* **2013**, 288, 6904-6911.

15. Lauc, G. and Zoldos, V. Protein glycosylation--an evolutionary crossroad between genes and environment. *Mol Biosyst.* **2010**, 6, 2373-2379.

16. Zoldos, V., Novokmet, M., Beceheli, I., and Lauc, G. Genomics and epigenomics of the human glycome. *Glycoconj. J.* **2013**, 30, 41-50.

17. Tharmalingam, T., Adamczyk, B., Doherty, M.A., Royle, L., and Rudd, P.M. Strategies for the profiling, characterisation and detailed structural analysis of N-linked oligosaccharides. *Glycoconj. J.* **2013**, 30, 137-146.

18. Rakus, J.F. and Mahal, L.K. New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome. *Annual Review of Analytical Chemistry* **2011**, 4, 367-392.

19. Vanderschaeghe, D., Festjens, N., Delanghe, J., and Callewaert, N. Glycome profiling using modern glycomics technology: technical aspects and applications. *Biol Chem* **2010**, 391, 149-161.

20. Mariño, K., Bones, J., Kattla, J.J., and Rudd, P.M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* **2010**, 6, 713-723.

21. Pabst, M. and Altmann, F. Glycan analysis by modern instrumental methods. *Proteomics* **2011**, 11, 631-643.

22. Yamada, K. and Kakehi, K. Recent advances in the analysis of carbohydrates for biomedical use. *Journal of Pharmaceutical and Biomedical Analysis* **2011**, 55, 702-727.

23. Zaia, J. Capillary electrophoresis-mass spectrometry of carbohydrates. *Methods Mol Biol* **2013**, 984, 13-25.

24. Mechref, Y. Analysis of glycans derived from glycoconjugates by capillary electrophoresis-mass spectrometry. *Electrophoresis* **2011**, 32, 3467-3481.

25. Vanderschaeghe, D., Guttman, A., and Callewaert, N. High-throughput profiling of the serum N-glycome on capillary electrophoresis microfluidics systems. *Methods Mol Biol* **2013**, 919, 87-96.

26. Hua, S., Lebrilla, C., and An, H.J. Application of nano-LC-based glycomics towards biomarker discovery. *Bioanalysis.* **2011**, 3, 2573-2585.

27. Behan, J.L. and Smith, K.D. The analysis of glycosylation: a continued need for high pH anion exchange chromatography. *Biomed. Chromatogr.* **2011**, 25, 39-46.

28. Wuhrer, M. Glycomics using mass spectrometry. *Glycoconj. J.* **2013**, 30, 11-22.

29. Lazar, I.M., Lee, W., and Lazar, A.C. Glycoproteomics on the rise: Established methods, advanced techniques, sophisticated biological applications. *Electrophoresis* **2013**, 34, 113-125.

30. Zaia, J. Mass Spectrometry and Glycomics. *OMICS: A Journal of Integrative Biology* **2010**, 14, 401-418.

31. Leymarie, N. and Zaia, J. Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis. *Anal. Chem.* **2012**, 84, 3040-3048.

32. Smith, D.F. and Cummings, R.D. Application of microarrays for deciphering the structure and function of the human glycome. *Mol Cell Proteomics* **2013**, 12, 902-912.

33. Gupta, G., Surolia, A., and Sampathkumar, S.G. Lectin microarrays for glycomic analysis. *OMICS: A Journal of Integrative Biology* **2010**, 14, 419-436.

34. Smith, D.F., Song, X., and Cummings, R.D. Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol.* **2010**, 480, 417-444.

35. Park, S., Gildersleeve, J.C., Blixt, O., and Shin, I. Carbohydrate microarrays. *Chem Soc. Rev* **2012**, 42, 4310-4326.

36. Rillahan, C.D. and Paulson, J.C. Glycan Microarrays for Decoding the Glycome. *Annu. Rev. Biochem.* **2011**, 80, 797-823.

37. Pasing, Y., Sickmann, A., and Lewandrowski, U. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biol Chem* **2012**, 393, 249-258.

38. Nilsson, J., Halim, A., Grahn, A., and Larson, G. Targeting the glycoproteome. *Glycoconj. J* **2013**, 30, 119-136.

39. Artemenko, N.V., Campbell, M.P., and Rudd, P.M. GlycoExtractor: a web-based interface for high throughput processing of HPLC-glycan data. *J. Proteome Res.* **2010**, 9, 2037-2041.

40. Woodin, C.L., Maxon, M., and Desaire, H. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *Analyst* **2013**, 138, 2793-2803.

41. Porter, A., Yue, T., Heeringa, L., Day, S., Suh, E., and Haab, B.B. A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology* **2010**, 20, 369-380.

42. Kletter, D., Singh, S., Bern, M., and Haab, B.B. Global comparisons of lectin-glycan interactions using a database of analyzed glycan array data. *Mol Cell Proteomics* **2013**, 12, 1026-1035.

43. Bennun, S.V., Yarema, K.J., Betenbaugh, M.J., and Krambeck, F.J. Integration of the transcriptome and glycome for identification of glycan cell signatures. *PLoS. Comput. Biol.* **2013**, 9, e1002813.

44. der Lieth, C.W., Freire, A.A., Blank, D., Campbell, M.P., Ceroni, A., Damerell, D.R., Dell, A., Dwek, R.A., Ernst, B., Fogh, R., Frank, M., Geyer, H., Geyer, R., Harrison, M.J., Henrick, K., Herget, S., Hull, W.E., Ionides, J., Joshi, H.J., Kamerling, J.P., Leeflang, B.R., Lütteke, T., Lundborg, M., Maass, K., Merry, A., Ranzinger, R., Rosen, J., Royle, L., Rudd, P.M., Schloissnig, S., Stenutz, R., Vranken, W.F., Widmalm, G., and Haslam, S.M. EUROCarbDB: An open-access platform for glycoinformatics. *Glycobiology* **2011**, 21, 493-502.

45. Chen, M.M., Bartlett, A.I., Nerenberg, P.S., Friel, C.T., Hackenberger, C.P., Stultz, C.M., Radford, S.E., and Imperiali, B. Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A* **2010**, 107, 22528-22533.

46. Banks, D.D. The effect of glycosylation on the folding kinetics of erythropoietin. *J. Mol. Biol.* **2011**, 412, 536-550.

47. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993**, 3, 97-130.

48. Dwek, R.A. Glycobiology: toward understanding the function of sugars. *Chem. Rev.* **1996**, 96, 683-720.

49. Jaeken, J. Congenital disorders of glycosylation. *Handb. Clin. Neurol.* **2013**, 113, 1737-1743.

50. Muthana, S.M., Campbell, C.T., and Gildersleeve, J.C. Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chem. Biol.* **2012**, 7, 31-43.

51. Adamczyk, B., Tharmalingam, T., and Rudd, P.M. Glycans as cancer biomarkers. *Biochim. Biophys. Acta* **2012**, 1820, 1347-1353.

52. Chandler, K.B. and Goldman, R. Glycoprotein disease markers and single protein-omics. *Mol Cell Proteomics* **2013**, 12, 836-845.

53. Mechref, Y., Hu, Y., Garcia, A., and Hussein, A. Identifying cancer biomarkers by mass spectrometry-based glycomics. *Electrophoresis* **2012**, 33, 1755-1767.

54. Angata, T., Fujinawa, R., Kurimoto, A., Nakajima, K., Kato, M., Takamatsu, S., Korekane, H., Gao, C.X., Ohtsubo, K., Kitazume, S., and Taniguchi, N. Integrated approach toward the discovery of glyco-biomarkers of inflammation-related diseases. *Ann. N. Y. Acad. Sci.* **2012**, 1253, 159-169.

55. An, H.J. and Lebrilla, C.B. A glycomics approach to the discovery of potential cancer biomarkers. *Methods Mol. Biol.* **2010**, 600, 199-213.

56. Zoldos, V., Horvat, T., and Lauc, G. Glycomics meets genomics, epigenomics and other high throughput omics for system biology studies. *Curr. Opin. Chem. Biol.* **2012**, 17, 34-40.

57. Neelamegham, S. and Liu, G. Systems glycobiology: biochemical reaction networks regulating glycan structure and function. *Glycobiology* **2011**, 21, 1541-1553.

58. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E. *Essentials of Glycobiology* **2009** (2nd edition), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

59. Avci, F.Y. and Kasper, D.L. How bacterial carbohydrates influence the adaptive immune system. *Annu. Rev Immunol.* **2010**, 28, 107-130.

60. Ghazarian, H., Idoni, B., and Oppenheimer, S.B. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **2011**, 113, 236-247.

61. Boyd, W.C. and Shapleigh, E. Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science* **1954**, 119, 419.

62. Stillmark, H. *Ricin, ein giftiges Ferment.* Thesis **1988**

63. Sumner, J.B., Howell, S.F., and Zeissig, A. Concanavalin A and hemagglutination. *Science* **1935**, 82, 65-66.

64. Watkins, W.M. and Morgan, W.T. Neutralization of the anti-H agglutinin in eel serum by simple sugars. *Nature* **1952**, 169, 825-826.

65. Sharon, N. and Lis, H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **2004**, 14, 53R-62R.

66. Drickamer, K. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J Biol Chem* **1988**, 263, 9557-9560.

67. Lis, H. and Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **1998**, 98, 637-674.

68. Wu, A.M., Lisowska, E., Duk, M., and Yang, Z. Lectins as tools in glycoconjugate research. *Glycoconj. J.* **2009**, 26, 899-913.

69. Rudiger, H. and Gabius, H.J. Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj. J* **2001**, 18, 589-613.

70. Chandra, N.R., Kumar, N., Jeyakani, J., Singh, D.D., Gowda, S.B., and Prathima, M.N. Lectindb: a plant lectin database. *Glycobiology* **2006**, 16, 938-946.

71. Weis, W.I. and Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **1996**, 65, 441-473.

72. Elgavish, S. and Shaanan, B. Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.* **1997**, 22, 462-467.

73. Quiocho, F.A. Probing the atomic interactions between proteins and carbohydrates. *Biochem. Soc. Trans.* **1993**, 21, 442-448.

74. Garcia-Pino, A., Buts, L., Wyns, L., and Loris, R. Interplay between metal binding and cis/trans isomerization in legume lectins: structural and thermodynamic study of P. angolensis lectin. *J. Mol. Biol.* **2006**, 361, 153-167.

75. Kogelberg, H., Solís, D., and Jiménez-Barbero, J. New structural insights into carbohydrate-protein interactions from NMR spectroscopy. *Curr. Opin. Struct. Biol.* **2003**, 13, 646-653.

76. Lundquist, J.J. and Toone, E.J. The cluster glycoside effect. *Chem. Rev.* **2002**, 102, 555-578.

77. Lee Y.C. and Lee R.T. Carbohydrate-Protein Interactions: Basis of Glycobiology. *Acc. Chem. Res.* **1995**, 28, 321-327.

78. Shinohara, Y., Kim, F., Shimizu, M., Goto, M., Tosu, M., and Hasegawa, Y. Kinetic measurement of the interaction between an oligosaccharide and lectins by a biosensor based on surface plasmon resonance. *Eur. J. Biochem.* **1994**, 223, 189-194.

79. Shinohara, Y., Hasegawa, Y., Kaku, H., and Shibuya, N. Elucidation of the mechanism enhancing the avidity of lectin with oligosaccharides on the solid phase surface. *Glycobiology* **1997**, 7, 1201-1208.

80. Monsigny, M., Mayer, R., and Roche, A.C. Sugar-lectin interactions: sugar clusters, lectin multivalency and avidity. *Carbohydr. Lett.* **2000**, 4, 35-52.

81. Maierhofer, C., Rohmer, K., and Wittmann, V. Probing multivalent carbohydrate-lectin interactions by an enzyme-linked lectin assay employing covalently immobilized carbohydrates. *Bioorg. Med. Chem.* **2007**, 15, 7661-7676.

82. Mo, H., Meah, Y., Moore, J.G., and Goldstein, I.J. Purification and characterization of Dolichos lablab lectin. *Glycobiology* **1999**, 9, 173-179.

83. Park, S., Lee, M.R., Pyo, S.J., and Shin, I. Carbohydrate chips for studying high-throughput carbohydrate-protein interactions. *J. Am. Chem. Soc.* **2004**, 126, 4812-4819.

84. Palmer, R.A. and Niwa, H. X-ray crystallographic studies of protein-ligand interactions. *Biochem. Soc. Trans.* **2003**, 31, 973-979.

85. Bewley, C.A. and Shahzad-Ul-Hussan, S. Characterizing carbohydrate-protein interactions by nuclear magnetic resonance spectroscopy. *Biopolymers* **2013**, 99, 796-806.

86. Aboitiz, N., Vila-Perelló, M., Groves, P., Asensio, J.L., Andreu, D., Cañada, F.J., and Jiménez-Barbero, J. NMR and modeling studies of protein-carbohydrate interactions: synthesis,

three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *Chembiochem.* **2004**, 5, 1245-1255.

87. Roldos, V., Canada, F.J., and Jiménez-Barbero, J. Carbohydrate-protein interactions: a 3D view by NMR. *Chembiochem* **2011**, 12, 990-1005.

88. Dam, T.K. and Brewer, C.F. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **2002**, 102, 387-429.

89. Gourdine, J.P., Cioci, G., Miguet, L., Unverzagt, C., Silva, D.V., Varrot, A., Gautier, C., Smith-Ravin, E.J., and Imberty, A. High affinity interaction between a bivalve C-type lectin and a biantennary complex-type N-glycan revealed by crystallography and microcalorimetry. *J. Biol. Chem.* **2008**, 283, 30112-30120.

90. Safina, G. Application of surface plasmon resonance for the detection of carbohydrates, glycoconjugates, and measurement of the carbohydrate-specific interactions: a comparison with conventional analytical techniques. A critical review. *Anal Chim. Acta* **2012**, 712, 9-29.

91. Cheng, C.I., Chang, Y.P., and Chu, Y.H. Biomolecular interactions and tools for their recognition: focus on the quartz crystal microbalance and its diverse surface chemistries and applications. *Chem Soc. Rev* **2012**, 41, 1947-1971.

92. Jiménez-Castells, C., de la Torre, B., Andreu, D., and Gutiérrez-Gallego, R. Neo-glycopeptides: the importance of sugar core conformation in oxime-linked glycoprobes for interaction studies. *Glycoconjugate Journal* **2008**, 25, 879-887.

93. Vila-Perelló, M., Gutiérrez Gallego, R., and Andreu, D. A simple approach to well-defined sugar-coated surfaces for interaction studies. *ChemBioChem* **2005**, 6, 1831-1838.

94. Mori, T., Toyoda, M., Ohtsuka, T., and Okahata, Y. Kinetic analyses for bindings of concanavalin A to dispersed and condensed mannose surfaces on a quartz crystal microbalance. *Anal. Biochem.* **2009**, 395, 211-216.

95. Day, Y.S.N., Baird, C.L., Rich, R.L., and Myszka, D.G. Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. *Protein Science* **2002**, 11, 1017-1025.

96. Liedberg, B., Nylander, C., and Lünstrom, I. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators* **1983**, 4, 299-304.

97. Couture, M., Zhao, S.S., and Masson, J.F. Modern surface plasmon resonance for bioanalytics and biophysics. *Phys. Chem Chem Phys.* **2013**, 15, 11190-11216.

98. www.biacore.com **2010**

99. Biacore Sensor Surface Handbook **2003**

100. Roos, H., Karlsson, R., Nilshans, H., and Persson, A. Thermodynamic analysis of protein interactions with biosensor technology. *J. Mol. Recognit.* **1998**, 11, 204-210.

101. Mattei, B., Borch, J., and Roepstorff., P. Peer Reviewed: Biomolecular Interaction Analysis and MS. *Anal. Chem.* **2004**, 76, 18-25.

102. Nelson, R.W. and Krone, J.R. Advances in surface plasmon resonance biomolecular interaction analysis mass spectrometry (BIA/MS). *J. Mol. Recognit.* **1999**, 12, 77-93.

103. Sönksen, C.P., Nordhoff, E., Jansson, O., Malmqvist, M., and Roepstorff, P. Combining MALDI mass spectrometry and biomolecular interaction analysis using a biomolecular interaction analysis instrument. *Anal. Chem.* **1998**, 70, 2731-2736.

104. Zhukov, A., Schurenberg, M., Jansson, O., Areskoug, D., and Buijs, J. Integration of surface plasmon resonance with mass spectrometry: automated ligand fishing and sample preparation for MALDI MS using a Biacore 3000 biosensor. *J. Biomol. Tech.* **2004**, 15, 112-119.

105. Bellon, S., Buchmann, W., Gonnet, F., Jarroux, N., Anger-Leroy, M., Guillonneau, F., and Daniel, R. Hyphenation of surface plasmon resonance imaging to matrix-assisted laser desorption ionization mass spectrometry by on-chip mass spectrometry and tandem mass spectrometry analysis. *Anal. Chem.* **2009**, 81, 7695-7702.

106. Natsume, T., Nakayama, H., Jansson, O., Isobe, T., Takio, K., and Mikoshiba, K. Combination of biomolecular interaction analysis and mass spectrometric amino acid sequencing. *Anal. Chem.* **2000**, 72, 4193-4198.

107. Nelson, R.W., Nedelkov, D., and Tubbs, K.A. Biosensor chip mass spectrometry: a chip-based proteomics approach. *Electrophoresis* **2000**, 21, 1155-1163.

108. Spiga, O., Bernini, A., Scarselli, M., Ciutti, A., Bracci, L., Lozzi, L., Lelli, B., Di Maro, D., Calamandrei, D., and Niccolai, N. Peptide-protein interactions studied by surface plasmon and nuclear magnetic resonances. *FEBS Lett.* **2002**, 511, 33-35.

109. Stockley, P.G. and Persson, B. Surface plasmon resonance assays of DNA-protein interactions. *Methods Mol. Biol.* **2009**, 543, 653-669.

110. Besenicar, M., Macek, P., Lakey, J.H., and Anderluh, G. Surface plasmon resonance in protein-membrane interactions. *Chemistry and Physics of Lipids* **2006**, 141, 169-178.

111. Duverger, E., Frison, N., Roche, A.C., and Monsigny, M. Carbohydrate-lectin interactions assessed by surface plasmon resonance. *Biochimie* **2003**, 85, 167-179.

112. Haseley, S.R., Talaga, P., Kamerling, J.P., and Vliegenthart, J.F. Characterization of the carbohydrate binding specificity and kinetic parameters of lectins by using surface plasmon resonance. *Anal. Biochem.* **1999**, 274, 203-210.

113. Gutiérrez Gallego R., Haseley, S.R., van Miegem, V.F., Vliegenthart, J.F., and Kamerling, J.P. Identification of carbohydrates binding to lectins by using surface plasmon resonance in combination with HPLC profiling. *Glycobiology* **2004**, 14, 373-386.

114. Hutchinson, A.M. Characterization of glycoprotein oligosaccharides using surface plasmon resonance. *Anal. Biochem.* **1994**, 220, 303-307.

115. O'Shannessy, D.J. and Wilchek, M. Immobilization of glycoconjugates by their oligosaccharides: use of hydrazido-derivatized matrices. *Anal. Biochem.* **1990**, 191, 1-8.

116. Vornholt, W., Hartmann, M., and Keusgen, M. SPR studies of carbohydrate-lectin interactions as useful tool for screening on lectin sources. *Biosens. Bioelectron.* **2007**, 22, 2983-2988.

117. Otamiri, M. and Nilsson, K.G. Analysis of human serum antibody-carbohydrate interaction using biosensor based on surface plasmon resonance. *Int. J. Biol. Macromol.* **1999**, 26, 263-268.

118. Linman, M.J., Taylor, J.D., Yu, H., Chen, X., and Cheng, Q. Surface plasmon resonance study of protein-carbohydrate interactions using biotinylated sialosides. *Anal. Chem.* **2008**, 80, 4007-4013.

119. Critchley, P. and Clarkson, G.J. Carbohydrate-protein interactions at interfaces: comparison of the binding of Ricinus communis lectin to two series of synthetic glycolipids using surface plasmon resonance studies. *Org. Biomol. Chem.* **2003**, 1, 4148-4159.

120. Zeng, X., Murata, T., Kawagishi, H., Usui, T., and Kobayashi, K. Analysis of specific interactions of synthetic glycopolypeptides carrying N-acetyllactosamine and related compounds with lectins. *Carbohydr. Res.* **1998**, 312, 209-217.

121. Gondran, C., Dubois, M.P., Fort, S., Cosnier, S., and Szunerits, S. Detection of carbohydrate-binding proteins by oligosaccharide-modified polypyrrole interfaces using electrochemical surface plasmon resonance. *Analyst* **2008**, 133, 206-212.

122. Tiefenbrunn, T.K. and Dawson, P.E. Chemoselective ligation techniques: modern applications of time-honored chemistry. *Biopolymers* **2010**, 94, 95-106.

123. Lee, D.J., Mandal, K., Harris, P.W., Brimble, M.A., and Kent, S.B. A one-pot approach to neoglycopeptides using orthogonal native chemical ligation and click chemistry. *Org. Lett.* **2009**, 11, 5270-5273.

124. Peri, F., Dumy, P., and Mutter, M. Chemo- and stereoselective glycosylation of hydroxylamino derivatives: A versatile approach to glycoconjugates. *Tetrahedron* **1998**, 54, 12269-12278.

125. Lee, M.R. and Shin, I. Facile preparation of carbohydrate microarrays by site-specific, covalent immobilization of unmodified carbohydrates on hydrazide-coated glass slides. *Org. Lett.* **2005**, 7, 4269-4272.

126. Gudmundsdottir, A.V., Paul, C.E., and Nitz, M. Stability studies of hydrazide and hydroxylamine-based glycoconjugates in aqueous solution. *Carbohydr. Res.* **2009**, 344, 278-284.

127. Carotenuto, A., Alcaro, M.C., Saviello, M.R., Peroni, E., Nuti, F., Papini, A.M., Novellino, E., and Rovero, P. Designed glycopeptides with different beta-turn types as synthetic probes for the detection of autoantibodies as biomarkers of multiple sclerosis. *J. Med. Chem.* **2008**, 51, 5304-5309.

128. El Hawiet, A., Kitova, E.N., and Klassen, J.S. Quantifying carbohydrate-protein interactions by electrospray ionization mass spectrometry analysis. *Biochemistry* **2012**, 51, 4244-4253.

129. Zdanov, A., Li, Y., Bundle, D.R., Deng, S.J., MacKenzie, C.R., Narang, S.A., Young, N.M., and Cygler, M. Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. *Proc. Natl. Acad. Sci. U. S. A* **1994**, 91, 6423-6427.

130. Kitova, E.N., Kitov, P.I., Bundle, D.R., and Klassen, J.S. The observation of multivalent complexes of Shiga-like toxin with globotriaoside and the determination of their stoichiometry by nanoelectrospray Fourier-transform ion cyclotron resonance mass spectrometry. *Glycobiology* **2001**, 11, 605-611.

131. El Hawiet, A., Kitova, E.N., and Klassen, J.S. Quantifying protein interactions with isomeric carbohydrate ligands using a catch and release electrospray ionization-mass spectrometry assay. *Anal. Chem.* **2013**, 85, 7637-7644.

132. Palcic, M.M., Zhang, B., Qian, X., Rempel, B., and Hindsgaul, O. Evaluating carbohydrate-protein binding interactions using frontal affinity chromatography coupled to mass spectrometry. *Methods Enzymol.* **2003**, 362, 369-376.

133. Suckau, D., Kohl, J., Karwath, G., Schneider, K., Casaretto, M., Bitter-Suermann, D., and Przybylski, M. Molecular epitope identification by limited proteolysis of an immobilized antigen-antibody complex and mass spectrometric peptide mapping. *Proc. Natl. Acad. Sci. U. S. A* **1990**, 87, 9848-9852.

134. Zhao, Y., Muir, T.W., Kent, S.B., Tischer, E., Scardina, J.M., and Chait, B.T. Mapping protein-protein interactions by affinity-directed mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A* **1996**, 93, 4020-4024.

135. Hochleitner, E.O., Gorny, M.K., Zolla-Pazner, S., and Tomer, K.B. Mass spectrometric characterization of a discontinuous epitope of the HIV envelope protein HIV-gp120 recognized by the human monoclonal antibody 1331A. *J. Immunol.* **2000**, 164, 4156-4161.

136. Przybylski, M., Moise, A., Siebert, H., and Gabius, H. CREDEX-MS: Molecular elucidation of carbohydrate recognition peptides in lectins and related proteins by proteolytic excision-mass spectrometry. *J. Pept. Sci.* **2008**, Supplement to Vol. 14, 40.

137. Przybylski, M., Moise, A., and Gabius, H. Identification of ligand recognition domains **2009**, EP2009/003495.

138. Moise, A., André, S., Eggers, F., Krzeminski, M., Przybylski, M., and Gabius, H.J. Toward bioinspired galectin mimetics: identification of ligand-contacting peptides by proteolytic-excision mass spectrometry. *J. Am. Chem. Soc.* **2011**, 133, 14844-14847.

139. López-Lucendo, M.F., Solís, D., André, S., Hirabayashi, J., Kasai, K., Kaltner, H., Gabius, H.J., and Romero, A. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J. Mol. Biol.* **2004**, 343, 957-970.

140. Satoh, A., Fukui, E., Yoshino, S., Shinoda, M., Kojima, K., and Matsumoto, I. Comparison of methods of immobilization to enzyme-linked immunosorbent assay plates for the detection of sugar chains. *Anal. Biochem.* **1999**, 275, 231-235.

141. Pepper, D.S. Some alternative coupling chemistries for affinity chromatography. *Mol. Biotechnol.* **1994**, 2, 157-178.

142. Hatakeyama, T., Murakami, K., Miyamoto, Y., and Yamasaki, N. An assay for lectin activity using microtiter plate with chemically immobilized carbohydrates. *Anal. Biochem.* **1996**, 237, 188-192.

143. Parker, C.E. and Tomer, K.B. MALDI/MS-based epitope mapping of antigens bound to immobilized antibodies. *Mol. Biotechnol.* **2002**, 20, 49-62.

144. Ulrich, S., Boturyn, D., Marra, A., Renaudet, O., and Dumy, P. Oxime ligation: a chemoselective click-type reaction for accessing multifunctional biomolecular constructs. *Chemistry.* **2014**, 20, 34-41.

145. Jiménez-Castells, C., de la Torre, B.G., Gutiérrez Gallego, R., and Andreu, D. Optimized synthesis of aminooxy-peptides as glycoprobe precursors for surface-based sugar–protein interaction studies. *Bioorganic & Medicinal Chemistry Letters* **2007**, 17, 5155-5158.

146. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **1995**, 4, 2411-2423.

147. Itakura, Y., Nakamura-Tsuruta, S., Kominami, J., Sharon, N., Kasai, K., and Hirabayashi, J. Systematic comparison of oligosaccharide specificity of Ricinus communis agglutinin I and Erythrina lectins: a search by frontal affinity chromatography. *J. Biochem.* **2007**, 142, 459-469.

148. Elgavish, S. and Shaanan, B. Structures of the Erythrina corallodendron lectin and of its complexes with mono- and disaccharides. *J. Mol. Biol.* **1998**, 277, 917-932.

149. Gupta, D., Cho, M., Cummings, R.D., and Brewer, C.F. Thermodynamics of carbohydrate binding to galectin-1 from Chinese hamster ovary cells and two mutants. A comparison with four galactose-specific plant lectins. *Biochemistry* **1996**, 35, 15236-15243.

150. De Boeck, H., Loontiens, F.G., Lis, H., and Sharon, N. Binding of simple carbohydrates and some N-acetyllactosamine-containing oligosaccharides to Erythrina cristagalli agglutinin as followed with a fluorescent indicator ligand. *Arch. Biochem. Biophys.* **1984**, 234, 297-304.

151. Sharma, S., Bharadwaj, S., Surolia, A., and Podder, S.K. Evaluation of the stoichiometry and energetics of carbohydrate binding to Ricinus communis agglutinin: a calorimetric study. *Biochem. J.* **1998**, 333 (Pt 3), 539-542.

152. Bonneil, E., Young, N.M., Lis, H., Sharon, N., and Thibault, P. Probing genetic variation and glycoform distribution in lectins of the Erythrina genus by mass spectrometry. *Arch. Biochem. Biophys.* **2004**, 426, 241-249.

153. Svensson, C., Teneberg, S., Nilsson, C.L., Kjellberg, A., Schwarz, F.P., Sharon, N., and Krengel, U. High-resolution crystal structures of Erythrina cristagalli lectin in complex with lactose and 2'-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J. Mol. Biol.* **2002**, 321, 69-83.

154. Shivaji, S., Scheit, K.H., and Bhargava, P.M. *Proteins of Seminal Plasma* **1990** (book chapter), New York: Wiley.

155. Aitken, R.J. and Nixon, B. Sperm capacitation: a distant landscape glimpsed but unexplored. *Mol. Hum. Reprod.* **2013**, 19, 785-793.

156. Bailey, J.L. Factors Regulating Sperm Capacitation. *Syst Biol Reprod Med* **2010**, 56, 334-348.

157. Yanagimachi, R. *Mammalian fertilization* **1994**, (2nd edition), New York: Raven Press Ltd, 189-317.

158. Amann, R.P., Hammerstedt, R.H., and Veeramachaneni, D.N. The epididymis and sperm maturation: a perspective. *Reprod. Fertil. Dev.* **1993**, 5, 361-381.

159. Rodríguez-Martínez, H., Kvist, U., Ernerudh, J., Sanz, L., and Calvete, J.J. Seminal plasma proteins: what role do they play? *Am. J. Reprod. Immunol.* **2011**, 66 Suppl 1, 11-22.

160. Calvete, J.J. and Sanz, L. Insights into structure-function correlations of ungulate seminal plasma proteins. *Soc. Reprod. Fertil. Suppl* **2007**, 65, 201-215.

161. Caballero, I., Parrilla, I., Almiñana, C., del Olmo, D., Roca, J., Martínez, E.A., and Vázquez, J.M. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. *Reproduction in Domestic Animals* **2012**, 47, 12-21.

162. Manjunath, P. and Sairam, M.R. Purification and biochemical characterization of three major acidic proteins (BSP-A1, BSP-A2 and BSP-A3) from bovine seminal plasma. *Biochem. J.* **1987**, 241, 685-692.

163. Kosanovic, M.M. and Jankovic, M.M. Molecular heterogeneity of gelatin-binding proteins from human seminal plasma. *Asian J. Androl* **2010**, 12, 363-375.

164. Leblond, E., Desnoyers, L., and Manjunath, P. Phosphorylcholine-binding proteins from the seminal fluids of different species share antigenic determinants with the major proteins of bovine seminal plasma. *Mol. Reprod. Dev.* **1993**, 34, 443-449.

165. Calvete, J.J., Mann, K., Schäfer, W., Sanz, L., Reinert, M., Nessau, S., Raida, M., and Töpfer-Petersen, E. Amino acid sequence of HSP-1, a major protein of stallion seminal plasma: effect of glycosylation on its heparin- and gelatin-binding capabilities. *Biochem. J.* **1995**, 310 (Pt 2), 615-622.

166. Calvete, J.J., Raida, M., Gentzel, M., Urbanke, C., Sanz, L., and Töpfer-Petersen, E. Isolation and characterization of heparin- and phosphorylcholine-binding proteins of boar and stallion seminal plasma. Primary structure of porcine pB1. *FEBS Lett.* **1997**, 407, 201-206.

167. Nauc, V. and Manjunath, P. Radioimmunoassays for bull seminal plasma proteins (BSP-A1/-A2, BSP-A3, and BSP-30-Kilodaltons), and their quantification in seminal plasma and sperm. *Biol. Reprod.* **2000**, 63, 1058-1066.

168. Desnoyers, L., Therien, I., and Manjunath, P. Characterization of the major proteins of bovine seminal fluid by two-dimensional polyacrylamide gel electrophoresis. *Mol. Reprod. Dev.* **1994**, 37, 425-435.

169. Esch, F.S., Ling, N.C., Bohlen, P., Ying, S.Y., and Guillemin, R. Primary structure of PDC-109, a major protein constituent of bovine seminal plasma. *Biochem. Biophys. Res. Commun.* **1983**, 113, 861-867.

170. Desnoyers, L. and Manjunath, P. Major proteins of bovine seminal plasma exhibit novel interactions with phospholipid. *J. Biol. Chem.* **1992**, 267, 10149-10155.

171. Manjunath, P., Marcel, Y.L., Uma, J., Seidah, N.G., Chretien, M., and Chapdelaine, A. Apolipoprotein A-I binds to a family of bovine seminal plasma proteins. *J. Biol. Chem.* **1989**, 264, 16853-16857.

172. Parrish, J.J., Susko-Parrish, J., Winer, M.A., and First, N.L. Capacitation of bovine sperm by heparin. *Biol. Reprod.* **1988**, 38, 1171-1180.

173. Thérien, I., Bleau, G., and Manjunath, P. Phosphatidylcholine-binding proteins of bovine seminal plasma modulate capacitation of spermatozoa by heparin. *Biol. Reprod.* **1995**, 52, 1372-1379.

174. Bergeron, A., Crete, M.H., Brindle, Y., and Manjunath, P. Low-density lipoprotein fraction from hen's egg yolk decreases the binding of the major proteins of bovine seminal plasma to sperm and prevents lipid efflux from the sperm membrane. *Biol. Reprod.* **2004**, 70, 708-717.

175. Calvete, J.J., Raida, M., Sanz, L., Wempe, F., Scheit, K.H., Romero, A., and Töpfer-Petersen, E. Localization and structural characterization of an oligosaccharide O-linked to

bovine PDC-109 Quantitation of the glycoprotein in seminal plasma and on the surface of ejaculated and capacitated spermatozoa. *FEBS Letters* **1994**, 350, 203-206.

176. Calvete, J.J., Varela, P.F., Sanz, L., Romero, A., Mann, K., and Töpfer-Petersen, E. A procedure for the large-scale isolation of major bovine seminal plasma proteins. *Protein Expression and Purification* **1996**, 8, 48-56.

177. Gerwig, G.L., Calvete, J.J., Töpfer-Petersen, E., and Vliegenthart, J.F. The structure of the O-linked carbohydrate chain of bovine seminal plasma protein PDC-109 revised by H-NMR spectroscopy A correction. *FEBS Lett.* **1996**, 387, 99-100.

178. Wah, D.A., Fernández-Tornero, C., Sanz, L., Romero, A., and Calvete, J.J. Sperm coating mechanism from the 1.8A crystal structure of PDC-109-phosphorylcholine complex. *Structure* **2002**, 10, 505-514.

179. Constantine, K.L., Madrid, M., Banyai, L., Trexler, M., Patthy, L., and Llinas, M. Refined solution structure and ligand-binding properties of PDC-109 domain b. A collagen-binding type II domain. *J. Mol. Biol.* **1992**, 223, 281-298.

180. Kim, H.J., Choi, M.Y., Kim, H.J., and Llinás, M. Conformational dynamics and ligand binding in the multi-domain protein PDC109. *PLoS ONE* **2010**, 5, e9180.

181. Gasset, M., Saiz, J.L., Laynez, J., Sanz, L., Gentzel, M., Töpfer-Petersen, E., and Calvete, J.J. Conformational features and thermal stability of bovine seminal plasma protein PDC-109 oligomers and phosphorylcholine-bound complexes. *European Journal of Biochemistry* **1997**, 250, 735-744.

182. Romero, A., Varela, P.F., Töpfer-Petersen, E., and Calvete, J.J. Crystallization and preliminary X-ray diffraction analysis of bovine seminal plasma PDC-109, a protein composed of two fibronectin type II domains. *Proteins* **1997**, 28, 454-456.

183. Calvete, J.J., Campanero-Rhodes, M., Raida, M., and Sanz, L. Characterisation of the conformational and quaternary structure-dependent heparin-binding region of bovine seminal plasma protein PDC-109. *FEBS Letters* **1999**, 444, 260-264.

184. Sankhala, R.S. and Swamy, M.J. The major protein of bovine seminal plasma, PDC-109, is a molecular chaperone. *Biochemistry* **2010**, 49, 3908-3918.

185. Laitaoja, M., Sankhala, R.S., Swamy, M.J., and Jänis, J. Top–down mass spectrometry reveals new sequence variants of the major bovine seminal plasma protein PDC-109. *J. Mass. Spectrom.* **2012**, 47, 853-859.

186. Aumüller, G., Vesper, M., Seitz, J., Kemme, M., and Scheit, K.H. Binding of a major secretory protein from bull seminal vesicles to bovine spermatozoa. *Cell Tissue Res.* **1988**, 252, 377-384.

187. Thérien, I., Moreau, R., and Manjunath, P. Bovine seminal plasma phospholipid-binding proteins stimulate phospholipid efflux from epididymal sperm. *Biol. Reprod.* **1999**, 61, 590-598.

188. Thérien, I., Moreau, R., and Manjunath, P. Major proteins of bovine seminal plasma and high-density lipoprotein induce cholesterol efflux from epididymal sperm. *Biol. Reprod.* **1998**, 59, 768-776.

189. Gwathmey, T.M., Ignotz, G.G., and Suarez, S.S. PDC-109 (BSP-A1/A2) promotes bull sperm binding to oviductal epithelium in vitro and may be involved in forming the oviductal sperm reservoir. *Biol. Reprod.* **2003**, 69, 809-815.

190. Sánchez-Luengo, S., Aumüller, G., Albrecht, M., Sen, P.C., Röhm, K., and Wilhelm, B. Interaction of PDC-109, the major secretory protein from bull seminal vesicles, with bovine sperm membrane Ca2+-ATPase. *J. Androl* **2004**, 25, 234-244.

191. Thérien, I., Soubeyrand, S., and Manjunath, P. Major proteins of bovine seminal plasma modulate sperm capacitation by high-density lipoprotein. *Biol. Reprod.* **1997**, 57, 1080-1088.

192. Lane, M., Therien, I., Moreau, R., and Manjunath, P. Heparin and high-density lipoprotein mediate bovine sperm capacitation by different mechanisms. *Biol. Reprod.* **1999**, 60, 169-175.

193. Sankhala, R.S., Damai, R.S., Anbazhagan, V., Kumar, C.S., Bulusu, G., and Swamy, M.J. Biophysical investigations on the interaction of the major bovine seminal plasma protein, PDC-109, with heparin. *J. Phys. Chem. B* **2011**, 115, 12954-12962.

194. Fiol, d.C., Vincenti, L.M., Martini, A.C., Ponce, A.A., and Ruiz, R.D. Effects of PDC-109 on bovine sperm functional activity in presence or absence of heparin. *Theriogenology* **2004**, 62, 207-216.

195. Damai, R.S., Sankhala, R.S., Anbazhagan, V., and Swamy, M.J. 31P NMR and AFM studies on the destabilization of cell and model membranes by the major bovine seminal plasma protein, PDC-109. *IUBMB. Life* **2010**, 62, 841-851.

196. Srivastava, N., Srivastava, S.K., Ghosh, S.K., Singh, L.P., Prasad, J.K., Kumar, A., Perumal, P., Jerome, A., and Thamizharasan, A. Sequestration of PDC-109 protein improves freezability of crossbred bull spermatozoa. *Anim Reprod. Sci.* **2012**, 131, 54-62.

197. Manjunath, P. and Therien, I. Role of seminal plasma phospholipid-binding proteins in sperm membrane lipid modification that occurs during capacitation. *J Reprod Immunol.* **2002**, 53, 109-119.

198. Srivastava, N., Jerome, A., Srivastava, S.K., Ghosh, S.K., and Kumar, A. Bovine seminal PDC-109 protein: An overview of biochemical and functional properties. *Anim Reprod. Sci.* **2013**, 138, 1-13.

199. Boerke, A., Tsai, P.S., Garcia-Gil, N., Brewis, I.A., and Gadella, B.M. Capacitation-dependent reorganization of microdomains in the apical sperm head plasma membrane: functional relationship with zona binding and the zona-induced acrosome reaction. *Theriogenology* **2008**, 70, 1188-1196.

200. Suarez, S.S. Formation of a reservoir of sperm in the oviduct. *Reprod Domest. Anim* **2002**, 37, 140-143.

201. Talevi, R. and Gualtieri, R. Molecules involved in sperm-oviduct adhesion and release. *Theriogenology* **2010**, 73, 796-801.

202. Lefebvre, R., Lo, M.C., and Suarez, S.S. Bovine sperm binding to oviductal epithelium involves fucose recognition. *Biol. Reprod.* **1997**, 56, 1198-1204.

203. Suarez, S.S., Revah, I., Lo, M., and Kolle, S. Bull sperm binding to oviductal epithelium is mediated by a Ca2+-dependent lectin on sperm that recognizes Lewis-a trisaccharide. *Biol. Reprod.* **1998**, 59, 39-44.

204. Revah, I., Gadella, B.M., Flesch, F.M., Colenbrander, B., and Suarez, S.S. Physiological State of Bull Sperm Affects Fucose- and Mannose-Binding Properties. *Biol. Reprod.* **2000**, 62, 1010-1015.

205. Ignotz, G.G., Lo, M.C., Perez, C.L., Gwathmey, T.M., and Suarez, S.S. Characterization of a fucose-binding protein from bull sperm and seminal plasma that may be responsible for formation of the oviductal sperm reservoir. *Biol. Reprod.* **2001**, 64, 1806-1811.

206. Thygesen, M.B., Munch, H., Sauer, J., Cló, E., Jorgensen, M.R., Hindsgaul, O., and Jensen, K.J. Nucleophilic catalysis of carbohydrate oxime formation by anilines. *J Org. Chem* **2010**, 75, 1752-1755.

207. Carrasco, M.R., Silva, O., Rawls, K.A., Sweeney, M.S., and Lombardo, A.A. Chemoselective alkylation of N-alkylaminooxy-containing peptides. *Org. Lett.* **2006**, 8, 3529-3532.

208. Mandal, D.K., Kishore, N., and Brewer, C.F. Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry* **1994**, 33, 1149-1156.

209. Pereira, M.E. and Kabat, E.A. Blood group specificity of the lectin from Lotus tetragonolobus. *Ann. N. Y. Acad. Sci.* **1974**, 234, 301-305.

210. Audette, G.F., Olson, D.J., Ross, A.R., Quail, J.W., and Delbaere, L.T. Examination of the structural basis for O(H) blood group specificity by Ulex europaeus Lectin I. *Can. J. Chem.* **2002**, 80, 1010-1021.

211. Knibbs, R.N., Goldstein, I.J., Ratcliffe, R.M., and Shibuya, N. Characterization of the carbohydrate binding specificity of the leukoagglutinating lectin from Maackia amurensis. Comparison with other sialic acid-specific lectins. *J. Biol. Chem.* **1991**, 266, 83-88.

212. Duverger, E., Coppin, A., Strecker, G., and Monsigny, M. Interaction between lectins and neoglycoproteins containing new sialylated glycosynthons. *Glycoconj. J.* **1999**, 16, 793-800.

213. Thomas, C.J., Anbazhagan, V., Ramakrishnan, M., Sultan, N., Surolia, I., and Swamy, M.J. Mechanism of membrane binding by the bovine seminal plasma protein, PDC-109: A Surface Plasmon Resonance study. *Biophysical Journal* **2003**, 84, 3037-3044.

214. Gasset, M., Magdaleno, L., and Calvete, J.J. Biophysical study of the perturbation of model membrane structure caused by seminal plasma protein PDC-109. *Archives of Biochemistry and Biophysics* **2000**, 374, 241-247.

215. Damai, R.S., Anbazhagan, V., Rao, K.B., and Swamy, M.J. Fluorescence studies on the interaction of choline-binding domain B of the major bovine seminal plasma protein, PDC-109 with phospholipid membranes. *Biochim. Biophys. Acta* **2009**, 1794, 1725-1733.

216. Sankhala, R.S., Damai, R.S., and Swamy, M.J. Correlation of membrane binding and hydrophobicity to the chaperone-like activity of PDC-109, the major protein of bovine seminal plasma. *PLoS ONE* **2011**, 6, e17330.

217. Carrasco, M.R., Nguyen, M.J., Burnell, D.R., MacLaren, M.D., and Hengel, S.M. Synthesis of neoglycopeptides by chemoselective reaction of carbohydrates with peptides containing a novel N '-methyl-aminooxy amino acid. *Tetrahedron Letters* **2002**, 43, 5727-5729.

218. Matsubara, N., Oiwa, K., Hohsaka, T., Sadamoto, R., Niikura, K., Fukuhara, N., Takimoto, A., Kondo, H., and Nishimura, S. Molecular design of glycoprotein mimetics: glycoblotting by engineered proteins with an oxylamino-functionalized amino acid residue. *Chemistry* **2005**, 11, 6974-6981.

219. Suarez, S.S. The oviductal sperm reservoir in mammals: mechanisms of formation. *Biol. Reprod.* **1998**, 58, 1105-1107.

220. Coy, P., García-Vázquez, F.A., Visconti, P.E., and Avilés, M. Roles of the oviduct in mammalian fertilization. *Reproduction* **2012**, 144, 649-660.

221. Holt, W.V. and Fazeli, A. The oviduct as a complex mediator of mammalian sperm function and selection. *Mol. Reprod. Dev.* **2010**, 77, 934-943.

222. Suarez, S.S. Regulation of sperm storage and movement in the mammalian oviduct. *Int. J Dev Biol* **2008**, 52, 455-462.

223. Burkitt, M., Walker, D., Romano, D.M., and Fazeli, A. Using computational modeling to investigate sperm navigation and behavior in the female reproductive tract. *Theriogenology* **2012**, 77, 703-716.

224. Coy, P., Cánovas, S., Mondéjar, I., Saavedra, M.D., Romar, R., Grullón, L., Matás, C., and Avilés, M. Oviduct-specific glycoprotein and heparin modulate sperm-*zona pellucida* interaction during fertilization and contribute to the control of polyspermy. *Proc. Natl. Acad. Sci. U. S. A* **2008**, 105, 15809-15814.

225. Avilés, M., Gutiérrez-Adán, A., and Coy, P. Oviductal secretions: will they be key factors for the future ARTs? *Mol Hum Reprod* **2010**, 16, 896-906.

226. Coy, P. and Avilés, M. What controls polyspermy in mammals, the oviduct or the oocyte? *Biol Rev Camb Philos. Soc.* **2010**, 85, 593-605.

227. Dell, A., Morris, H.R., Easton, R.L., Patankar, M., and Clark, G.F. The glycobiology of gametes and fertilization. *Biochim. Biophys. Acta* **1999**, 1473, 196-205.

228. Töpfer-Petersen, E. Carbohydrate-based interactions on the route of a spermatozoon to fertilization. *Human Reproduction Update* **1999**, 5, 314-329.

229. Diekman, A.B. Glycoconjugates in sperm function and gamete interactions: how much sugar does it take to sweet-talk the egg? *Cell Mol Life Sci.* **2003**, 60, 298-308.

230. Benoff, S. Carbohydrates and fertilization: an overview. *Mol Hum Reprod* **1997**, 3, 599-637.

231. Töpfer-Petersen, E. Molecules on the sperm's route to fertilization. *J Exp. Zool.* **1999**, 285, 259-266.

232. Zhu, X. and Naz, R.K. Fertilization antigen-1: cDNA cloning, testis-specific expression, and immunocontraceptive effects. *Proc. Natl. Acad. Sci. U. S. A* **1997**, 94, 4704-4709.

233. Hunter, R.H., Nichol, R., and Crabtree, S.M. Transport of spermatozoa in the ewe: timing of the establishment of a functional population in the oviduct. *Reprod Nutr. Dev* **1980**, 20, 1869-1875.

234. Hunter, R.H. Sperm release from oviduct epithelial binding is controlled hormonally by peri-ovulatory graafian follicles. *Mol Reprod Dev* **2008**, 75, 167-174.

235. Suarez, S.S. Sperm transport and motility in the mouse oviduct: observations in situ. *Biol. Reprod.* **1987**, 36, 203-210.

236. Smith, T.T. and Yanagimachi, R. Attachment and release of spermatozoa from the caudal isthmus of the hamster oviduct. *Journal of Reproduction and Fertility* **1991**, 91, 567-573.

237. Hunter, R.H., Barwise, L., and King, R. Sperm transport, storage and release in the sheep oviduct in relation to the time of ovulation. *Br. Vet J* **1982**, 138, 225-232.

238. Hunter, R.H. Pre-ovulatory arrest and peri-ovulatory redistribution of competent spermatozoa in the isthmus of the pig oviduct. *J Reprod Fertil* **1984**, 72, 203-211.

239. Hunter, R.H. and Wilmut, I. Sperm transport in the cow: peri-ovulatory redistribution of viable cells within the oviduct. *Reprod Nutr. Dev* **1984**, 24, 597-608.

240. Lefebvre, R., Chenoweth, P.J., Drost, M., LeClear, C.T., MacCubbin, M., Dutton, J.T., and Suarez, S.S. Characterization of the oviductal sperm reservoir in cattle. *Biol. Reprod.* **1995**, 53, 1066-1074.

241. Thomas, P.G., Ball, B.A., and Brinsko, S.P. Interaction of equine spermatozoa with oviduct epithelial cell explants is affected by estrous cycle and anatomic origin of explant. *Biol. Reprod.* **1994**, 51, 222-228.

242. Pacey, A.A., Hill, C.J., Scudamore, I.W., Warren, M.A., Barratt, C.L., and Cooke, I.D. The interaction in vitro of human spermatozoa with epithelial cells from the human uterine (fallopian) tube. *Hum Reprod* **1995**, 10, 360-366.

243. Suarez, S.S. Carbohydrate-mediated formation of the oviductal sperm reservoir in mammals. *Cells Tissues. Organs* **2001**, 168, 105-112.

244. DeMott, R.P., Lefebvre, R., and Suarez, S.S. Carbohydrates mediate the adherence of hamster sperm to oviductal epithelium. *Biol. Reprod.* **1995**, 52, 1395-1403.

245. Cortés, P.P., Orihuela, P.A., Zúñiga, L.M., Velásquez, L.A., and Croxatto, H.B. Sperm Binding to Oviductal Epithelial Cells in the Rat: Role of Sialic Acid Residues on the Epithelial Surface and Sialic Acid-Binding Sites on the Sperm Surface. *Biol. Reprod.* **2004**, 71, 1262-1269.

246. Dobrinski, I., Ignotz, G.G., Thomas, P.G., and Ball, B.A. Role of carbohydrates in the attachment of equine spermatozoa to uterine tubal (oviductal) epithelial cells in vitro. *Am J Vet Res* **1996**, 57, 1635-1639.

247. Ekhlasi-Hundrieser, M., Gohr, K., Wagner, A., Tsolova, M., Petrunkina, A., and Töpfer-Petersen, E. Spermadhesin AQN1 is a candidate receptor molecule involved in the formation of the oviductal sperm reservoir in the pig. *Biol. Reprod.* **2005**, 73, 536-545.

248. Kadirvel, G., Machado, S.A., Korneli, C., Collins, E., Miller, P., Bess, K.N., Aoki, K., Tiemeyer, M., Bovin, N., and Miller, D.J. Porcine Sperm Bind to Specific 6-Sialylated Biantennary Glycans to Form the Oviduct Reservoir. *Biol. Reprod.* **2012**, 87, 147.

249. Apichela, S.A., Valz-Gianinet, J.N., Schuster, S., Jiménez-Díaz, M.A., Roldán-Olarte, E.M., and Miceli, D.C. Lectin binding patterns and carbohydrate mediation of sperm binding to llama oviductal cells in vitro. *Anim Reprod. Sci.* **2010**, 118, 344-353.

250. Gwathmey, T.M., Ignotz, G.G., Mueller, J.L., Manjunath, P., and Suarez, S.S. Bovine seminal plasma proteins PDC-109, BSP-A3, and BSP-30-kDa share functional roles in storing sperm in the oviduct. *Biol. Reprod.* **2006**, 75, 501-507.

251. Hung, P.H. and Suarez, S.S. Alterations to the bull sperm surface proteins that bind sperm to oviductal epithelium. *Biol. Reprod.* **2012**, 87, 88.

252. Ignotz, G.G., Cho, M.Y., and Suarez, S.S. Annexins are candidate oviductal receptors for bovine sperm surface proteins and thus may serve to hold bovine sperm in the oviductal reservoir. *Biol. Reprod.* **2007**, 77, 906-913.

253. Lefebvre, R. and Suarez, S.S. Effect of capacitation on bull sperm binding to homologous oviductal epithelium. *Biol. Reprod.* **1996**, 54, 575-582.

254. Chian, R.C., Lapointe, S., and Sirard, M.A. Capacitation in vitro of bovine spermatozoa by oviduct epithelial cell monolayer conditioned medium. *Mol. Reprod. Dev.* **1995**, 42, 318-324.

255. Mahmoud, A.I. and Parrish, J.J. Oviduct fluid and heparin induce similar surface changes in bovine sperm during capacitation: A flow cytometric study using lectins. *Mol. Reprod. Dev.* **1996**, 43, 554-560.

256. Talevi, R. and Gualtieri, R. Sulfated Glycoconjugates Are Powerful Modulators of Bovine Sperm Adhesion and Release from the Oviductal Epithelium In Vitro. *Biol. Reprod.* **2001**, 64, 491-498.

257. Talevi, R., Zagami, M., Castaldo, M., and Gualtieri, R. Redox Regulation of Sperm Surface Thiols Modulates Adhesion to the Fallopian Tube Epithelium. *Biol. Reprod.* **2007**, 76, 728-735.

258. Gualtieri, R., Mollo, V., Duma, G., and Talevi, R. Redox control of surface protein sulphhydryls in bovine spermatozoa reversibly modulates sperm adhesion to the oviductal epithelium and capacitation. *Reproduction* **2009**, 138, 33-43.

259. Eisenbach, M. and Tur-Kaspa, I. Do human eggs attract spermatozoa? *Bioessays* **1999**, 21, 203-210.

260. Chang, H. and Suarez, S.S. Rethinking the relationship between hyperactivation and chemotaxis in mammalian sperm. *Biol. Reprod.* **2010**, 83, 507-513.

261. Kim, E., Baba, D., Kimura, M., Yamashita, M., Kashiwabara, S., and Baba, T. Identification of a hyaluronidase, Hyal5, involved in penetration of mouse sperm through cumulus mass. *Proc. Natl. Acad. Sci. U. S. A* **2005**, 102, 18028-18033.

262. Schubert, C. WORLD OF REPRODUCTIVE BIOLOGY: Only One Sperm Allowed. *Biol. Reprod.* **2012** (DOI: 10.1095/biolreprod.112.101568).

263. Snell, W.J. and White, J.M. The molecules of mammalian fertilization. *Cell* **1996**, 85, 629-637.

264. Stetson, I., Izquierdo-Rico, M.J., Moros, C., Chevret, P., Lorenzo, P.L., Ballesta, J., Rebollar, P.G., Gutiérrez-Gallego, R., and Avilés, M. Rabbit *zona pellucida* composition: a molecular, proteomic and phylogenetic approach. *J. Proteomics.* **2012**, 75, 5920-5935.

265. Florman, H.M. and Wassarman, P.M. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell* **1985**, 41, 313-324.

266. Yurewicz, E.C., Sacco, A.G., Gupta, S.K., Xu, N., and Gage, D.A. Hetero-oligomerization-dependent binding of pig oocyte *zona pellucida* glycoproteins ZPB and ZPC to boar sperm membrane vesicles. *J. Biol. Chem.* **1998**, 273, 7488-7494.

267. Kanai, S., Yonezawa, N., Ishii, Y., Tanokura, M., and Nakano, M. Recombinant bovine *zona pellucida* glycoproteins ZP3 and ZP4 coexpressed in Sf9 cells form a sperm-binding active hetero-complex. *FEBS J.* **2007**, 274, 5390-5405.

268. Yonezawa, N., Kanai-Kitayama, S., Kitayama, T., Hamano, A., and Nakano, M. Porcine *zona pellucida* glycoprotein ZP4 is responsible for the sperm-binding activity of the ZP3/ZP4 complex. *Zygote.* **2012**, 20, 389-397.

269. Baibakov, B., Boggs, N.A., Yauger, B., Baibakov, G., and Dean, J. Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *J Cell Biol* **2012**, 197, 897-905.

270. Clark, G.F. The mammalian zona pellucida: a matrix that mediates both gamete binding and immune recognition? *Syst Biol Reprod Med* **2010**, 56, 349-364.

271. Gupta, S.K., Bhandari, B., Shrestha, A., Biswal, B.K., Palaniappan, C., Malhotra, S.S., and Gupta, N. Mammalian *zona pellucida* glycoproteins: structure and function during fertilization. *Cell Tissue Res* **2012**, 349, 665-678.

272. Wassarman, P.M., Jovine, L., and Litscher, E.S. A profile of fertilization in mammals. *Nat Cell Biol* **2001**, 3, E59-E64.

273. Hoodbhoy, T. and Dean, J. Insights into the molecular basis of sperm-egg recognition in mammals. *Reproduction* **2004**, 127, 417-422.

274. Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., and Dean, J. Gamete recognition in mice depends on the cleavage status of an egg's *zona pellucida* protein. *Science* **2010**, 329, 216-219.

275. Rankin, T.L., Coleman, J.S., Epifano, O., Hoodbhoy, T., Turner, S.G., Castle, P.E., Lee, E., Gore-Langton, R., and Dean, J. Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Dev. Cell* **2003**, 5, 33-43.

276. Burkart, A.D., Xiong, B., Baibakov, B., Jiménez-Movilla, M., and Dean, J. Ovastacin, a cortical granule protease, cleaves ZP2 in the *zona pellucida* to prevent polyspermy. *J. Cell Biol.* **2012**, 197, 37-44.

277. Visconti, P.E. and Florman, H.M. Mechanisms of sperm-egg interactions: between sugars and broken bonds. *Sci. Signal.* **2010**, 3, e35.

278. Clark, G.F. Molecular models for mouse sperm-oocyte binding. *Glycobiology* **2011**, 21, 3-5.

279. Kate, A.R., Aitken, R.J., and Brett, N. *More than a simple lock and key mechanism: unraveling the intricacies of sperm-zona pellucida binding* **2012** (book chapter), Binding Protein,PhD. Kotb Abdelmohsen (Ed.), ISBN:978-953-51-0758-3, InTech, DOI:10.5772/50499.

280. Avella, M.A., Xiong, B., and Dean, J. The molecular basis of gamete recognition in mice and humans. *Mol Hum Reprod* **2013**, 19, 279-289.

281. Clark, G.F. The role of carbohydrate recognition during human sperm-egg binding. *Hum Reprod* **2013**, 28, 566-577.

282. Brewis, I.A. and Wong, C.H. Gamete recognition: sperm proteins that interact with the egg zona pellucida. *Rev Reprod* **1999**, 4, 135-142.

283. Noguchi, S., Yonezawa, N., Katsumata, T., Hashizume, K., Kuwayama, M., Hamano, S., Watanabe, S., and Nakano, M. Characterization of the *zona pellucida* glycoproteins from bovine ovarian and fertilized eggs. *Biochim. Biophys. Acta* **1994**, 1201, 7-14.

284. Yonezawa, N., Fukui, N., Kuno, M., Shinoda, M., Goko, S., Mitsui, S., and Nakano, M. Molecular cloning of bovine *zona pellucida* glycoproteins ZPA and ZPB and analysis for sperm-binding component of the zona. *European Journal of Biochemistry* **2001**, 268, 3587-3594.

285. Katsumata, T., Noguchi, S., Yonezawa, N., Tanokura, M., and Nakano, M. Structural characterization of the N-linked carbohydrate chains of the *zona pellucida* glycoproteins from bovine ovarian and fertilized eggs. *Eur. J Biochem.* **1996**, 240, 448-453.

286. Amari, S., Yonezawa, N., Mitsui, S., Katsumata, T., Hamano, S., Kuwayama, M., Hashimoto, Y., Suzuki, A., Takeda, Y., and Nakano, M. Essential role of the nonreducing terminal alpha-mannosyl residues of the N-linked carbohydrate chain of bovine *zona pellucida* glycoproteins in sperm-egg binding. *Mol Reprod Dev* **2001**, 59, 221-226.

287. Sakaguchi, Y., Iwata, H., Kuwayama, T., and Monji, Y. Effect of N-acetyl-D-glucosamine on bovine sperm-oocyte interactions. *J Reprod Dev* **2009**, 55, 676-684.

288. Velásquez, J.G., Canovas, S., Barajas, P., Marcos, J., Jiménez-Movilla, M., Gallego, R.G., Ballesta, J., Avilés, M., and Coy, P. Role of sialic acid in bovine sperm-*zona pellucida* binding. *Mol Reprod Dev* **2007**, 74, 617-628.

289. Pang, P.C., Chiu, P.C., Lee, C.L., Chang, L.Y., Panico, M., Morris, H.R., Haslam, S.M., Khoo, K.H., Clark, G.F., Yeung, W.S., and Dell, A. Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. *Science* **2011**, 333, 1761-1764.

290. Waters, S.I. and White, J.M. Biochemical and molecular characterization of bovine fertilin alpha and beta (ADAM 1 and ADAM 2): a candidate sperm-egg binding/fusion complex. *Biol. Reprod.* **1997**, 56, 1245-1254.

291. Coonrod, S.A., Westhusin, M.E., and Naz, R.K. Monoclonal antibody to human fertilization antigen-1 (FA-1) inhibits bovine fertilization in vitro: application in immunocontraception. *Biol. Reprod.* **1994**, 51, 14-23.

292. Fayrer-Hosken, R.A., Caudle, A.B., and Shur, B.D. Galactosyltransferase activity is restricted to the plasma membranes of equine and bovine sperm. *Mol. Reprod. Dev.* **1991**, 28, 74-78.

293. Martin-DeLeon, P.A. Epididymal SPAM1 and its impact on sperm function. *Mol. Cell Endocrinol.* **2006**, 250, 114-121.

294. Morin, G., Lalancette, C., Sullivan, R., and Leclerc, P. Identification of the bull sperm p80 protein as a PH-20 ortholog and its modification during the epididymal transit. *Mol. Reprod. Dev.* **2005**, 71, 523-534.

295. Töpfer-Petersen, E., Romero, A., Varela, P.F., Ekhlasi-Hundrieser, M., Dostàlovà, Z., Sanz, L., and Calvete, J.J. Spermadhesins: a new protein family. Facts, hypotheses and perspectives. *Andrologia* **1998**, 30, 217-224.

296. Tardif, S., Wilson, M.D., Wagner, R., Hunt, P., Gertsenstein, M., Nagy, A., Lobe, C., Koop, B.F., and Hardy, D.M. Zonadhesin is essential for species specificity of sperm adhesion to the egg zona pellucida. *J. Biol. Chem.* **2010**, 285, 24863-24870.

297. Hardy, D.M. and Garbers, D.L. A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand factor. *J. Biol. Chem.* **1995**, 270, 26025-26028.

298. Redgrove, K.A., Anderson, A.L., Dun, M.D., McLaughlin, E.A., O'Bryan, M.K., Aitken, R.J., and Nixon, B. Involvement of multimeric protein complexes in mediating the capacitation-dependent binding of human spermatozoa to homologous zonae pellucidae. *Dev Biol* **2011**, 356, 460-474.

299. Aebersold, R. and Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198-207.

300. Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, 20, 3551-3567.

301. Pappin, D.J., Hojrup, P., and Bleasby, A.J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **1993**, 3, 327-332.

302. Wöhlbrand, L., Trautwein, K., and Rabus, R. Proteomic tools for environmental microbiology - a roadmap from sample preparation to protein identification and quantification. *Proteomics* **2013**, 13, 2700-2730.

303. Wu, L. and Han, D.K. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert. Rev Proteomics* **2006**, 3, 611-619.

304. de Mateo, S., Estanyol, J.M., and Oliva, R. Methods for the analysis of the sperm proteome. *Methods Mol Biol* **2013**, 927, 411-422.

305. Oliva, R., de Mateo, S., and Estanyol, J.M. Sperm cell proteomics. *Proteomics* **2009**, 9, 1004-1017.

306. Baker, M.A. The 'omics revolution and our understanding of sperm cell biology. *Asian J Androl* **2011**, 13, 6-10.

307. du Plessis, S.S., Kashou, A.H., Benjamin, D.J., Yadav, S.P., and Agarwal, A. Proteomics: a subcellular look at spermatozoa. *Reprod Biol Endocrinol.* **2011**, 9, 36.

308. Brewis, I.A. and Gadella, B.M. Sperm surface proteomics: from protein lists to biological function. *Mol Hum Reprod* **2010**, 16, 68-79.

309. Holt, W.V., Elliott, R.M., Fazeli, A., Satake, N., and Watson, P.F. Validation of an experimental strategy for studying surface-exposed proteins involved in porcine sperm-oviduct contact interactions. *Reprod Fertil Dev* **2005**, 17, 683-692.

310. Stein, K.K., Go, J.C., Lane, W.S., Primakoff, P., and Myles, D.G. Proteomic analysis of sperm regions that mediate sperm-egg interactions. *Proteomics* **2006**, 6, 3533-3543.

311. Flesch, F.M., Wijnand, E., van de Lest, C.H., Colenbrander, B., van Golde, L.M., and Gadella, B.M. Capacitation dependent activation of tyrosine phosphorylation generates two sperm head plasma membrane proteins with high primary binding affinity for the zona pellucida. *Mol Reprod Dev* **2001**, 60, 107-115.

312. Aitken, R.J., Nixon, B., Lin, M., Koppers, A.J., Lee, Y.H., and Baker, M.A. Proteomic changes in mammalian spermatozoa during epididymal maturation. *Asian J. Androl* **2007**, 9, 554-564.

313. Platt, M.D., Salicioni, A.M., Hunt, D.F., and Visconti, P.E. Use of differential isotopic labeling and mass spectrometry to analyze capacitation-associated changes in the phosphorylation status of mouse sperm proteins. *J. Proteome. Res.* **2009**, 8, 1431-1440.

314. Lefievre, L., Chen, Y., Conner, S.J., Scott, J.L., Publicover, S.J., Ford, W.C., and Barratt, C.L. Human spermatozoa contain multiple targets for protein S-nitrosylation: an alternative mechanism of the modulation of sperm function by nitric oxide? *Proteomics* **2007**, 7, 3066-3084.

315. Lalancette, C., Faure, R.L., and Leclerc, P. Identification of the proteins present in the bull sperm cytosolic fraction enriched in tyrosine kinase activity: a proteomic approach. *Proteomics* **2006**, 6, 4523-4540.

316. Runnebaum, I.B., Schill, W.B., and Töpfer-Petersen, E. ConA-binding proteins of the sperm surface are conserved through evolution and in sperm maturation. *Andrologia* **1995**, 27, 81-90.

317. Ensslin, M., Calvete, J.J., Thole, H.H., Sierralta, W.D., Adermann, K., Sanz, L., and Töpfer-Petersen, E. Identification by affinity chromatography of boar sperm membrane-associated proteins bound to immobilized porcine zona pellucida. Mapping of the phosphorylethanolamine-binding region of spermadhesin AWN. *Biol Chem Hoppe Seyler* **1995**, 376, 733-738.

318. Lefièvre, L., Bedu-Addo, K., Conner, S.J., Machado-Oliveira, G.S., Chen, Y., Kirkman-Brown, J.C., Afnan, M.A., Publicover, S.J., Ford, W.C., and Barratt, C.L. Counting sperm does not add up any more: time for a new equation? *Reproduction* **2007**, 133, 675-684.

319. Stewart-Savage, J. Effect of bovine serum albumin concentration and source on sperm capacitation in the golden hamster. *Biol. Reprod.* **1993**, 49, 74-81.

320. Bedu-Addo, K., Lefievre, L., Moseley, F.L., Barratt, C.L., and Publicover, S.J. Bicarbonate and bovine serum albumin reversibly 'switch' capacitation-induced events in human spermatozoa. *Mol Hum Reprod* **2005**, 11, 683-691.

321. Visconti, P.E., Bailey, J.L., Moore, G.D., Pan, D., Olds-Clarke, P., and Kopf, G.S. Capacitation of mouse spermatozoa. I. Correlation between the capacitation state and protein tyrosine phosphorylation. *Development* **1995**, 121, 1129-1137.

322. Parrish, J.J., Susko-Parrish, J.L., and First, N.L. Capacitation of bovine sperm by heparin: inhibitory effect of glucose and role of intracellular pH. *Biol. Reprod.* **1989**, 41, 683-699.

323. Kasekarn, W., Kanazawa, T., Hori, K., Tsuchiyama, T., Lian, X., Garenaux, E., Kongmanas, K., Tanphaichitr, N., Yasue, H., Sato, C., and Kitajima, K. Pig sperm membrane microdomains contain a highly glycosylated 15-25-kDa wheat germ agglutinin-binding protein. *Biochem. Biophys. Res. Commun.* **2012**, 426, 356-362.

324. Liu, D.Y. and Baker, H.W. A simple method for assessment of the human acrosome reaction of spermatozoa bound to the zona pellucida: lack of relationship with ionophore A23187-induced acrosome reaction. *Hum. Reprod.* **1996**, 11, 551-557.

325. Olson, G.E., Winfrey, V.P., Garbers, D.L., and Noland, T.D. Isolation and characterization of a macromolecular complex associated with the outer acrosomal membrane of bovine spermatozoa. *Biol. Reprod.* **1985**, 33, 761-779.

326. Baenziger, J.U. and Fiete, D. Structure of the complex oligosaccharides of fetuin. *J. Biol. Chem.* **1979**, 254, 789-795.

327. Kim, E., Park, K.E., Kim, J.S., Baek, D.C., Lee, J.W., Lee, S.R., Kim, M.S., Kim, S.H., Kim, C.S., Koo, D.B., Kang, H.S., Ryoo, Z.Y., and Chang, K.T. Importance of the porcine ADAM3 disintegrin domain in sperm-egg interaction. *J Reprod Dev* **2009**, 55, 156-162.

328. Key, M. A tutorial in displaying mass spectrometry-based proteomic data using heat maps. *BMC. Bioinformatics.* **2012**, 13 Suppl 16, S10.

329. Leahy, T. and Gadella, B.M. Sperm surface changes and physiological consequences induced by sperm handling and storage. *Reproduction* **2011**, 142, 759-778.

330. Schöneck, C., Braun, J., and Einspanier, R. Sperm viability is influenced in vitro by the bovine seminal protein aSFP: effects on motility, mitochondrial activity and lipid peroxidation. *Theriogenology* **1996**, 45, 633-642.

331. Roncoletta, M., Morani, E.S., Esper, C.R., Barnabe, V.H., and Franceschini, P.H. Fertility-associated proteins in Nelore bull sperm membranes. *Anim Reprod. Sci.* **2006**, 91, 77-87.

332. Dostàlovà, Z., Calvete, J.J., Sanz, L., Hettel, C., Riedel, D., Schöneck, C., Einspanier, R., and Töpfer-Petersen, E. Immunolocalization and quantitation of acidic seminal fluid protein (aSFP) in ejaculated, swim-up, and capacitated bull spermatozoa. *Biol. Chem. Hoppe Seyler* **1994**, 375, 457-461.

333. Tedeschi, G., Oungre, E., Mortarino, M., Negri, A., Maffeo, G., and Ronchi, S. Purification and primary structure of a new bovine spermadhesin. *Eur. J. Biochem.* **2000**, 267, 6175-6179.

334. Moura, A.A., Koc, H., Chapman, D.A., and Killian, G.J. Identification of proteins in the accessory sex gland fluid associated with fertility indexes of dairy bulls: a proteomic approach. *J. Androl* **2006**, 27, 201-211.

335. Varki, A. Sialic acids as ligands in recognition phenomena. *FASEB J.* **1997**, 11, 248-255.

336. Deppe, M., Morales, P., and Sánchez, R. Effect of protease inhibitors on the acrosome reaction and sperm-*zona pellucida* binding in bovine sperm. *Reprod. Domest. Anim* **2008**, 43, 713-719.

337. Ding, Z., Qu, F., Guo, W., Ying, X., Wu, M., and Zhang, Y. Identification of sperm forward motility-related proteins in human seminal plasma. *Mol. Reprod. Dev.* **2007**, 74, 1124-1131.

338. Lu, C.H., Lee, R.K., Hwu, Y.M., Chu, S.L., Chen, Y.J., Chang, W.C., Lin, S.P., and Li, S.H. SERPINE2, a serine protease inhibitor extensively expressed in adult male mouse reproductive tissues, may serve as a murine sperm decapacitation factor. *Biol. Reprod.* **2011**, 84, 514-525.

339. Finkelstein, M., Megnagi, B., Ickowicz, D., and Breitbart, H. Regulation of sperm motility by PIP2(4,5) and actin polymerization. *Dev. Biol.* **2013**, 381, 62-72.

340. Breitbart, H., Cohen, G., and Rubinstein, S. Role of actin cytoskeleton in mammalian sperm capacitation and the acrosome reaction. *Reproduction* **2005**, 129, 263-268.

341. Finkelstein, M., Etkovitz, N., and Breitbart, H. Role and regulation of sperm gelsolin prior to fertilization. *J. Biol. Chem.* **2010**, 285, 39702-39709.

342. Lee, W.M. and Galbraith, R.M. The extracellular actin-scavenger system and actin toxicity. *N. Engl. J. Med.* **1992**, 326, 1335-1341.

343. Park, S.M., Hwang, I.K., Kim, S.Y., Lee, S.J., Park, K.S., and Lee, S.T. Characterization of plasma gelsolin as a substrate for matrix metalloproteinases. *Proteomics.* **2006**, 6, 1192-1199.

344. McCauley, T.C., Zhang, H.M., Bellin, M.E., and Ax, R.L. Identification of a heparin-binding protein in bovine seminal fluid as tissue inhibitor of metalloproteinases-2. *Mol. Reprod. Dev.* **2001**, 58, 336-341.

345. Mondéjar, I., Grullón, L.A., García-Vázquez, F.A., Romar, R., and Coy, P. Fertilization outcome could be regulated by binding of oviductal plasminogen to oocytes and by releasing of plasminogen activators during interplay between gametes. *Fertil Steril.* **2012**, 97, 453-461.

346. Coy, P., Jiménez-Movilla, M., García-Vázquez, F.A., Mondéjar, I., Grullón, L., and Romar, R. Oocytes use the plasminogen-plasmin system to remove supernumerary spermatozoa. *Hum Reprod* **2012**, 27, 1985-1993.

347. Fusi, F.M. and Bronson, R.A. Sperm surface fibronectin. Expression following capacitation. *J Androl* **1992**, 13, 28-35.

348. Armstrong, J.S., Rajasekaran, M., Hellstrom, W.J., and Sikka, S.C. Antioxidant potential of human serum albumin: role in the recovery of high quality human spermatozoa for assisted reproductive technology. *J. Androl* **1998**, 19, 412-419.

349. Akerlof, E., Fredricsson, B., Gustafson, O., Lunell, N.O., Nylund, L., Rosenborg, L., Slotte, H., and Pousette, A. Serum factors stimulate the motility of human spermatozoa. *Int. J. Androl* **1989**, 12, 124-130.

350. Singleton, C.L. and Killian, G.J. A study of phospholipase in albumin and its role in inducing the acrosome reaction of guinea pig spermatozoa in vitro. *J. Androl* **1983**, 4, 150-156.

351. Gonçalves, R.F., Staros, A.L., and Killian, G.J. Oviductal fluid proteins associated with the bovine *zona pellucida* and the effect on in vitro sperm-egg binding, fertilization and embryo development. *Reprod Domest. Anim* **2008**, 43, 720-729.

352. Sitaram, N., Subbalakshmi, C., Krishnakumari, V., and Nagaraj, R. Identification of the region that plays an important role in determining antibacterial activity of bovine seminalplasmin. *FEBS Lett.* **1997**, 400, 289-292.

353. Clark, E.N., Corron, M.E., and Florman, H.M. Caltrin, the calcium transport regulatory peptide of spermatozoa, modulates acrosomal exocytosis in response to the egg's zona pellucida. *J. Biol. Chem.* **1993**, 268, 5309-5316.

354. Winnica, D.E., Novella, M.L., Dematteis, A., and Coronel, C.E. Trypsin/acrosin inhibitor activity of rat and guinea pig caltrin proteins. Structural and functional studies. *Biol. Reprod.* **2000**, 63, 42-48.

355. Dematteis, A., Miranda, S.D., Novella, M.L., Maldonado, C., Ponce, R.H., Maldera, J.A., Cuasnicu, P.S., and Coronel, C.E. Rat caltrin protein modulates the acrosomal exocytosis during sperm capacitation. *Biol. Reprod.* **2008**, 79, 493-500.

356. Bharshankar, R.N. and Bharshankar, J.R. Relationship of seminal plasma transferrin with seminal parameters in male infertility. *Indian J. Physiol Pharmacol.* **2000**, 44, 456-460.

357. Hinton, B.T., Palladino, M.A., Rudolph, D., Lan, Z.J., and Labus, J.C. The role of the epididymis in the protection of spermatozoa. *Curr. Top. Dev. Biol.* **1996**, 33, 61-102.

358. Ibrahim, N.M., Gilbert, G.R., Loseth, K.J., and Crabo, B.G. Correlation between clusterin-positive spermatozoa determined by flow cytometry in bull semen and fertility. *J. Androl* **2000**, 21, 887-894.

359. Bilodeau, J.F., Chatterjee, S., Sirard, M.A., and Gagnon, C. Levels of antioxidant defenses are decreased in bovine spermatozoa after a cycle of freezing and thawing. *Mol. Reprod. Dev.* **2000**, 55, 282-288.

360. Bailey, R. and Griswold, M.D. Clusterin in the male reproductive system: localization and possible function. *Mol. Cell Endocrinol.* **1999**, 151, 17-23.

361. Wang, G., Wu, Y., Zhou, T., Guo, Y., Zheng, B., Wang, J., Bi, Y., Liu, F., Zhou, Z., Guo, X., and Sha, J. Mapping of the N-Linked Glycoproteome of Human Spermatozoa. *J. Proteome Res.* **2013**, 12, 5750-5759.

362. Kierszenbaum, A.L. Sperm axoneme: a tale of tubulin posttranslation diversity. *Mol. Reprod. Dev.* **2002**, 62, 1-3.

363. Soggiu, A., Piras, C., Hussein, H.A., De Canio, M., Gaviraghi, A., Galli, A., Urbani, A., Bonizzi, L., and Roncada, P. Unravelling the bull fertility proteome. *Mol Biosyst.* **2013**, 9, 1188-1195.

364. Ashrafzadeh, A., Nathan, S., and Karsani, S.A. Comparative Analysis of Mafriwal (Bos taurus x Bos indicus) and Kedah Kelantan (Bos indicus) Sperm Proteome Identifies Sperm Proteins Potentially Responsible for Higher Fertility in a Tropical Climate. *Int. J Mol Sci.* **2013**, 14, 15860-15877.

365. Park, Y.J., Kwon, W.S., Oh, S.A., and Pang, M.G. Fertility-related proteomic profiling bull spermatozoa separated by percoll. *J. Proteome Res.* **2012**, 11, 4162-4168.

366. D'Amours, O., Frenette, G., Fortier, M., Leclerc, P., and Sullivan, R. Proteomic comparison of detergent-extracted sperm proteins from bulls with different fertility indexes. *Reproduction* **2010**, 139, 545-556.

367. Peddinti, D., Nanduri, B., Kaya, A., Feugang, J.M., Burgess, S.C., and Memili, E. Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC. Syst Biol* **2008**, 2, 19.

368. Byrne, K., Leahy, T., McCulloch, R., Colgrave, M.L., and Holland, M.K. Comprehensive mapping of the bull sperm surface proteome. *Proteomics* **2012**, 12, 3559-3579.

369. Druart, X., Rickard, J.P., Mactier, S., Kohnke, P.L., Kershaw-Young, C.M., Bathgate, R., Gibb, Z., Crossett, B., Tsikis, G., Labas, V., Harichaux, G., Grupen, C.G., and de Graaf, S.P. Proteomic characterization and cross species comparison of mammalian seminal plasma. *J Proteomics* **2013**, 91C, 13-22.

370. Kelly, V.C., Kuy, S., Palmer, D.J., Xu, Z., Davis, S.R., and Cooper, G.J. Characterization of bovine seminal plasma by proteomics. *Proteomics* **2006**, 6, 5826-5833.

371. Belleannee, C., Labas, V., Teixeira-Gomes, A.P., Gatti, J.L., Dacheux, J.L., and Dacheux, F. Identification of luminal and secreted proteins in bull epididymis. *J Proteomics* **2011**, 74, 59-78.

372. Moura, A.A., Souza, C.E., Stanley, B.A., Chapman, D.A., and Killian, G.J. Proteomics of cauda epididymal fluid from mature Holstein bulls. *J Proteomics* **2010**, 73, 2006-2020.

373. Moura, A.A., Chapman, D.A., Koc, H., and Killian, G.J. A comprehensive proteomic analysis of the accessory sex gland fluid from mature Holstein bulls. *Anim Reprod Sci.* **2007**, 98, 169-188.

374. Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **2010**, 38, W64-W70.

375. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* **2003**, 31, 258-261.

376. Oliva, R. Protamines and male infertility. *Hum. Reprod. Update* **2006**, 12, 417-435.

377. Ganguly, I., Gaur, G.K., Kumar, S., Mandal, D.K., Kumar, M., Singh, U., Kumar, S., and Sharma, A. Differential expression of protamine 1 and 2 genes in mature spermatozoa of normal and motility impaired semen producing crossbred Frieswal (HFxSahiwal) bulls. *Res. Vet. Sci.* **2013**, 94, 256-262.

378. Rashid, S., Grzmil, P., Drenckhahn, J.D., Meinhardt, A., Adham, I., Engel, W., and Neesen, J. Disruption of the murine dynein light chain gene Tcte3-3 results in asthenozoospermia. *Reproduction* **2010**, 139, 99-111.

379. Joti, P., Ghosh-Roy, A., and Ray, K. Dynein light chain 1 functions in somatic cyst cells regulate spermatogonial divisions in Drosophila. *Sci. Rep.* **2011**, 1, 173.

380. Baker, M.A., Reeves, G., Hetherington, L., Muller, J., Baur, I., and Aitken, R.J. Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics Clin. Appl.* **2007**, 1, 524-532.

381. Johnston, D.S., Wooters, J., Kopf, G.S., Qiu, Y., and Roberts, K.P. Analysis of the human sperm proteome. *Ann. N. Y. Acad. Sci.* **2005**, 1061, 190-202.

382. Wang, G., Guo, Y., Zhou, T., Shi, X., Yu, J., Yang, Y., Wu, Y., Wang, J., Liu, M., Chen, X., Tu, W., Zeng, Y., Jiang, M., Li, S., Zhang, P., Zhou, Q., Zheng, B., Yu, C., Zhou, Z., Guo, X., and Sha, J. In-depth proteomic analysis of the human sperm reveals complex protein compositions. *J Proteomics* **2013**, 79, 114-122.

383. Marín-Briggiler, C.I., Lapyckyj, L., González Echeverría, M.F., Rawe, V.Y., Alvarez, S.C., and Vazquez-Levin, M.H. Neural cadherin is expressed in human gametes and participates in sperm-oocyte interaction events. *Int. J. Androl* **2010**, 33, e228-e239.

384. Benoff, S., Cooper, G.W., Hurley, I., Napolitano, B., Rosenfeld, D.L., Scholl, G.M., and Hershlag, A. Human sperm fertilizing potential in vitro is correlated with differential expression of a head-specific mannose-ligand receptor. *Fertil. Steril.* **1993**, 59, 854-862.

385. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, 227, 680-685.

386. McMeekin, T.L., Wilenski, M., and Groves, M.L. Refractive indices of proteins in relation to amino acid composition and specific volume. *Biochem. Biophys. Res. Commun.* **1962**, 7, 151-156.

387. Harpaz, Y., Gerstein, M., and Chothia, C. Volume changes on protein folding. *Structure.* **1994**, 2, 641-649.

APPENDIX I:

Sira Defaus, Preeti Gupta, David Andreu, and Ricardo Gutiérrez-Gallego. "Mammalian protein glycosylation – structure versus function". *Analyst* (2014), (DOI: 10.1039/C3AN02245E).

Analyst

CRITICAL REVIEW

View Article Online
View Journal

Mammalian protein glycosylation – structure *versus* function

Cite this: DOI: 10.1039/c3an02245e

S. Defaus,[ab] P. Gupta,[b] D. Andreu[a] and R. Gutiérrez-Gallego*[abc]

Carbohydrates fulfil many common as well as extremely important functions in nature. They show a variety of molecular displays – *e.g.*, free mono-, oligo-, and polysaccharides, glycolipids, proteoglycans, glycoproteins, *etc.* – with particular roles and localizations in living organisms. Structure-specific peculiarities are so many and diverse that it becomes virtually impossible to cover them all from an analytical perspective. Hence this manuscript, focused on mammalian glycosylation, rather than a complete list of analytical descriptors or recognized functions for carbohydrate structures, comprehensively reviews three central issues in current glycoscience, namely (i) structural analysis of glycoprotein glycans, covering both classical and novel approaches for teasing out the structural puzzle as well as potential pitfalls of these processes; (ii) an overview of functions attributed to carbohydrates, covering from monosaccharide to complex, well-defined epitopes and full glycans, including post-glycosylational modifications, and (iii) recent technical advances allowing structural identification of glycoprotein glycans with simultaneous assignation of biological functions.

Received 4th December 2013
Accepted 6th March 2014

DOI: 10.1039/c3an02245e

www.rsc.org/analyst

1. Introduction

The world of carbohydrates is extremely complex, rendering it both fascinating and challenging to those facing the task of unraveling their structural features. The term carbohydrate spans many different disciplines from large-scale industrial applications to fine-tuned biomedical uses, and the science of carbohydrates has experienced ups and downs over the last few decades in terms of attention paid, importance attributed, and level of understanding reached. Currently, the field of carbohydrate (bio)chemistry is enjoying renewed interest at both basic and applied (biomedical, pharmaceutical) levels, as clearly evidenced by the >500 reviews on the subject over the past 18 months. Most efforts are devoted to the study of carbohydrate-mediated biomolecular interactions and glycoprotein engineering but the structural analysis of carbohydrates, in all

[a]*Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, 08003 Barcelona, Spain. E-mail: ricardogutierrez@upf.edu*

[b]*Bio-analysis Group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona Biomedical Research Park, 08003 Barcelona, Spain*

[c]*Anapharm Biotech, 08038 Barcelona, Spain*

Sira Defaus received her B.S. (2008) and M.Sc. (2009) degrees from the Department of Organic Chemistry at the University of Barcelona. Later she joined the Bioanalysis and Analytical Services Research Group, and the Protein Chemistry and Proteomics laboratory at the Barcelona Biomedical Research Park (PRBB) to obtain her PhD degree. During her study she investigated the structure–activity relationships of glycans and glycopeptides and developed analytical tools for protein detection in biological sources. Her research interests are peptide chemistry and glycobiology.

Preeti Gupta obtained her M.Sc. degree in Biotechnology and M.Tech degree in Biochemical Engineering from Banaras Hindu University, India. In 2012 she joined the Bio-analysis Group at the Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona through the Agency for Management of University and Research Grants (AGAUR) offered by the Government of Catalonia, Spain to pursue her PhD. She is interested in characterizing protein glycosylation, glycan structure and glycoprotein identification and quantification in plasma samples using HPLC and MALDI-TOF and LC-MS.

its aspects, remains the basis of nearly all the developments of recent times. The goal of this review is to highlight relevant aspects of structural analysis of carbohydrates with focus on mammalian protein glycosylation and insights into its relevance. The final section deals with recent advances paving the way towards structural analysis within actual biological settings, ideally, without any external interference.

2. Structural analysis of glycoprotein glycans

Glycoproteins are fundamental in most important biological processes including fertilization, immune response, inflammation, viral replication, parasite infection, cell growth, cell–cell adhesion, or glycoproteins clearance. Whereas protein synthesis follows a well-defined, genetically encoded linear process, glycosylation is a non-template-driven, secondary gene event initiated during protein synthesis and involving a large collection of redundant and overlapping enzymes (glycosidases and glycosyltransferases) partially compartmentalized throughout the endoplasmic reticulum (ER) and the Golgi system.[1] Various competing reactions in the processing pathways, plus the need for enzyme, acceptor and substrate concurrence, as well as other physiological factors contribute to glycan microheterogeneity, *i.e.*, glycoprotein isoforms resulting from different glycans at a given site. This heterogeneity may be relatively simple, such as for RNAse B,[2] or rather complex as in the case of CD59 where at least 123 different desialylated glycan variants have been identified at a single site.[3] Thus, carbohydrate diversity and consequent complexity arises from several factors. Firstly, from the structural variety at the monosaccharide level, where multidirectional combinations of different monosaccharide building blocks, linkages,[4] anomericity, and branching generate a vast number of complex glycan structures (polysaccharides) that can be further modified by sulfation, acetylation, methylation, phosphorylation, *etc.*, and linked covalently to aglycones such as peptides (in different ways) or lipids forming the corresponding glycoconjugates (see Table 1). Secondly, from the influence of the peptide sequence in determining potential glycosylation sites, the effect of the 3D protein display in subsequent glycan processing events, and the spatial distribution or multivalent presentation leading to the avidity principle.[5] Thirdly, from microheterogeneity and macroheterogeneity phenomena inherent to carbohydrate chemistry resulting from the fact that in an individual glycoprotein a specific glycosylation site is not always associated with the same glycan structure and that not all *N*-glycan sequons are necessarily glycosylated.

Table 1 Different types of glycosylation. The letters in the sequence correspond to the 1-letter annotation of amino acids[a]

Linkage	Type	Sequence
Man–α-Trp	*C*-Mannosylation	W–X–X–W
GlcNAc–β-Asn	*N*-Glycosylation	N–X–(S/T) (X ≠ P) N–X–C, N–G, N–X– (rare)
GalNAc–α-Ser/Thr	*O*-Glycosylation	Various ppGalNAcT act concertedly
GlcNAc–β-Ser/Thr		Any S or T
GlcNAc–α-Thr		T (near P residues)
Glc–α-Tyr		GYG (glycogenin)
Glc–β-Ser		C–X–S–X–P–C
Glc–β-Asn		N–X–(S/T)
Gal–Thr		G–X–T (X = A, R, P, hP, S)
Gal–β-Hyl		X–Hyl–G
Fuc–α-Ser/Thr		C–X–X–G–G–(S/T)–C X–X–X–X–(S/T)
Man–α-Ser/Thr		I–X–P–T–(P/X)–T–X–P– X–X–X–X–P–T–X–(T/X)–X–X
Man–α-1-P–Ser		S rich domains
Xyl–β-Ser		–G–S–G– (near acidic residues)

[a] X may be any amino acid.

David Andreu studied chemistry at the University of Barcelona, where he obtained his PhD (1981, with E. Giralt). After further training in peptide chemistry during a postdoc with Bruce Merrifield at Rockefeller University (1982–1985), in 1985 he returned to Spain as Associate Professor of organic chemistry at the University of Barcelona. In 2001 he was appointed Professor of Chemistry at the Pompeu Fabra University where he heads the Protein Chemistry and Proteomics laboratory at the Barcelona Biomedical Research Park. In addition to glycan–protein interactions, his other scientific endeavours include cell-penetrating and antimicrobial peptides, and peptide-based synthetic vaccines.

Ricardo Gutiérrez-Gallego studied chemistry at Utrecht University (NL) and obtained a PhD in 2001 (with J. Vliegenthart and J. Kamerling). Subsequently, he joined the IMIM-Parc de Salut Mar where he developed the analytical methodology program for (glyco) protein hormone detection in the context of antidoping policies. Simultaneously, in 2001, he was appointed assistant professor of chemistry at Pompeu Fabra University where he expanded his research on the structure–function relationship of protein glycosylation with special emphasis on glycan-mediated biomolecular interactions. In 2013 he amplified his scope of activities joining Anapharm Biotech to provide analytical support in drug-development programs.

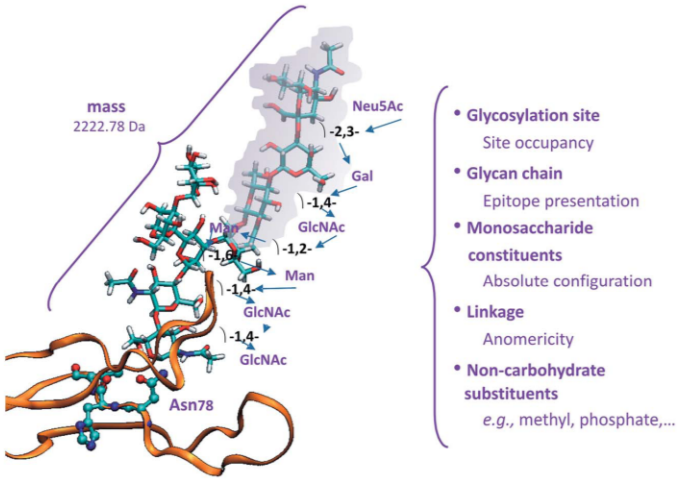


Fig. 1 As an example in the structural elucidation of glycoproteins an *N*-glycan in the human chorionic gonadotropin (hCG) glycoprotein α-chain is shown. Elements to be specified are listed on the right and some of them displayed. The shaded part represents the epitope potentially recognized in a carbohydrate-driven interaction.

Eventually, such diversity gives rise to a set of glycoforms, in both soluble and membrane-anchored forms that are as essential to life as a genetic code, and constitute an evolutionary conserved feature of all living cells.[6] The identification of the number, structure, and function of glycans in a particular biological context, initiated decades before the "omics" boom, was recently termed glycomics, and substantial progress has been made in understanding how glycans are directly involved in almost every biological process or human disease.[7] Still, the glycome is far more complex than the genome, transcriptome, or proteome, due to a much more dynamic character that varies considerably not only with the cell or tissue type, but also with the developmental stage,[8] metabolic state, or changes such as disease,[9] aging,[10,11] environmental factors,[12] or evolution.[13,14] For instance, epigenetic regulation may induce novel glycan structures that make the organism fitter in a specific environment without altering genetic information.[15] It is therefore of utmost importance to know what carbohydrate structures decorate which glycoproteins under particular conditions.

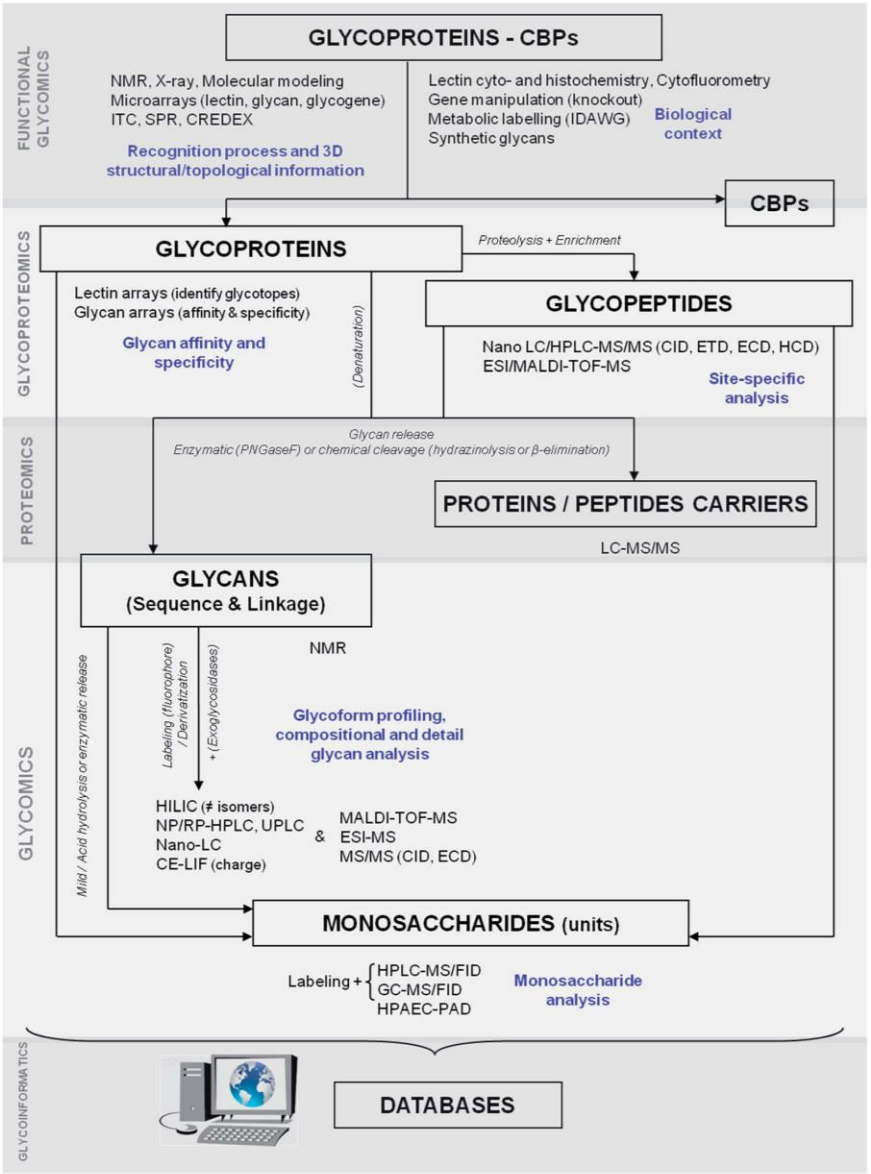
Even for dedicated specialists analysis of protein glycosylation remains an extremely challenging task due to many different physical parameters that must be established before a structural characterization can be considered complete (Fig. 1). As a consequence, there is no single analytical method capable of providing all the necessary information for fast and reliable identification and quantification of a particular structure, let alone to also establish its particular functionality. Rather, a multidimensional approach involving several orthogonal, physical, chemical, and biochemical techniques as depicted in Scheme 1 is required.

In the following pages an overview is provided of the techniques employed in structural analysis of protein glycosylation, their shortcomings and particular virtues, and the latest trends in this field.

2.1. Analyzing glycosylation

Over the last 2–3 decades the continuous refinement of analytical tools has greatly facilitated glycan analysis; numerous reviews[16–21] and papers cover the main technologies routinely used today for *N*- and *O*-linked glycan analysis, including capillary electrophoresis (CE),[22–24] liquid chromatography (LC),[25,26] mass spectrometry (MS)[27–30] and microarray-based[31–35] approaches to glycomics and glycoproteomics.[28,36,37] It is important to stress that in all these techniques a compromise exists between analytical sensitivity and the degree of structural detail provided. None of these tools, or any other for that matter, can single-handedly reveal all the features (see Fig. 1) necessary for full characterization. Hence, an unambiguously structural analysis must be conducted at different levels, namely intact glycoprotein, glycopeptides and released glycans, and in each case the most appropriate technique for deciphering that part of the puzzle must be chosen. This, in turn, entails another compromise between the degree of information obtained *vs.* the amount of (purified) material required.

2.1.1. Analysis of intact glycoproteins. In the first evaluation of protein glycosylation it is recommended to assess the microheterogeneity at the glycoprotein level as it provides an excellent starting point. Quite often this is done by means of conventional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and comparing the resulting bands to a protein molecular weight standard (Fig. 5). Such evaluation, when conducted with non-specific staining techniques using coomassie, silver, or Pro-Q emerald dyes, should provide an unbiased view of the glycoform distribution. Alternatively, the detection could be performed through specific biomolecular recognition (using lectins, antibodies, *etc.*) at much better sensitivity than the non-specific staining. However, one should bear in mind that such biorecognition may be biased towards



Scheme 1 Different levels of glycan analysis include compositional and detailed glycan structure, glycan affinity and specificity, glycoform profiling, site-specific analysis and 3D structural and topological studies. Moreover, determination of carbohydrate-binding protein (CBP) structures and characterization of glycan–CBP recognition and complex formation are required, particularly in biological contexts. Advanced glyco-informatic resources are essential for analytical data collection, annotation, and analysis of the large-scale data generated.

particular glycoforms because of steric effects or other factors hampering interactions with other glycoforms. For its part, SDS-PAGE does not provide an accurate molecular weight determination as separation is governed by the hydrodynamic volume of the migrating species. One alternative technique is gel-based isoelectric focusing (IEF) (Fig. 2). This technique gained much momentum in the early days of proteomics as part of the two-dimensional gel-electrophoretic sample preparation and provides a rough charge distribution of the glycoprotein. Given the limited number of charged modifications of amino acid

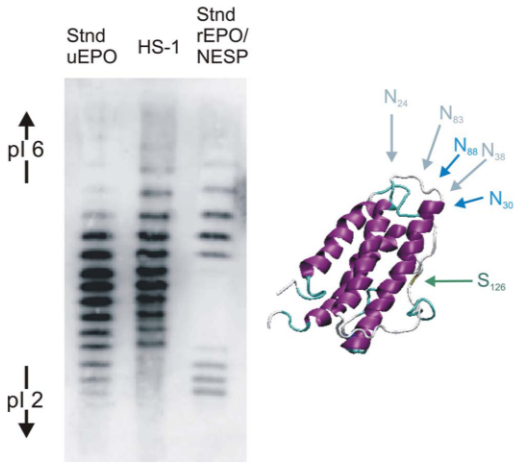


Fig. 2 Isoelectric focusing profiles of an endogenous erythropoietin (EPO) standard (left lane), EPO from a human urinary sample (center lane) and a recombinant erythropoietin mixture (right lane) composed of Eprex (3 *N*-glycans and migrating just below pI 6) and Darbepoetin alpha (5 *N*-glycans and migrating just above pI 2). On the right the crystal structure of erythropoietin (1BUY) and its glycosylation sites are shown.

residues, the IEF profile usually provides a reliable sketch of the degree of sialylation, sulfation, phosphorylation and/or glucuronidation of the glycoprotein.

Similar information can also be obtained from the mass spectrum of the intact glycoprotein. Matrix Assisted Laser Desorption Ionization Time Of Flight (MALDI-TOF) mass spectrometry is particularly suited for this purpose, being capable of handling complex mixtures and fairly tolerant of impurities, aside from detergents, which produce significant ion suppression.[38] Fig. 3 compares the MALDI-TOF spectra of a glycoprotein (rAT-III) and a non-glycosylated protein (rGH) and shows how the peak width provides information on the heterogeneity of the protein, and the peak number on the prevalent glycoforms. Depending on the purity of the glycoprotein, on its structural complexity, and on instrument resolution, information on microheterogeneity can be quite exhaustive.

A subsequent step in the structural interrogation, still at the glycoprotein level, concerns the evaluation of the monosaccharide residues present. One may distinguish different levels of analysis, all requiring the chemical hydrolysis of the glycoprotein. Typically, a first level of analysis addresses sialic acid (Sia) residues. The relevance of Sia was acknowledged nearly six decades ago[39] and, at the time, a specific colorimetric protocol named Bial's reaction and based on orcinol was employed for its detection. With time, the number of residues in the Sia family has increased and currently more than 50 structurally different sialic acid residues,[40] with a variety of associated functions,[41] are known. Even though the analysis of Sia has been pursued through many different approaches, it was selective conjugation of the released α-keto-acids with *ortho*-diamines to form quinoxaline derivatives that allowed both sensitive and specific analysis by liquid chromatography of this family of compounds. In particular, 1,2-diamino-4,5-methylenedioxybenzene (DMB) has found widespread use due to its fluorescent properties (ex: 373 nm, em: 448 nm),[42] which grants the protocol a yet unmatched sensitivity, and also because the mild acid conditions required to release Sia residues do not cause migration of the labile acetyl at C–O7.[43] For example, the detection of *N*-glycolyl neuraminic acid in erythropoietin – found at picomolar concentrations in human specimens – is an unambiguous evidence for a doping violation that could be established by this protocol[44] but not with MS analysis of the same sample (personal communication). Hence, obtaining a complete picture in terms of Sia speciation will usually require the DMB protocol, though care must be exercised as, in addition to the release-related degradation, other α-keto acids or 1,2-diketones in biological samples, *e.g.*, α-ketoglutaric, pyruvic or *p*-hydroxyphenyl-pyruvic acids, can interfere. In such cases, hyphenation of liquid chromatography with fluorescence detection (LC-FLD) to mass spectrometry is possibly the best solution.

The next level of analysis, namely determining all monosaccharides present in a glycoprotein, usually requires a strong acid (*e.g.*, 1 M HCl in methanol, 65 °C, overnight, in the presence of an internal standard) to hydrolyze all glycosidic linkages – except that between the first GlcNAc residue and Asn in *N*-linked glycosylation – and convert glycosidic acetals into the corresponding methylglycosides. The procedure, however, will also irretrievably cleave most post-glycosylational modifications. Following neutralization and evaporation, free hydroxyls are further derivatized with trimethylsilyl (or analogous) functionalities for both qualitative (four characteristic peaks for each monosaccharide) and quantitative evaluation using gas chromatography flame ionization detection (GC-FID) or gas chromatography mass spectrometry (GC-MS) and comparison with a standard monosaccharide mix.[45,46] Alternatively, following acid release, monosaccharides may be separated chromatographically by high performance anion exchange chromatography (HPAEC) (CarboPack PA-100) and detected by pulsed amperometric detection (PAD).[47] The latter protocol offers the advantage of a single peak per monosaccharide and of direct analysis without derivatization, but the basic LC conditions may induce C2-epimerization in GlcNAc to yield ManNAc,[48] or peeling reactions where some monosaccharides are degraded from the reducing end. Altogether, monosaccharide analysis offers the possibility of identifying which type of glycosylation is present; Man and GalNAc being representative of *N*- and *O*-glycosylation, respectively. Furthermore, the stoichiometry of the different sugars allows an educated guess on the type of *N*-linked glycans present by considering the ratio between the distinct monosaccharides with respect to Man. A similar approach can be employed to estimate substitution profiles (*i.e.* α/β1–2,3,4,6) in glycans. In this case, carbohydrates and other functional moieties are permethylated using the Hakomori protocol,[49] subsequently the monosaccharides are released by acid hydrolysis (leaving the methyl-ether bonds intact) and the resulting hydrolyzed monomers are reduced and acetylated to give volatile, partially methylated alditol acetates

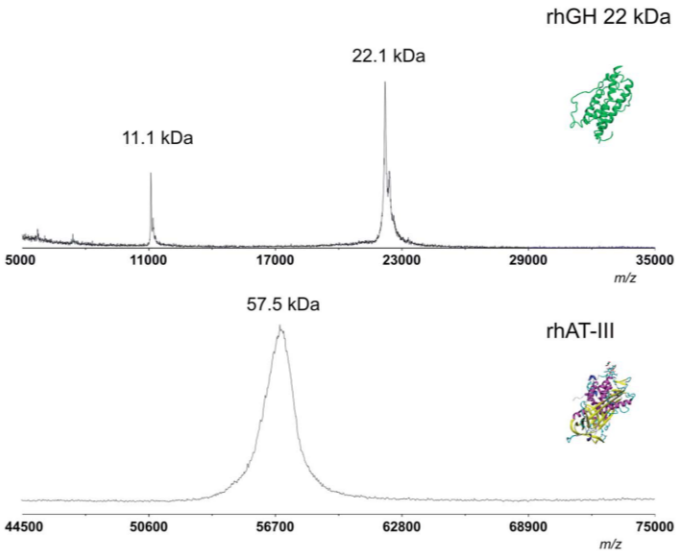


Fig. 3 Top: MALDI-TOF spectrum of recombinant human growth hormone (not glycosylated); bottom: recombinant human antithrombin-III (tri/tetra glycosylated). Microheterogeneity due to glycosylation can be clearly appreciated from the peak width.

(PMAA), again analyzed by GC-MS. This procedure provides unambiguous information on the linkage pattern as well as the ring size of the corresponding sugar, but it is important to emphasize that it is unable to distinguish between a 4-*O*-linked aldopyranose and a 5-*O*-linked aldofuranose.

A more recent development to assess glycosylation at the glycoprotein level consists in the interrogation of a particular glycoprotein or complex biological sample through a lectin array. Even though lectins have long been recognized as tools in the study of glycosylation, their systematic application in array format to detect the glycotopes in a given sample is relatively new.[50–53] Even if the information obtained cannot be compared with a thorough structural analysis (*vide infra*), it has the advantage of analyzing a crude biological sample, *e.g.* the cellular glycome,[54] without too much manipulation, and has demonstrated its value in assessing glycosylation changes in cancer cells on the basis of a direct or an antibody-assisted evanescent-field fluorescence detection scanner.[55,56] In addition, combination with antibodies allows the changes in glycosylation to be pinpointed to specific proteins,[57] adding one more level of specificity to the analysis. Hence, dynamic glycome analysis can be undertaken by means of differentially labeled CBPs.[58] Even so, there are several drawbacks to lectin arrays. For one, while current plant lectin-based arrays[32] cover most monosaccharides in the mammalian glycome, mammalian lectins would obviously provide a more representative glycoprofile. Also, one should not ignore that most lectins are promiscuous to a certain degree and that this behavior, different for each lectin and with different affinities for different sugars, will complicate glycome readout. Ultimately, it appears that lectin–carbohydrate interactions are not always straightforward and that glycoclusters, of either homo- or heterogenic nature, will strongly influence the interaction, and in this case, the analytical data.[5,59] The latter phenomenon appears to be, at least in part, responsible for the fine-tuning of biological communication processes and will as such be very difficult to interpret in terms of precise structural entities.

2.1.2. Analysis of glycoprotein glycans. Evaluation of intact glycans almost inevitably requires their release from the peptide backbone. While some high resolution approaches, in particular those based on MS, are capable of addressing microheterogeneity at the intact glycoprotein level, this is restricted to those entities with a very limited number of glycans and glycoforms such as apolipoprotein C3.[60] For more complex entities, separation of carbohydrate from the protein backbone is needed. Since, in this process, both the site-specificity and the protein origin of the glycans are lost, it is crucial to ensure the maximum degree of protein purity before the procedure is initiated. Deglycosylation can be achieved by either chemical or biological means, each with their respective dis/advantages.

The most widely used chemical method is hydrazinolysis, a procedure that releases the two major classes of glycans (Fig. 4) yet requires highly skilled staff and strict conditions for success, and is invariably accompanied by side-reactions and byproducts. In addition, re-acetylation is necessary to avoid *N*-glycans being lost during the process but may also induce *O*-acetylation. Selective and sequential release of oligosaccharides is achieved by mild hydrazinolysis of *O*-linked oligosaccharides at 60 °C, followed by that of *N*-linked oligosaccharides at 95 °C, but there may be a significant overlap between both processes depending on the protein and the degree of glycosylation. In addition, the procedure will destroy the protein backbone so that if both

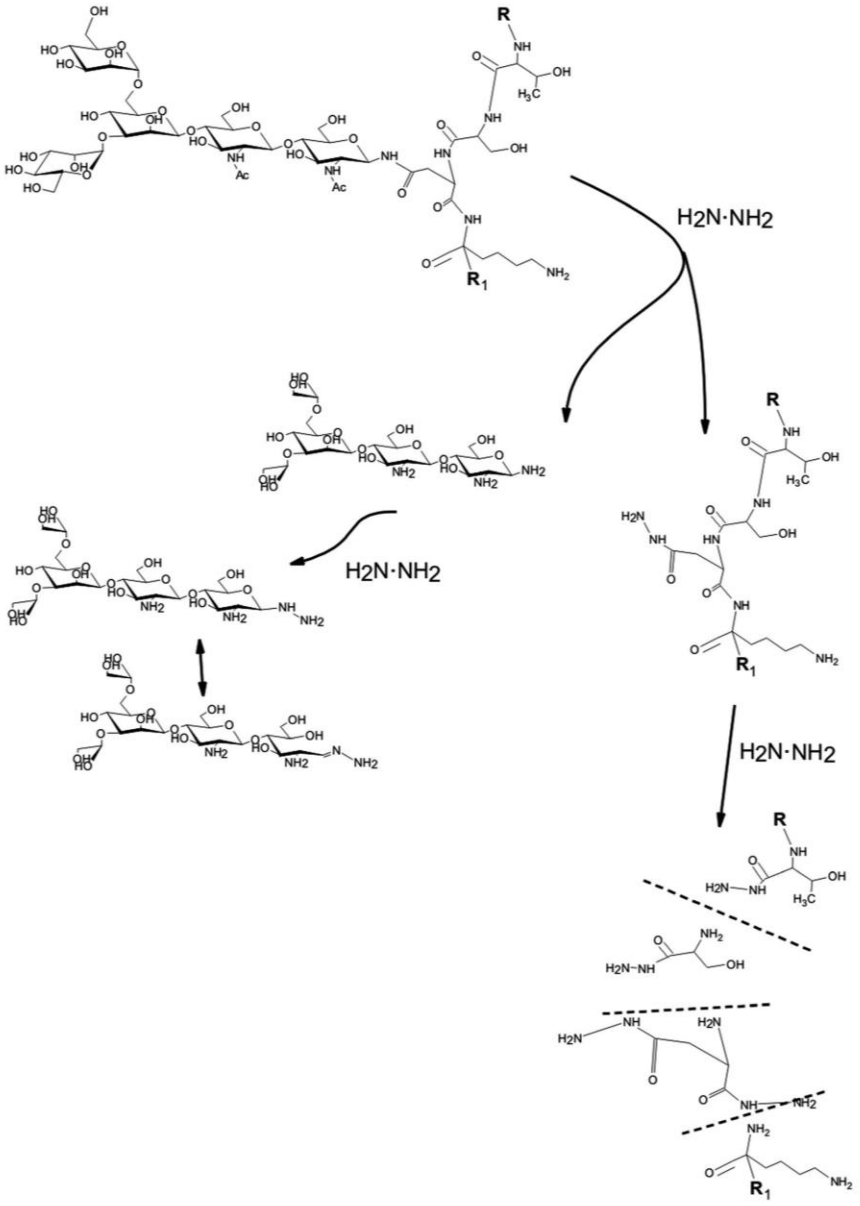


Fig. 4 Release of carbohydrates from the protein backbone following hydrazine treatment. An undesired side effect is protein destruction as indicated in the bottom right.

glycans and protein sequence are to be investigated, hydrazinolysis is not the method of choice. Another chemical procedure, *i.e.*, alkaline β-elimination (0.05 to 0.1 M NaOH or KOH, 60 °C, 12 h), can be applied for *O*-linked carbohydrates attached to Ser or Thr (except those at the carboxy-terminus), but not to Tyr, hydroxy-Pro or hydroxy-Lys. In this case, *N*-linked carbohydrates are unaffected. To prevent base-catalyzed peeling (*vide supra*), sugars must be immediately captured,[61] the alkaline solution carefully prepared,[62] or a reducing agent (*e.g.* 1 M $NaBH_4$) added which forms an alditol that precludes

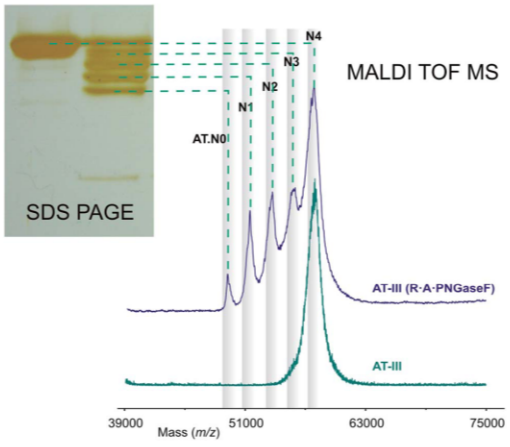


Fig. 5 SDS-PAGE and MALDI-TOF-MS analysis of human antithrombin III (AT-III) after conventional PNGaseF de-*N*-glycosylation. From the number of bands in the SDS-PAGE and peaks in the MALDI-TOF spectrum it is evident that deglycosylation is inefficient and does not reach completion.

reducing-end derivatization. If tagging is intended, it is best performed during release.[63] As with hydrazinolysis, the protein backbone is destroyed in the process.

The only strategy that preserves both protein and carbohydrate is enzymatic deglycosylation, which has been successfully developed for *N*-linked glycans using several endoglycosidases. For mammalian glycoproteins, peptide-N4-(*N*-acetyl-beta-glucosaminyl)asparagine amidase (PNGase F) is the enzyme of choice; it liberates nearly all *N*-linked carbohydrates under standard conditions (*e.g.*, phosphate buffer, 50 mM pH 7.3, 16 h, 37 °C). Nevertheless, the efficiency of this procedure needs to be checked to ensure a correct assessment of the subsequent analysis. One example where conventional conditions do not result in full release is human antithrombin III (AT-III) (Fig. 5), where only partial release of the four *N*-linked structures of the α-variant is achieved if the procedure is not optimized. For plant or invertebrate glycosylations, PNGase A is the preferred choice. In contrast to PNGase F, this enzyme, although of poorer overall efficiency, is capable of releasing α1–3-fucose-bearing core structures. Other endoglycanases (endo F1 to F3 or endo H) can be employed to release the carbohydrate chains, except for Asn-bound GlcNAc, as the enzymes specifically target chitobiose units. In sum, the full repertoire of *N*-glycans can be released by enzymatic means but caution is still advisable. Co- and post-release glycosylamine modification to functionalities other than C1-hydroxyl, such as urea,[64] glycerol[65] or thiol[66] or incomplete conversion in the presence of ammonium may obscure the final analysis. For *O*-glycans, one single deglycosylation enzyme has been identified thus far,[67] and its activity is restricted to T- and Tn-antigenic structures on Ser or Thr. As such, its main application is in unveiling the presence of either epitope, without providing further evidence on the presence of other *O*-glycans.

Once the release from the protein has been completed, the carbohydrates must be purified from the protein and buffer components prior to analysis. Separation into simpler glycan mixtures, a discipline in itself, can significantly facilitate subsequent analyses, but loss of low-abundance glycans may inevitably bias structural identification. On average, every separation step may cause a 10–50% loss of starting material. Generally, separation of the mixture is done by filtration, CE,[23,68] HPAEC,[26] or high-performance liquid chromatography (HPLC). The latter is one of the most versatile, as separation can be based on charge (weak or strong cation or anion exchange), hydrophobicity[69] or hydrophilic interaction (HILIC), and can be performed on either conventional, micro- or nanosized platforms.[25] Separation is typically performed on normal-phase (NP), but reverse-phase (RP) analysis is also possible after permethylation, as demonstrated recently in a comparative study of RP-LC-electrospray (ESI-MS), RP-LC-MALDI-MS, and MALDI-MS[70] using model proteins as well as human blood serum. This study concluded that, for complex samples such as serum, RP-LC-ESI-MS yielded the confident detection of more and lower-abundance glycans, and also permitted the separation of several structural isomers. Another type of derivatization, *i.e.*, selective incorporation of a reporter group at the reducing end of every glycan, is one major step forward in the field of carbohydrate profiling. Research at the Oxford Glycobiology Institute pioneered this approach for comprehensive glycosylation analysis when the starting material is scarce.[71,72] In this approximation, carbohydrates are labeled with fluorescent 2-amino benzamide (2AB), profiled by both weak anion exchange (WAX) and NP (nowadays HILIC) HPLC, and elution times standardized against a partial acid hydrolysate of 2AB-labeled dextran. The resulting glucose unit (GU) values[73,74] allow a preliminary assignment that can then be corroborated by targeted and sequential exoglycosidase digestions, followed by another round of HPLC profiling. Subsequent glycan trimming is of particular interest as not only does it provide confirmation of structural assignments, it simplifies the glycan pool, ultimately contributing to unveiling epitopes that are obscured in the overall microheterogeneity.[75] Fluorescence labeling at the reducing end is not restricted to 2AB as several other tags are described[76] and not only reduces sample requirement to the low femtomole level of individual structures,[77] it also allows accurate relative and absolute quantitation of the glycans present in a given glycoprotein.[78] It has become one of the standard techniques in carbohydrate profiling,[71,79] which can be amplified with internal standards if a different fluorescent tag is used for dextran and the sample, and can be easily extended with back-end MS evaluation when the material is not required for further exoglycosidase treatment.[80] Despite these advantages, it is a laborious approach that requires considerable care, especially during the 2AB labeling (in 30% acetic acid, 65 °C) to avoid desialylation that may confound structural assignments.[78] Automated sample preparation, *i.e.* both the fluorescence labeling and the post-release and post-labeling purification steps, greatly reduces analysis variability, providing robust and reliable glycomics data.[16,19,81]

Arguably one of the more powerful and versatile analytical techniques for all sorts of compounds, including carbohydrates, MS has become the cutting-edge technology for glycomics, linking mass with composition and providing precise characterization of complex structures. A wide range of MS equipments are available for glycan analysis. The introduction of MALDI-TOF instruments allowed rapid and straightforward evaluation of complex mixtures[82,83] and was a giant leap forward in MS evaluation of carbohydrates, hitherto restricted to the cumbersome, low-sensitivity fast atom bombardment mass spectrometry (FAB-MS). Improved analyses are made possible through combination with well-known derivatization strategies (*e.g.* permethylation or peracetylation) that reduce polarity and improve sensitivity by either MALDI-TOF-MS[84] or LC-ESI-MS.[69] Isotope-based differential derivatization protocols, *e.g.*, using CH_3I and CD_3I, allow exact determination of the number of free hydroxyls in a given structure, from which valuable information on carbohydrate composition can be inferred.[85–87] When glycan sequencing is the goal, analyses must include tandem MS experiments where structure-revealing ions are obtained by a combination of ion activation/fragmentation strategies such as collision-induced dissociation (CID), electron transfer dissociation (ETD) or electron capture dissociation (ECD). For instance, a recent study involving a series of oligosaccharide-derived oxonium fragment ions generated by CID enabled simultaneous characterization of IgG glycoforms at both Fc and Fab glycosylation sites by combining multiple reaction monitoring (MRM) MS with energy-resolved structural analysis.[88] In another study, CID-MS was employed to selectively monitor the generation of a *m/z* 284.053 fragment, consistent with GlcNAc phosphorylation in a mouse brain dataset, and unveiled this new post-glycosylational modification.[89] Within the last few years, ECD and ETD have enabled the assignment of *O*-GlcNAc sites at the proteomic scale and greatly facilitated protein-specific studies of single *O*-GlcNAc events. Particularly, ETD has been used to identify *O*-GlcNAc sites and PTMs such as phosphorylation and Arg methylation, on host cell factor C1 (HCF-1), a chromatin-associated protein involved in transcriptional regulation and cell proliferation, and one of the most highly *O*-GlcNAc-rich proteins found in cells.[90] Although MS is clearly indispensable in glycomic analysis, some techniques still present limitations such as susceptibility to salts, difficult assignment of isomeric and isobaric monosaccharides – even though the evolution of ion-mobility strategies are addressing this –,[91] complicated behaviour of acyl groups on glycans, and ionization efficiency dependence. Moreover, interpretation of MS^n fragmentation datasets remains a limiting factor with regard to throughput, user-dependent variability in discrimination and/or interpretation and complete identification of all glycoforms.

When sample complexity is limited to only a few glycan structures, the analytical technique of choice is nuclear magnetic resonance (NMR), the only one providing both qualitative and quantitative information on the glycan without being destructive. While mono-dimensional (^1H, ^{15}N, ^{13}C) experiments readily provide information on structural reporter groups,[92,93] multi-dimensional, both homo- and heteronuclear experiments yield information on the spatial orientation of the glycotope. Moreover, NMR may provide unambiguous information on the presence and position of post-glycosylational modifications such as sulfation, methylation, acetylation or phosphorylation.[94] Ironically for a technique that had proven crucial in the early development of the carbohydrate field, NMR had gradually lost influence due to the often prohibitive amounts of natural material required. Nevertheless, recent developments enabling analysis of picomoles[95] may reinvigorate a technique which in fact has never lost its appeal for the analysis of carbohydrate biological interactions[96,97] or the effect of changes in glycosylation.[98,99]

Altogether glycomics can be addressed through a variety of strategies and technologies that turn out to be orthogonal rather than parallel. While all of them rapidly generate very large amounts of data, differences between platforms can turn data analysis into a complex, time-consuming task requiring bio-informatics tools and databases to facilitate data processing and interpretation. Most of these glycoinformatic tools have particular focuses, *e.g.*, data from HPLC,[74] MS,[100] NMR or microarray[101,102] experiments. Initiatives for cross-linking data from different techniques and integrating multiple datasets are prospering and extremely useful,[103,104] although in the use of database search outputs critical interrogation is advisable. Additionally, the field of glycobiology would greatly benefit from a single glycan structural annotation, easy and of worldwide access, and supported by public agencies such as NCBI or EMBL. Limited public initiatives in this regard (*e.g.* Consortium for Functional Glycomics – http://www.functionalglycomics.org/static/consortium/consortium.shtml) are at risk of being overshadowed by commercial enterprises (*e.g.* Waters & NIBRT – http://www.waters.com/waters/promotionDetail.htm?id=134654015), most likely with ensuing limitations in accessing data, let alone seeking a universal output.

2.1.3. Analysis of site-specific glycosylation. With the increasing awareness of the importance of site specific glycosylation much effort is invested in addressing the glycoforms, enrichment of glycopeptides, and evaluating glycans at their site of attachment.[36,37] The prerequisite of preserving the peptide backbone eliminates the possibility of quantitative glycan profiling through the 2AB protocol, or any other procedure involving tagging of the reducing end. Furthermore, the analytical strategy is limited to MS as the only technique capable of differentiating peptide and carbohydrate sequences. However, a main drawback of glycopeptide MS analysis is that glycosidic bonds are less stable than amide bonds, so that predominant cleavage of the former leads to deglycosylated peptides with no information on the attachment site. The problem has been solved by simply varying the collision energy, so that fragmentation is selectively directed to either carbohydrate or peptide, and information on either part is obtained.[105] Another useful approach, requiring as above no hardware modification, is switching between high and low cone voltage during the LC-MS analysis. Whereas high voltage promotes glycan fragmentation, low voltage produces intact glycopeptides that are identified through

accurate mass measurements and signal intensity.[106] When applied to complex mixtures, deconvolution of the data is of the utmost importance for precise identification and quantification of singular glycopeptides. A significant advancement in the analysis of labile posttranslational modifications, including glycans, has resulted from the implementation of ECD or ETD. In these experiments, electron transfer from a radical anion to the peptide backbone results in preferential cleavage of the N–Cα bond, hence preserving the modification and allowing reliable analysis of both permanent[107] and transient glycosylation.[90] This high-accuracy mass spectrometric characterization combined with a strategy based on "filter aided sample preparation" (FASP) technology and multi-lectin affinity enrichment recently allowed the characterization of more than 5500 new glycosylation sites, confirming 74% of known sites in different mouse tissues and revealing their topological organization.[108] Still another strategy, named "in-gel non-specific proteolysis for elucidating glycoproteins" (INPEG), includes gel-based separation and subsequent digestion with a protease cocktail. With the reduced sample complexity afforded by SDS-PAGE and the help of a software package (Glycopeptide Finder), complex samples such as crude bovine milk or human serum can be evaluated.[109] It seems clear that standardized analysis protocols[79,110] as well as dedicated software applications[100] will be necessary to accurately and reproducibly assess glycosylation at the glycopeptide level, and to extract biologically relevant conclusions, *e.g.*, differentiation between hepatic and liver cell-surface gamma-glutamyl peptidases,[111] site-specific alteration of haptoglobin glycosylation related to hepatocarcinoma and liver cirrhosis,[112] or how a particular congenital glycosylation disorder (CDG-Id) is associated with site-specific glycan deficiencies.[113]

3. Functional analysis of glycans

The chemical and biological diversity of carbohydrates gives rise to a structural complexity that underlies their functional variety. Thus, glycosylation is not only important for protein folding and stability[114,115] but also plays important roles in various biological processes and recognition events (Fig. 6). These roles may be unrelated to the close structural environment where glycosylation occurs or, in contrast, very stringent in terms of glycotope structure and protein localization. Also, the functions exerted are very diverse including: (i) structural, organizational and stabilizing roles, (ii) protective or barrier functions, (iii) provision of specific receptors for microorganisms, toxins or antibodies to attack, shield or lure, (iv) modulation of protein functions in a glycosylation-dependent manner, (v) intra- and intercellular trafficking roles, and (vi) mediation of cell–matrix or cell–cell interactions.[116,117] Therefore, no particular function can or should be attributed to a given oligosaccharide, so that general statements on the subject are practically impossible. The only common general principle emerging from the numerous functions is that glycans generate important functional diversity required for the development, differentiation, and crosstalk in complex organisms as well as for their interactions with other organisms in the environment.

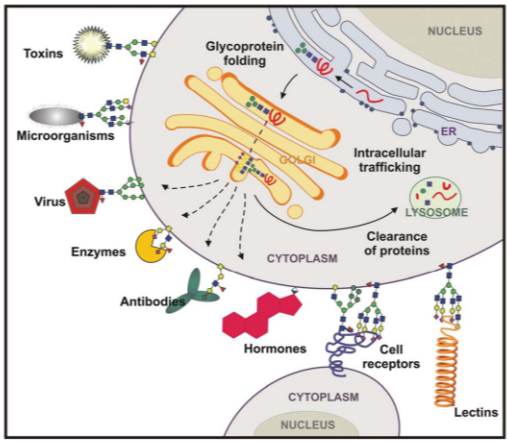


Fig. 6 Glycans participate in multiple mechanisms of cellular regulation. The general functions of glycans span from nascent protein folding and intracellular trafficking to roles in extracellular compartments such as cell–cell communication, providing specific receptors for noxious agents, protecting from microorganisms and antibodies or regulating myriad receptor–ligand interactions.

In the following sections, functions attributed to carbohydrates are reviewed through studies going from the smallest entity to larger glycosidic structures and finally including post-glycosylational modifications (see Fig. 1 and 7).

3.1. Glycosyltransferases

The majority of proteins synthesized in the rough ER undergo glycosylation and the carbohydrate chains attached to these target proteins serve a variety of structural and functional roles in membrane-anchored and secreted proteins. Glycosylation increases proteome diversity, because almost every aspect of glycosylation can be modified, including glycan composition, structure, bond and length.

The cellular glycome assembly *i.e.* the biosynthesis of disaccharides, oligosaccharides and polysaccharides, involves the action of hundreds of different glycosyltransferases (GTs), the enzymes that catalyze the regio- and stereospecific transfer of sugar moieties from activated donor molecules to a variety of acceptor biomolecules including glycans, lipids, peptides, and small molecules forming glycosidic bonds.[118] The complex glycans synthesized by these mammalian GTs are known to play crucial roles in cell–cell, cell–matrix and cell–pathogen interactions, which impact growth and development, infection and immunity, signaling, malignancy, and metabolic disorders. For instance, congenital disorders of glycosylation (CDG) are genetic diseases causing defects in the synthesis or the attachment of the glycan moiety of glycoproteins and glycolipids. Of the more than 40 CDG reported in humans, some 80% affect the nervous system and no effective treatment is known for any of these disorders.

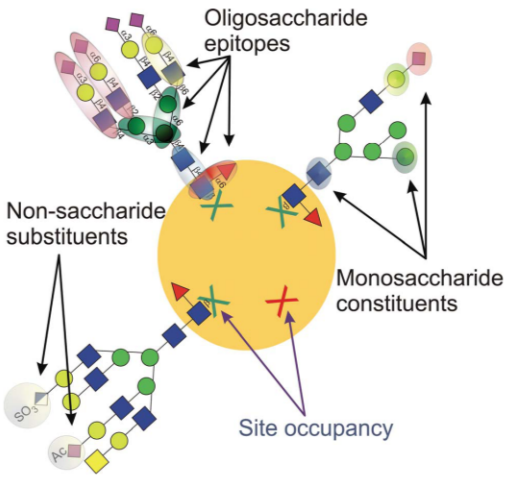


Fig. 7 Different levels at which carbohydrates contribute to glycoconjugate heterogeneity: *i.e.* by occupancy, the monosaccharides that build-up the structure, the specific epitopes composed of the monosaccharides, and ultimately, the non-carbohydrate substituents.

Given their importance in both normal development and pathological conditions, GTs are targets for inhibition and specific small-molecule inhibitors have long been sought to manipulate their activity in cells and to determine the functional roles of glycans. Although recent, structural, kinetic and inhibitor studies have provided important information about the evolution and reaction mechanism of GTs,[119] virtually nothing is known about their donor and acceptor specificity. Therefore, functional characterization remains the greatest challenge in the GT field as there is presently no easy way to assign functions to the many uncharacterized GTs.

3.2. Carbohydrate determinants (glycotopes)

While the complexity and diversity of the totality of glycan structures in an organism is almost impossible to calculate, some 7000 glycan determinants (glycotopes) recognized by CBPs including lectins, receptors, toxins, antibodies, and enzymes have been reckoned for the human glycome.[120] This value is probably underestimated but it provides an idea of the dimension generated by the approximately 700 proteins that make up the mammalian glycan repertoire, and sets the boundaries for glycan–CBP interaction studies[121] where the use of lectins, receptors, antibodies, enzymes, and glycan microarray technologies is crucial for elucidating carbohydrate-specific functions.

3.2.1. Monosaccharide constituents. In this text, "monosaccharide" refers to the simplest form of a sugar, found either as a stand-alone residue or as a terminal or internal part of a polysaccharide. Sialic acids (Sias) are a family of nine-carbon sugars typically attached to the outermost ends of glycoconjugate chains as well as on secreted glycoproteins. The high prevalence of Sias terminating glycan extensions suggests that their predominant function is modulating interactions with the environment. For example, receptor 2B4 of human natural killer (NK) cells has sialic acid residues on both *N*- and *O*-linked glycans. Removal of predicted 2B4 *N*-glycosylation sites decreases binding to its ligand CD48 suggesting that *N*-linked sugars are essential for binding, yet sialylation of 2B4 has a negative impact on ligand binding and 2B4-mediated NK cell cytotoxicity.[122] Similarly, Sias on human corticosteroid-binding globulin (CBG) *N*-glycans were shown to modulate its function, specifically by restricting the binding of CBG to its receptor through steric and/or electrostatic means. Removal of CBG NeuAc residues, or the entire *N*-glycan, increased cAMP production significantly, which was used to evaluate the CBG–receptor interaction.[123]

O-Glycosylation of the Notch extracellular domains in epidermal growth factor (EGF)-like repeats is essential for activity, and tissue-specific alterations in the glycan structures are known to regulate activity. As such, *O*-fucose and *O*-glucose-initiated glycans modulate Notch signaling events critical to cell fate determination and tissue development. More specifically, *O*-fucose-initiated glycans modulate the strength of Notch binding to DSL Notch ligands, while *O*-glucose-initiated glycans facilitate juxta-membrane cleavage, generating the substrate for intramembrane cleavage and Notch activation.[124,125] Moreover, increasing both sialylation and terminal α1–3-linked fucosylation in *N*-glycans could lead to suppression of EGF receptor (EGFR) dimerization and activation in lung cancer cells, which could in turn affect the metastatic ability of cancer cells, EGFR-mediated signaling, and cellular behavior. In particular, the Sia and Fuc residues in the Asn420 *N*-glycan could be critical in inhibiting EGFR dimerization and phosphorylation. In contrast, core fucosylation would promote EGFR dimerization and phosphorylation.[126]

Another prominent example of *O*-glycosylation is the intracellular modification of cytoplasmic and nuclear proteins with *O*-linked-*N*-acetylglucosamine (*O*-GlcNAc) that regulates basic and multiple cellular functions such as transcription and translation, neuronal function, nutrient sensing, cell cycle, and stress. Moreover, it is involved in the etiology of diabetes and neurodegeneration.[127] Indeed, CREB, a central transcription factor in the brain, is highly *O*-GlcNAc monoglycosylated in neurons and influences gene expression by inhibiting both basal and activity-induced CREB-mediated transcription, neuronal function regulation and long-term memory.[128] One of the earliest examples of *O*-GlcNAc modification was found over 25 years ago in nuclear proteins,[129] and since then numerous studies have suggested the existence of dynamic interaction networks, whereby *O*-GlcNAc simultaneously senses and modulates metabolic flow through essential pathways. For instance, histones are modified with *O*-GlcNAc within the nucleosomal core *in vivo*. In particular, histone H2B is GlcNAcylated at S112, and this PTM facilitates K120 monoubiquitination, presumably for transcriptional activation and is responsive to serum glucose levels and/or cellular energy states in certain cell types.[130] Moreover, histone *O*-GlcNAcylation levels change during mitosis and with heat shock showing that *O*-GlcNAc cycles dynamically on histones and can be considered

part of the histone code.[131] This modification is not confined to the nuclear environment as demonstrated by the dynamic induction of *O*-GlcNAc at Ser529 of phosphofructokinase 1 (PFK1) in response to hypoxia. Here the modification inhibits PFK1 activity and redirects the glucose-flux from glycolysis through the pentose phosphate pathway (PPP), thereby conferring a selective growth advantage to cancer cells. This was confirmed by blocking glycosylation of PFK1 in cancer cells resulting in reduced proliferation *in vitro* and impaired tumor formation *in vivo*.[132] Extracellular *O*-GlcNAcylation of secreted and membrane glycoproteins also occurs and mediates cell–cell or cell–matrix interactions at the cell surface.[133] Several recent reviews on *O*-GlcNAcylation have been published providing more details and studies on different aspects of this PTM[134,135] but it is certainly worth mentioning that modulation of these cellular processes by *O*-GlcNAcylation involves a very extensive cross-talk with phosphorylation[136] and that combinations of both, *i.e.* *O*-GlcNAc-6-phosphate have been proposed recently as a novel PTM of mammalian proteins with a variety of possible cellular functions.[89]

3.2.2. Oligosaccharides. As described above, glycans are mostly constituted by multiple monosaccharides and biological activity may be traced to single building blocks. However, with time, evidence has accumulated that in carbohydrate-mediated interactions larger entities (di to hexasaccharides) add yet another level of complexity. Thus, CBPs may recognize complex and relatively large structures that may be either linear or branched homo- or heteropolymeric in nature. One of the very first examples in this context are the ABO(H) major blood group antigens,[137,138] where the absence (O) or presence of an α-Gal (B) or α-GalNAc (A) on Fuc(α1–2)Gal is of paramount importance. Of similar size is the Sd^a-antigen, comprising a Neu5Ac(α2–3)[GalNAc(β1–4)]Gal(β1–R) trisaccharide, expressed in a donor-specific manner in males, and with no particular function hitherto attributed.[139] More recently, this glycotope has been coined as a potential biomarker for colon cancer and its absence related to downregulation of β-1,4-*N*-acetylgalactosaminyltransferase II (β4GalNAcT-II).[140] In close relationship to Sd^a-downregulation stands upregulation of sialyl Lewis^x expression, as α-1,3-fucosyltransferase activity directly competes with β4GalNAcT-II for the acceptor substrate. The Lewis type carbohydrate sequences (Lewis^a, Lewis^b, Lewis^x, Lewis^y, sulfo-Lewis^a, and pseudo-Lewis^y antigens) are expressed on many human glycoproteins and have been assigned a myriad of functions. Just to cite a recent description, terminal Lewis^x and Lewis^y antigens have been reported to be abundantly expressed on *N*-glycans in human seminal plasma glycoproteins and to bind specifically with the lectin domains of DC-SIGN in both male and female to maintain immune homeostasis.[141] The sialyl Lewis^x moiety is also of utmost importance in the interaction between P-selectin glycoprotein ligand 1 (PSGL-1) and P-selectin during the initial phases of inflammatory response.[142] While this interaction is promoted by the *N*-glycan in PSGL-1, in combination with upstream tyrosine sulfation, P-selectin itself is also functionally glycosylated. On a broader scope, the specific *N*-glycosylation status of a particular endothelial adhesion molecule (P/E-selectins, ICAM-1, VCAM-1, or PECAM-1) may regulate protein function during inflammation, affecting both leukocyte capturing and endothelial signalling functions. Adhesion molecule *N*-glycosylation is a dynamic process regulated during inflammation by mechanisms that operate in parallel, but independent of up-regulation of protein expression, and only under those conditions where the appropriate adhesion molecule protein and the corresponding *N*-glycan are expressed will efficient leukocyte adhesion be achieved.[143] For example, the presence of polysialic acid, long chains of α2–8-linked sialic acid residues, on neural cell adhesion molecules (NCAMs) has been demonstrated to decrease cell adhesion, and it is critical for a variety of processes including brain development; synaptic plasticity; axon guidance and path-finding; neurite outgrowth; and general cell migration.[144] Another unique carbohydrate structure characteristically expressed on a series of cell adhesion molecules (L1, myelin associated glycoprotein, TAG1, P0, *etc.*) is the human natural killer (HNK-1) epitope. Initially targeted by an antibody raised to natural killer cells, the epitope was soon recognized to consist of a sulfated trisaccharide, SO_4-3GlcA(β1–3)Gal(β1–4)GlcNAc(β1–R), that is expressed in a spatio-temporally regulated manner during the development and maintenance of the peripheral nervous system. Particularly, the single glycan moiety contained in P0 plays an important role in cell–cell adhesion.[82]

Finally, in the phenomenon of carbohydrate-mediated biological recognition, an extra level of complexity can be added when the carbohydrate binding event is potentiated by a multivalent expression of glycotopes that result in stronger CBP recognition. This phenomenon has been extensively studied under laboratory conditions,[59,145,146] much less in biological settings. A clear example is a recent study on the requirements for neuronal interactions and subsequent axon growth, where clustered presentation of *N*-glycans with *N*-acetyllactosamine (LacNAc) epitopes at branch ends of neural cell adhesion molecule L1 is required for neuronal galectin-4/L1 binding. Impairing the maturation of these epitopes precludes Gal-4/L1 association resulting in a failure of L1 membrane cluster organization, required for proper axon growth.[147]

The above mentioned examples are merely a glimpse of recent descriptive studies and illustrate the increasing relevance of glycotopes. One hopes that the information flow will grow exponentially to meet the vast challenge posed by glycomics and to establish a comprehensive functional appreciation of the human glycome.

3.3. Glycosylation site occupancy

Glycosylation impacts significantly on the physico-chemical properties of the glycoprotein and may thereby exert influence on its viability and activity. These effects are apparently independent of the structural modification but the modification *per se* is necessary. One of the better known examples in this respect is the folding of the nascent polypeptide chain where the monoglucosylated oligomannose structure serves as an anchor point for the chaperone-assisted event.[1] Persistent failure to fold properly, possibly due to the absence of the carbohydrate chain, ultimately results in lysosomal targeting. The common

approach to site occupancy issues involves studying the biological role after (enzymatic or chemical) glycan removal, or upon inhibition of glycosylation, alteration of oligosaccharide processing, or elimination of specific glycosylation sites. The consequences of altering, decreasing or abrogating glycan site occupancy are variable and unpredictable, ranging from nearly undetectable to decreased protein function, production level, stability, or even complete loss of function; in tune with this, the functional interpretation of the absence/defects in glycosylation is not always straightforward.

Several recent reports stress the need of glycosylation for viability; for example, in human chymotrypsin C (CTRC) it is required for efficient folding and secretion. Elimination of *N*-glycosylation by mutation of the single glycan (N52S) reduced CTRC secretion by about 10-fold.[148] Similarly, OATP1B1, an organic anion transporting polypeptide expressed in the human liver and containing three *N*-linked glycans, underwent dramatically decreased expression and was retained within the endoplasmic reticulum when all three sites were mutated to Gln.[149] A further example is BRI2, a type-II transmembrane protein where inhibition of its single *N*-glycosylation reduced cell surface trafficking and led to intracellular accumulation, although the mutation did not affect cleavage by furin or ADAM10.[150]

For glycoproteins whose glycosylation is relevant for functionality, different functional levels can be attained, as shown by numerous reports in recent literature. For instance, blocking the glycosylation of the hepatocyte growth factor receptor (c-Met), a transmembrane tyrosine kinase, attenuates c-Met function through inhibiting its cell membrane targeting.[151] CREB-H, a liver-abundant bZIP transcription factor, requires *N*-glycosylation at three sites in its luminal C-terminal domain for optimal activation.[152] Another example is glycoprotein KCC4, a K^+Cl^- co-transporter isoform involved in maintaining protein stability, regulation of cell volume, anchorage-independent cell growth, tumor formation, and lung colonization by tumors. Deglycosylated KCC4 forms decrease tumor formation and lung colonization in mice. Also, site-directed mutagenesis on the four putative *N*-glycosylation sites established that KCC4 localization to the cell surface depends on the central N331 and N344 sites.[153] This example serves to introduce a next level of functionality, namely when glycosylation of one or more, but not all, sites is required for proper functioning. While this type of study is much more informative, it is also harder to perform as multiple mutant strains must be produced or, alternatively, selective deglycosylation must be achieved. An example can be found in human acetylcholinesterase ($AChE_T$), with three putative *N*-glycosylation sites that are very important for maintaining the catalytically active conformation. Mutants $AChE_T^{N381Q}$, $AChE_T^{N495Q}$ and $AChE_T^{N296Q/N381Q/N495Q}$, particularly the former, showed a dramatic decrease in enzymatic activity compared with $AChE_T^{WT}$. In contrast, glycan removal did not change the sedimentation properties or proportions of AChE, indicating that *N*-linked glycosylation does not affect oligomerization.[154] Similarly, human serum carnosinase CN-1, involved in diabetic nephropathy, contains three potential *N*-glycosylation sites which, if deleted, result in impaired protein secretion; enzyme activity, for its part, is already reduced when two sites are deleted.[155] Finally, myeloperoxidase, a lysosomal protein of neutrophils with five *N*-glycans (N323, N355, N391, N483, and N729), undergoes significant loss of activity upon deglycosylation at N355.[156]

The *N*-glycosylation cases described above need to be completed with a few equally important examples of *O*-glycosylation. In addition to the well-established protective role of *O*-glycosylation in mucins, several more specific functions have been recently discovered. Thus, both regulated and aberrant glycosylation modulate the electrical signaling of the I_{Ks} channel, a macromolecular complex composed of a pore-forming α-(KCNQ1)-subunit and a modulatory β-(KCNE1)-subunit that is crucial for repolarization of the cardiac action potential.[157] Moreover, *O*-glycosylation at Thr-7 in the KCNE1 subunit is essential for proper biosynthesis and trafficking of the complex.[158] Similar examples are two-pore-domain potassium (K(2P)) channels, where disruption of glycosylation reduced current through decreasing the number of channels on the cell surface and hence influencing cellular depolarization.[159]

3.4. Site-specific glycosylation

As described in the previous section, the sole occupation of one or multiple glycosylation sites may affect glycoprotein functionality. It is being increasingly recognized that the selective and specific glycosylation of particular domains, among multiple potential sites, may be key in the regulation of the protein function. Many examples can already be found where glycosylation is impeded through mutations or eliminated after expression yielding a change in functioning. However, examples where glycosylation has been characterized at the structural level and the function studied are still scarce, and it is even more difficult to find structural studies at the site-specific level. In the following section, we review the state of the knowledge with several examples.

The human protein disulfide isomerase family A member 2 (PDIA2), an ER enzyme involved in protein folding and maturation, contains three *N*-glycans, one of which modulates PDIA2 homodimer formation and subsequent chaperoning activity. When devoid of carbohydrate, dimerization was highly efficient and *vice versa*.[160] The precise glycan structure of PDIA2 has not yet been elucidated but it is plausible that a decrease in glycan complexity accelerates protein folding as required. Similarly, upon investigating the role of glycosylation in E-cadherin (four *N*-glycans) and cancer, the N633 glycan is shown to be required for proper folding, trafficking, and expression whereas other glycans are related to stability of adheren junctions. Furthermore, the presence of (α1–6)-fucosylation on Asn-linked GlcNAc promotes cell–cell adhesion in both cancer and downstream signaling pathways.[161] Another study involved the melanocortin 1 receptor, the main determinant of skin pigmentation and phototype, which is *N*-glycosylated at N15 and N29. Mutagenesis and proteolytic studies showed that the N15-bound glycan was not essential while the N29-linked counterpart was crucially involved in ligand binding and normal cell surface expression.[162] In a recent paper on the tumor-associated antigen

Table 2 Examples of glycosylation changes in disease context

Protein/substrate	Alteration	Related disease	Ref.
AMPA receptor GluR2 subunit	Altered *N*-linked glycosylation suggests abnormal trafficking of AMPA receptors from the ER to the synaptic membrane	Schizophrenia	175
Amyloid-beta (Abeta) peptides	The sulfated galactose moiety of sulfatides is essential for Abeta peptide clearance. A deficiency of sulfatides in conjunction with ceramide elevation is associated with AD pathology and is present by the very earliest clinical stage of AD	Alzheimer's disease (AD)	176 and 177
Haptoglobin (Hp)	Unusual hyper-fucosylated site specific glycoforms of Hp	Liver cirrhosis and hepatocellular carcinoma (HCC)	112
Heparan sulfate (HS)	*N*-Sulfation and 2-*O*-sulfation *vs.* lipoprotein binding. Binding and uptake of lipoproteins depends on the degree of sulfation of the chains. Clearance appears to depend on *N*-sulfation based on loss of inhibitory activity of *N*-desulfated	Hepatic clearance of triglyceride-rich lipoproteins	178
Human serum and cerebrospinal fluid (CSF) proteins	Tetraantennary tetrasialylated glycan with a polylactosamine extension shows a 2-fold increase in patient sera. Triantennary trisialylated glycan containing the sLe^x epitope is significantly increased. Levels of bisecting and sialylated glycans in the cerebrospinal fluid show a general downregulation	Schizophrenia	179
Leukemia inhibitory factor (LIF)	Mannose phosphorylation of LIF mediates its internalization thereby reducing extracellular levels and stimulating embryonic stem cell differentiation	Leukemia	180
Lymphoblasts, glycoproteins and gangliosides	Enhanced expression of 9-*O*-acetylated sialoglycoproteins and 9-*O*-acetylated disialoganglioside on lymphoblasts	Childhood acute lymphoblastic leukemia (ALL)	181
Mucosal addressin cell adhesion molecule 1 (MAdCAM-1)	Sulfation of MAdCAM-1 protein with L-selectin ligand carbohydrates (6-sulfo sialyl Lewis X-capped *O*-glycans) regulates UC disease activity	Ulcerative colitis (UC) disease	182
Sialyl-Le(x)-positive mucins	Decrease of *O*-acetylation contributes to colon carcinoma-associated overexpression of sialyl-Le^x	Colorectal carcinoma	183
Sulfated mucins	Cystic fibrosis mucins contain a higher proportion of sialylated and sulfated *O*-glycans compared with non-pathogenic mucins	Cystic fibrosis (CF)	83
Thyroglobulin antibody (TgAb)	HT patients have significantly lower core fucose content on TgAb. Increasing trend of sialylation was found in PTC sera. In all patients, the sialic acid content and TgAb IgG levels showed negative correlation	Thyroid diseases: Hashimoto's thyroiditis (HT), Graves' disease (GD), papillary thyroid carcinoma (PTC), and PTC with histological lymphocytic thyroiditis (PTC-T)	184
α-Dystroglycan (α-DG)	*O*-Mannosyl phosphorylation of α-DG plays critical roles in the pathogenesis of dystroglycanopathy and is a key determinant of α-DG functional expression as a laminin receptor in normal tissues and cells. T192 → M mutation caused deficiencies in α-DG glycosylation and a marked reduction in its ability to bind extracellular-matrix components	Limb-girdle and congenital muscular dystrophy; and muscle–eye–brain disease	94,185 and 186

CD147 (*N*-glycans at N44, N152, and N186), enzymatic deglycosylation and permethylation followed by high-resolution MS analysis revealed the presence of Man_3 to Man_7 structures and barely processed bi-antennary *N*-glycans in which core-fucosylated Man_3 accounted for ~30% of the structures. All glycans were found to stabilize tertiary and quaternary structures and to maintain the active conformation essential for CD147 activity.[163] In addition, N152 was crucial for cell-surface

expression and {β1–6}-GlcNAc {~14%} residues were crucial for translocation to the plasma membrane. These same authors speculate that elevated core-fucosylation, as in E-cadherin, combined with metastasis-associated GnT-V overexpression, could potentiate the role of CD147 in hepatocellular carcinoma cells.

Slightly more dated literature on the subject includes myelin P0 protein, involved in myelin sheet formation, which is glycosylated at a single site and whose microheterogeneity has been fully elucidated and in which the sulfated HNK-1 epitope, crucial for homophilic binding, is only a minor component.[82] Another example, thoroughly studied from both structural and functional perspectives, is human chorionic gonadotropin {hCG},[164] a heterodimeric, cysteine-knot-type glycoprotein that was the first of its kind produced for medical purposes. Both hCG subunits contain two *N*-glycans each {αN52 and αN78; βN13 and βN30} in addition to several *O*-glycans. Oligosaccharides comprise {α2–3}-monosialylated di/monoantennary complex type structures with partial core fucosylation, as well as {α2–3}-monosialylated hybrid type structures. Core fucosylation is found only in the β-subunit and both *N*-glycans are of diantennary complex type, while in the α-subunit N78 lacks hybrid type structures but are instead predominant at N52. This site-specific glycosylation is required for efficient recombination of both α and β subunits to form the active hormone.[165] One should also mention the laborious work on the glycosylation of Tamm–Horsfall glycoprotein,[166] where only one {N14} out of eight potential sites was non-occupied, and the rest exhibited remarkable diversity: the N489 site included di- and tri-charged oligosaccharides exposing, among others, the $4HSO_3$-GalNAc(β1–4)GlcNAc epitope; N251 contained only oligomannose-type chains ranging from $Man_5GlcNAc_2$ to $Man_8GlcNAc_2$, while N208 was quite heterogeneous, with multiply charged complex glycan structures terminated by sulfate groups, Sia residues, and/or the Sd^a-determinant. A final example is human erythropoietin {EPO}, possibly the most extensively studied cytokine, for which full glycan profiling {of the endogenous form} is, after nearly 50 years of research effort, still incomplete. The recombinant version used as a pharmaceutical is less negatively charged despite being fully sialylated,[167] its three N-sites contain complex type tetra-antennary *N*-acetyllactosamine repeats with acetylated Sia residues.[78]

3.5. Modifications of carbohydrates

While glycosylation is unrivalled as PTM in terms of abundance, complexity, and relevance, carbohydrates themselves may be subject to yet another level of structural multifariousness. Post-glycosylational modifications {PGMs} of specific sites {mostly hydroxyl or amino groups} within the glycan chain occur after the oligomer has been assembled and include sulfation, acylation, phosphorylation, methylation or epimerization that may modulate the biological function of the carbohydrate and as such play a critical role in many normal and pathological processes.[168]

Several examples of PGM have already been mentioned above such as the sulfation of the HNK-epitope, or sulfo-Sd^a variant, which may be more common than anticipated. One prime candidate here is EPO, for which sulfation may explain the difference between the fully sialylated recombinant version and the even more charged endogenous variant,[167] but unambiguous evidence is hard to collect in view of endogenous EPO levels and current analytical sensitivity. Sulfation is not only important in glycoprotein glycans; a prominent group of carbohydrates bearing this modification are glycosaminoglycans {GAGs}, unbranched polysaccharides made up of repeating disaccharide units of hexosamine and uronic acid, found on the extracellular matrix of cell surfaces and classified in different types depending on sulfation patterns. These molecules participate directly or indirectly in many different physiological processes ranging from the balance between morphogenetic protein and fibroblast growth factor signaling to maintain cartilage homeostasis[169] to axon growth inhibition after central nervous system injury by specific chondroitin sulfate-E {CS-E} sulfation motif within chondroitin sulfate proteoglycans {CSPGs};[170] activation of the antithrombin–thrombin complex by heparin to promote fibrinogen cleavage;[171] and many others.[172] Another PGM is hydroxyl acylation, particularly, *O*-acetylation of sialic acids in positions 4, 7, 8, and/or 9 that gives rise to many different variants hence communicational possibilities. For instance, 9-*O*-acetylation of Sia regulates the function of CD22 {Siglec-2} *in vivo* as an inhibitor of B cell receptor signaling. Enzymatic acetylation and deacetylation of cell surface α2–6-linked Sia residues controls B cell development, signaling, and immunological tolerance.[173] Phosphorylation, arguably one of the best studied PTMs transiently affecting protein charge, is also a PGM with particular functions in glycans. For instance, mannose-6-phosphate {M6P} is the key targeting signal for acid hydrolase precursor proteins destined for lysosome transport. The M6P tag enables recognition by the M6P receptor, and NMR analysis has revealed the role of phosphodiester-containing lysosomal enzymes in the process.[174] As shown above, the effects of altered oligosaccharides on glycoconjugate functions are highly variable and quite unpredictable, and the resulting aberrant glycome composition is often associated with specific diseases. As an illustration, Table 2 summarizes the diversity of pathological states in which altered glycosylation has been implicated.

Hitherto in this section the relevance of glycosylation as a general phenomenon, its occurrence at a specific position, the site-specific presentation of a particular epitope, and the importance of glycosylation modifications have been presented. From this, it easily follows the considerable interest in the identification of glycan profiles of particular glycoproteins, body fluids or tissues under healthy or disease conditions. For instance, since in most cancers fucosylation and sialylation levels are significantly modified, such aberrant glycan structures can become useful glyco-biomarkers.[187–191] High-throughput discovery and new analytical approaches, including those addressed to PGMs,[192] are becoming essential for unraveling the biological significance of carbohydrate modification and for developing candidate biomarkers for particular conditions. One of the driving forces in the current biomarker research is "single protein-omics", namely, elucidating the

association between disease and site-specific glycoform variants of a protein rather than full-proteome coverage. The following section summarizes the state of the art in this field.

4. Simultaneous structural and functional analysis of glycans

As described in the previous part, glycoproteins are fundamental in many important biological processes and it is quite clear that no single function can be attributed to a particular oligosaccharide. Likewise, there is no single method that routinely provides all the information required for fast and reliable identification and quantification of a particular structure, let alone its particular functionality. One must also bear in mind that, from a biological point of view, identifying the carbohydrate binding entity is as important as deciphering the cognate sugar epitope. The vast majority of structural studies conducted today are performed within the constraints dictated by either physiological or technical boundaries. In an ideal situation, the analysis of biological interactions with glycoprotein participation and directly triggering a physiological response would be performed *in situ*, in real time, and without external intervention (Fig. 8).

However, this goal is as yet unattainable and state-of-the-art approaches still require the use of chemical and/or biological labeling strategies or the analysis under *in vitro* conditions where the biological context is greatly reduced to the cellular level. In the following sections, the latest scientific achievements, focused on the biological functionality, are summarized, often also referring to analytical innovations not included hitherto in this review. Thus, henceforth we will review approaches to (i) investigating glycosylation diversity under physiological conditions, (ii) biological interactions promoted by a particular glycotope, or (iii) introducing well-defined glycosylation by biological (*vs.* chemical) means to explore its functionality.

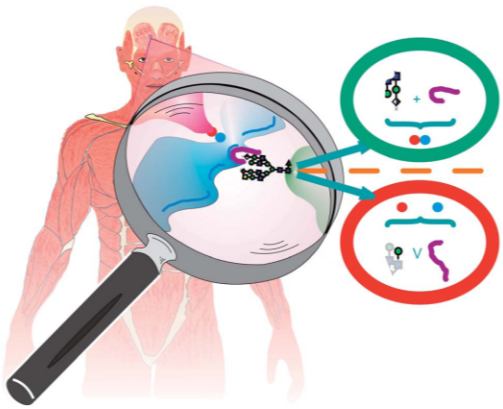


Fig. 8 An ideal analytical setup for monitoring carbohydrate-driven biological functionality. In a productive scenario (green circle), specific binding of a particular carbohydrate epitope and its cognate CBP (purple) triggers further interaction (blue and red circles). In an unproductive scenario (red circle), in contrast, the lack of a sulfated GlcA in the carbohydrate epitope or a non-matching CBP structure precludes binding and subsequent interaction.

4.1. Analysis through biomolecular interactions

There seems to be little doubt that MS-based applications, with their excellent trade-off between analytical capacity, information flow, and sensitivity, will remain a key tool in glycoproteomics studies.[28,36,37] However, these approaches have as an inherent downside the non-natural environment in which analysis takes place, and the fact that comparative analysis (*i.e.* glycosylation *vs.* pathological state or glycan interactions with different binding partners) always requires independent runs. In this regard, array-type experiments, either with a single specimen analyzed over a panel of glycans or lectins, or the reverse format where a panel of glycans or lectins is run over a single sample, constitute interesting alternatives for functional glycomics studies.

One of the classical approaches, *i.e.* use of lectin histochemistry to map the presence and localization of reactive glycan epitopes, as well as detect subtle glycosylation alterations that attend both transformation to malignancy and tumor progression in cells and tissues, has regained prominence of late.[193,194] This technique relies on the readout of biomolecular interactions with surface-exposed carbohydrates, can be employed directly on complex tissue specimens, and is of particular value in extracellular explorations. As such, it has been used exhaustively in the evaluation of gametes. For instance, experiments on fixed sections of the adult murine testis and epididymis revealed that Leydig cells react specifically with SNA or CD22 lectins, both recognizing α2–6-linked Sia; and that the same sugar, but α2–3-linked, allowed differentiation between basal (no Sia) and apical (MAA lectin staining) cells of the epididymis.[195] Also, the application of human galectins as tools for glycophenotyping has been demonstrated by a detailed comparison of their staining properties in the different layers of the *zona pellucida* extracellular matrix using confocal laser scanning microscopy.[196] In the reverse situation, carbohydrate binding molecules are fixed to a solid support and samples are interrogated.[32,54] One noticeable contribution in this context is the development of a dual-color ratiometric readout, similar to that used in gene microarrays. Briefly, two differentially labeled samples mixed in equal amounts compete for lectin binding, allowing the detection of subtle differences in glycosylation expression among many samples by comparing them with a common reference. The versatility, consistency, reproducibility and sensitivity of this approach is nicely illustrated by its application in the comparison of whole mammalian glycomes and the examination of dynamic glycosylation changes upon cell differentiation.[58] Another level of specificity can be added to lectin microarrays by combining them with protein specific antibodies in a sandwich-type approach.[57] This has been successfully applied to establish the prevalence and carriers of particular glycosylation patterns in pancreatic cancer.[56] In these assays, issues such as orientation, native multimeric quaternary

structure, clustering, and metal ion requirements influence the recognition and binding ability of lectins and should not be ignored. Other aspects to consider when working with lectin microarrays are overlapping substrate specificities, which complicate relative quantification, or the detrimental effect of washing steps on weak-to-moderate interactions. Current challenges in the lectin microarray field are expanding the lectin repertoire to include non-plant (and novel plant) lectins, developing recombinant lectins, as well as introducing affinity rather than just specificity parameters in the read-out.

Attention must also be paid to CBPs from a functional perspective. In many situations, the binding partners for particular glycans are not known and, for simplicity, are regarded as constants rather than variables in most studies.[102] In any case, CBP study has become a discipline on its own, mainly fueled by efforts from the Consortium for Functional Glycomics (CFG) (http://www.functionalglycomics.org/static/consortium/consortium.shtml). The generation of mammalian glycan arrays (currently with more than 650 structures) has greatly enabled the systematic study of carbohydrate–protein-binding interactions, despite the intrinsic limitations discussed below. The CFG has developed databases containing not only structures but also experimental data generated with their microarrays. Although microarrays do not reveal glycosylation patterns, they constitute a powerful technology to characterize CBP binding specificities.[31] As an example, both carbohydrate microarray and computational modeling approaches have been used for the rapid screening of glycosaminoglycan (GAG) interactions with proteins and multimeric protein complexes. Novel interactions between a specific sulfated epitope, CS-E, and the neurotrophin family of growth factors have been identified with these methods, providing unique molecular-level insights into the diverse biological functions of GAGs.[197] In another recent study, glycosphingolipids (GSLs) extracted from bovine brain gangliosides and fluorescently labeled were bound to a microarray for subsequent interrogation by biologically relevant CBPs like cholera toxin, antibodies and sera from individuals with Lyme disease.[198] More recently, glyconanoparticles (GNPs) in microarray format have been used to study glycan–lectin interactions. The GNPs were made by conjugating carbohydrate ligands on silica nanoparticles and microarrays were generated by conventional photocoupling chemistry. They were then probed with fluorescein-labeled lectins and with fluorescein-doped silica nanoparticles.[199] The above studies are representative of glycofunctional approaches but several aspects of glycan array design and biomolecular interaction assay should not be overlooked.[34] In addition to the limitations in oligosaccharide synthesis[200] or in glycan isolation from natural sources, aspects such as oligosaccharide density,[201] spacing, and orientation achieved upon immobilization, as well as the nature, flexibility and length of the linker are key parameters in optimizing array strategies. Indeed, since many CBPs achieve their specificity and affinity through multivalent interactions with glycans,[202] glycoarrays should aim at faithful replication of multivalent sugar display, and at capturing the physiological avidity of such interactions in as native-like fashion as possible. For instance, a bead-modified surface providing multivalency (*i.e.* the cluster effect) was used for probing carbohydrate–protein interactions mimicking a cellular environment.[203] These clusters can be of identical or mixed sugar composition to assess both valency and hetero-cluster effects.[5] Also, a new class of end-functionalized polymers mimicking the multivalent architecture of chondroitin sulfate (CS) proteoglycans have been designed, providing insights into how multivalency within and between GAG chains enhances the avidity, specificity and cooperativity of GAG–protein interactions.[204]

Carbohydrates immobilized to solid surfaces have also been employed in slightly different formats and with various other objectives, such as affinity-based systems to detect specific CBP structures or to ascertain other interaction characteristics. For instance, a novel glycan-affinity method combining proteolytic digestion of protein–glycan complexes and mass spectrometry (CREDEX-MS, "Carbohydrate REcognition Domain EXcision Mass Spectrometry") has proven useful in the structural definition of CBPs of two human galectins with lactose as the binding partner.[205]

An extremely valuable biophysical tool for carbohydrate–protein interaction studies is surface plasmon resonance (SPR). In studies aimed at detecting multiple sclerosis autoantibodies in sera, the glycopeptide antigen CSF114(Glc) was immobilized at the 3D-sensor surface and real-time specific autoantibody detection was achieved.[206] SPR is particularly valuable because it allows simultaneous monitoring of several surfaces (a limited form of arraying) and because non-specific binding can be discarded by using a non-glycosylated peptide as a reference surface. A refined version of SPR, Au nano-island based localized surface plasmon resonance, has been used to characterize the specific recognition between concavalin A (Con A) and mannose.[207] The exquisite potential of SPR for analyzing carbohydrate binding proteins was also highlighted in a study of galactose-specific *Erythrina crista-galli* agglutinin binding to several galactoside-epitopes exposed in a well-defined manner[86] at the 3D-sensor surface.[208] Results complemented those of CREDEX-MS, demonstrating that the combination of both techniques can provide good insights into CBPs in various settings. Also, apparent differences in the binding preference of carbohydrate ligands have been observed by quantitative SPR analysis, suggesting that glycan presentation and the conformational space it occupies plays an important role in binding, regardless of affinity.[209]

4.2. Glycoengineering (genetic, chemoenzymatic, chemical)

One alternative approach to assess the importance of glycosylation is selective modification of the carbohydrate decoration and study of its effects. In this context, gene targeting uses homologous recombination to change an endogenous gene. The method can be used to delete a gene, remove exons, add a gene, and introduce point mutations. Indeed, gene targeting has been widely used in glycomics research by removing ("knocking out") or adding ("knocking in") specific mutations of interest to a variety of models. The regulation of genes whose protein products are involved in glycan synthesis and glycan–

protein interaction provides insights into glycan structural diversity and function in complex biological systems.[210] Recently, a strategy for developing cell lines that produce simplified homogenous *O*-glycan structures and thus interrogating the human *O*-glycoproteome has been presented. Named 'Simple Cell', it uses zinc-finger nuclease (ZFN)-based gene targeting of COSMC gene to glycoengineer stable human cell lines displaying only truncated Tn and STn *O*-glycans. More than 100 *O*-glycoproteins and up to 350 glycosylation sites, including a previously unidentified linkage to tyrosine, were elucidated by this approach.[211] The strategy has been used to analyze the function of a single GalNAc-transferase (GalNAc-T) isoform and its role in congenital diseases and disorders.[212]

An attractive approach toward predefined glycoforms is *in vitro* chemoenzymatic glycosylation, *i.e.*, remodeling natural or recombinant glycoproteins by addition of sugar units through sequential glycosyltransferase-catalyzed reactions, or by endoglycosidase-catalyzed transglycosylation and *en bloc* transfer of pre-assembled large oligosaccharides to the protein in a single step under the catalysis of an endo-β-*N*-acetylglucosaminidase (ENGase).[213] Alternatively, site-specific glycosylation can be achieved by chemoselective ligation of proteins to appropriately tagged glycans. For instance, cysteine residues in the protein can be reacted with a thiol-reactive group pre-installed in the sugar moiety to give a disulfide or thioether-linked glycoconjugate. Other strategies involve ligation (oxime, hydrazone) between amino and carbonyl groups, or azide–alkyne cycloadditions under mild, bio-compatible conditions. For instance, using recently developed chemoenzymatic strategies, *N*-glycans containing core-fucose substitution and/or bisecting GlcNAc with otherwise ordinary complex-type antennae terminated in α2–3- or α2–6-linked sialic acid were synthesized and neoglycoproteins produced. With these ultra-defined entities *in vivo* bio-distribution was assessed showing that core substitutions alter glycan ligand properties through conformational changes which act as molecular switches for target affinity and influence glycoprotein-mediated cell binding and serum clearance.[214,215]

4.3. Evaluating glycosylation within the biological context

As mentioned above, evaluating glycosylation within a given biological context and without external manipulation is complicated; hence strategies with minimal impact on the system are actively sought. One approach in this direction is selective targeting, under physiological conditions, of particular glycans that can be subsequently evaluated. A recent study employs an engineered β-1,4-galactosyltransferase to specifically transfer a keto-Gal functionality to *O*-GlcNAc-modified proteins. The ketone moiety was subsequently reacted with various aminooxy-functionalized polyethylene glycol tags of defined mass and the resulting samples were analyzed by gel-based methods. In this way, a direct read-out of *O*-GlcNAc stoichiometry *vs.* state (*e.g.* mono-, di-, tri-, *etc.*) was possible, with insights into the complex interplay between *O*-GlcNAc glycosylation and phosphorylation.[216] Another chemoenzymatic strategy enabled rapid, sensitive and selective detection of the (Fuc(α1–2)Gal) disaccharide motif involved in processes such as learning and memory, inflammation, asthma, and tumorigenesis. By exploiting the restricted substrate tolerance of a blood group A GalNAc-transferase, the disaccharide is targeted with azido-functionalized UDP–GalNAc that is later captured from the complex sample mixture. This labeling strategy provides a variety of different enrichment strategies and imaging read-outs for a variety of Fuc(α1–2)Gal motifs.[217] Broader applications would of course require a supply of such restricted enzymes, currently unavailable at a larger scale. In another example of azide–alkyne chemoselective ("click chemistry") conjugation,[218] by introducing tetraacetylated *N*-azidoacetyl-D-mannosamine in the Sia biosynthetic pathway; mature glycoproteins containing azido-Sia were produced and targeted at the cellular level by capture with a biotinylated alkyne reagent and subsequent MS evaluation. Sias have also been targeted by periodate oxidation, in which vicinal hydroxyls in a *cis* configuration (present only in terminal Man, Gal(NAc) or non *O*-acetylated Sia residues) are converted to aldehydes. Subsequent oxime ligation with aminooxy-biotin labels glycoprotein subpopulations with high efficiency and cell viability, after which samples can be evaluated by MS.[219] The authors use ultra-mild conditions to assess only Sia and target terminal GalNAc through an enzymatic protocol to allow differentiation between Sia-containing and deficient cells.

Another elegant example of glycome comparison, based on stable isotope labeling with amino acids in cell culture (SILAC), labels amine-containing monosaccharides in cells using ^{14}N or ^{15}N glutamine as the sole nitrogen source. Named Isotopic Detection of Aminosugars With Glutamine (IDAWG),[220] the technique shows great promise for analyzing glycome dynamics under different conditions. However, interpretation of the data may not be straightforward as the protocol targets ManNAc, GlcNAc, GalNAc and NeuAc simultaneously, setting an equation with at least four variables added to the intrinsic microheterogeneity of glycosylation. On the backflip is the fact that the proteome may be targeted in a synchronized fashion. A rather different approach was chosen in a recent study where cell surface amino groups were reacted with unsaturated aldehyde yielding dihydropyridines products without affecting cell viability and simultaneously introducing the Hilyte Fluor 750 tag to perform noninvasive whole body fluorescence imaging. Examples included labeling of colon and gastric cancer cell lines in BALB/c nude mice to monitor tumor metastasis.[221]

5. Outlook

In this review we have not aimed at an exhaustive examination of all aspects related to a greater or lesser extent to protein glycosylation. Rather, we sought to provide a taste of some of the disciplines involving glycoscience that will landmark the future. One of the major challenges that glycoscience faced since its very beginning remains: handling the glycoproteome at the endogenous level, addressing its complexity in an automated high-throughput mode, and analyzing glycoproteins in complex samples with simultaneous characterization of both the glycan moieties and the corresponding protein carriers. Novel instrumental developments, such as ion mobility mass

spectrometry, to name only one, or the intelligent hyphenation of orthogonal existing techniques such as combining front-end biomolecular interaction analysis with in-line mass spectrometric evaluation, will be required to meet this challenge which will always constitute the first step in understanding the biological function of a glycoconjugate. In this respect, integration of glycomics with other –omics fields such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics[222] will certainly rank glycomics according to its merits. Current efforts towards systems glycobiology modelling, *i.e.* coupling biochemical knowledge and mathematics into *in silico* models of the cellular glycosylation system, will no doubt be decisive in this respect.[223] Evidently, a broad picture of how glycosylation is regulated through omics-data acquisition and systematic integration will be an enormously valuable asset to gain understanding of glycan functions as well as to develop clinical diagnostics and glyco-biomarker discoveries.[190] Such systems-level studies will help establish novel quantitative and mechanistic links between gene expression, protein expression, enzyme activity, carbohydrate structure and glycoconjugate function.

Abbreviations

2AB	2-Amino benzamide
AChE	Acetylcholinesterase
AD	Alzheimer disease
αDG	Alpha dystroglycan
ALL	Acute lymphoblastic leukemia
Arg	Arginine
Asn	Asparagine
AT	Antithrombin
CBPs	Carbohydrate binding proteins
CD147	Cluster of differentiation 147
CD22	Cluster of differentiation 22
CDG	Congenital disorder of glycosylation
CE	Capillary electrophoresis
CF	Cystic fibrosis
CFG	Consortium for Functional Glycomics
CID	Collision-induced dissociation
CREDEX	Carbohydrate REcognition Domain EXcision
CS-E	Chondroitin sulfate E
CSF	Cerebrospinal fluid
CSPG	Chondroitin sulfate proteoglycans
CTRC	Chymotrypsin C
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DMB	1,2-Diamino-4,5-methylenedioxybenzene
DTT	Dithiothreitol
ECD	Electron capture dissociation
EGFR	Epidermal growth factor receptor
EMBL	European Molecular Biology Laboratory
EPO	Erythropoietin
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FAB	Fast atom bombardment
FASP	Filter aided sample preparation
FID	Flame ionization detector
Fuc	Fucose
GAGs	Glycosaminoglycans
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
GC	Gas chromatography
GCMS	Gas chromatography mass spectrometry
GH	Growth hormone
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	*N*-Acetylglucosamine
Gln	Glutamine
Glu	Glutamic acid
GnT-V	*N*-Acetylglucosaminyl transferase V
GT	Glycosyltransferase
GU	Glucose unit
HCC	Hepatocellular carcinoma
HCD	Higher-energy collisional dissociation
HCl	Hydrochloric acid
HILIC	Hydrophilic interaction liquid chromatography
HNK-1	Human natural killer epitope 1
Hp	Haptoglobin
HPAEC	High performance anion exchange chromatography
HPAEC	High pH anion exchange chromatography
HPLC	High performance liquid chromatography
HS	Heparin sulfate
HSA	Human serum albumin
ICAM	Inter cellular adhesion molecule
IDAWG	Isotopic detection of aminosugars with glutamine
IEF	Isoelectric focusing
INPEG	In-gel non-specific proteolysis for elucidating glycoproteins
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
LC-FLC	Liquid chromatography with post-column fluorescence derivatization
Lewis A	Galβ1–3(Fucα1–4)GlcNAc
Lewis B	(Fucα1–2)Galβ1–3(Fucα1–4)GlcNAc
Lewis X	Galβ1–4(Fucα1–3)GlcNAc
Lewis Y	(Fucα1–2)Galβ1–4(Fucα1–3)GlcNAc
LIF	Laser-induced fluorescence
LIF	Leukemia inhibitory factor
Lys	Lysine
M	Molar – mole per liter
M6P	Mannose-6-phosphate
MAA	*Maackia amurensis* agglutinin
MALDI-TOF	Matrix assisted laser desorption ionization time of flight
Man	Mannose
ManNAc	*N*-Acetylmannosamine
mM	Millimolar
MRM	Multiple reaction monitoring
MS	Mass spectrometry
NCAM	Neural cell adhesion molecule
NCBI	National Center for Biotechnology Information
NeuAc	*N*-Acetylneuraminic acid
NIBRT	National institute for bioprocessing and training

NK	Natural killer
nm	Nanometer
NMR	Nuclear magnetic resonance
NP	Normal-phase
OXM	Oxyntomodulin
PAD	Pulsed amperometric detection
PDIA	Protein disulfide isomerase
PECAM	Platelet endothelial cell adhesion molecule
PFK	Phosphofructokinase
PGMs	Post-glycosylational modifications
PMAA	Partially methylated alditol acetate
PNGase F	Peptide-N4-(*N*-acetyl-beta-glucosaminyl) asparagine amidase
Pro	Proline
PTC	Papillary thyroid carcinoma
PTM	Post-translational modification
Q-TOF	Quadrupole time-of-flight
RP	Reverse-phase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
Sia	Sialic acid
Sialyl-Lewis X	Neu5Acα2–3Galβ1–4(Fucα1–3)GlcNAc
SNA	*Sambucus nigra* agglutinin
SPR	Surface plasmon resonance
S-Tn	Neu5Acα2–6GalNAcα1-*O*-Ser/Thr
T antigen	Galβ1–3-GalNAcα-*O*-Ser/Thr
Thr	Threonine
Tn antigen	GalNAcα1-*O*-Ser/Thr
Tyr	Tyrosine
UC	Ulcerative colitis
UPLC	Ultra performance liquid chromatography
VCAM	Vascular cell adhesion molecule
WAX	Weak anion exchange
Xyl	Xylose
ZFN	Zinc-finger nuclease

References

1 K. W. Moremen, M. Tiemeyer and A. V. Nairn, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 448–462.
2 H. C. Joao and R. A. Dwek, *Eur. J. Biochem.*, 1993, **218**, 239–244.
3 P. M. Rudd, B. P. Morgan, M. R. Wormald, D. J. Harvey, C. W. van den Berg, S. J. Davis, M. A. Ferguson and R. A. Dwek, *J. Biol. Chem.*, 1997, **272**, 7229–7244.
4 R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
5 M. Gómez-García, J. M. Benito, D. Rodríguez-Lucena, J. X. Yu, K. Chmurski, M. C. Ortiz, R. Gutiérrez-Gallego, A. Maestre, J. Defaye and J. M. García Fernández, *J. Am. Chem. Soc.*, 2005, **127**, 7970–7971.
6 A. Varki, *Cold Spring Harbor Perspect. Biol.*, 2011, **3**, 1–14.
7 G. W. Hart and R. J. Copeland, *Cell*, 2010, **143**, 672–676.
8 R. S. Haltiwanger and J. B. Lowe, *Annu. Rev. Biochem.*, 2004, **73**, 491–537.
9 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.
10 L. R. Ruhaak, H. W. Uh, M. Beekman, C. A. Koeleman, C. H. Hokke, R. G. Westendorp, M. Wuhrer, J. J. Houwing-Duistermaat, P. E. Slagboom and A. M. Deelder, *PLoS One*, 2010, **5**, e12566.
11 M. Pucic, A. Muzinic, M. Novokmet, M. Skledar, N. Pivac, G. Lauc and O. Gornik, *Glycobiology*, 2012, **22**, 975–982.
12 J. Wang, C. I. Balog, K. Stavenhagen, C. A. Koeleman, H. U. Scherer, M. H. Selman, A. M. Deelder, T. W. Huizinga, R. E. Toes and M. Wuhrer, *Mol. Cell. Proteomics*, 2011, **10**, M110.
13 S. A. Springer and P. Gagneux, *J. Biol. Chem.*, 2013, **288**, 6904–6911.
14 G. Lauc and V. Zoldos, *Mol. BioSyst.*, 2010, **6**, 2373–2379.
15 V. Zoldos, M. Novokmet, I. Beceheli and G. Lauc, *Glycoconjugate J.*, 2013, **30**, 41–50.
16 T. Tharmalingam, B. Adamczyk, M. A. Doherty, L. Royle and P. M. Rudd, *Glycoconjugate J.*, 2013, **30**, 137–146.
17 J. F. Rakus and L. K. Mahal, *Annu. Rev. Anal. Chem.*, 2011, **4**, 367–392.
18 D. Vanderschaeghe, N. Festjens, J. Delanghe and N. Callewaert, *Biol. Chem.*, 2010, **391**, 149–161.
19 K. Mariño, J. Bones, J. J. Kattla and P. M. Rudd, *Nat. Chem. Biol.*, 2010, **6**, 713–723.
20 M. Pabst and F. Altmann, *Proteomics*, 2011, **11**, 631–643.
21 K. Yamada and K. Kakehi, *J. Pharm. Biomed. Anal.*, 2011, **55**, 702–727.
22 J. Zaia, *Methods Mol. Biol.*, 2013, **984**, 13–25.
23 Y. Mechref, *Electrophoresis*, 2011, **32**, 3467–3481.
24 D. Vanderschaeghe, A. Guttman and N. Callewaert, *Methods Mol. Biol.*, 2013, **919**, 87–96.
25 S. Hua, C. Lebrilla and H. J. An, *Bioanalysis*, 2011, **3**, 2573–2585.
26 J. L. Behan and K. D. Smith, *Biomed. Chromatogr.*, 2011, **25**, 39–46.
27 M. Wuhrer, *Glycoconjugate J.*, 2013, **30**, 11–22.
28 I. M. Lazar, W. Lee and A. C. Lazar, *Electrophoresis*, 2013, **34**, 113–125.
29 J. Zaia, *OMICS: J. Integr. Biol.*, 2010, **14**, 401–418.
30 N. Leymarie and J. Zaia, *Anal. Chem.*, 2012, **84**, 3040–3048.
31 D. F. Smith and R. D. Cummings, *Mol. Cell. Proteomics*, 2013, **12**, 902–912.
32 G. Gupta, A. Surolia and S. G. Sampathkumar, *OMICS: J. Integr. Biol.*, 2010, **14**, 419–436.
33 D. F. Smith, X. Song and R. D. Cummings, *Methods Enzymol.*, 2010, **480**, 417–444.
34 S. Park, J. C. Gildersleeve, O. Blixt and I. Shin, *Chem. Soc. Rev.*, 2012, **42**, 4310–4326.
35 C. D. Rillahan and J. C. Paulson, *Annu. Rev. Biochem.*, 2011, **80**, 797–823.
36 Y. Pasing, A. Sickmann and U. Lewandrowski, *Biol. Chem.*, 2012, **393**, 249–258.
37 J. Nilsson, A. Halim, A. Grahn and G. Larson, *Glycoconjugate J.*, 2013, **30**, 119–136.
38 C. Reichel, *Drug Test. Anal.*, 2010, **2**, 603–619.
39 H. J. Müller-Eberhard and H. G. Kunkel, *J. Exp. Med.*, 1956, **104**, 253–269.

40 H. Nie, Y. Li and X. L. Sun, *J. Proteomics*, 2012, **75**, 3098–3112.
41 A. Varki and P. Gagneux, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 16–36.
42 K. Hayakawa, C. De Felice, T. Watanabe, T. Tanaka, K. Iinuma, K. Nihei, S. Higuchi, T. Ezoe, I. Hibi and K. Kurosawa, *J. Chromatogr.*, 1993, **620**, 25–31.
43 R. Schauer, G. V. Srinivasan, D. Wipfler, B. Kniep and R. Schwartz-Albiez, *Adv. Exp. Med. Biol.*, 2011, **705**, 525–548.
44 J. Mallorquí, E. Llop, C. de Bolòs, R. Gutiérrez-Gallego, J. Segura and J. A. Pascual, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2010, **878**, 2117–2122.
45 J. Kamerling and J. Vliegenthart, *Carbohydrates*, Walter de Gruyter, Berlin, 1989, pp. 176–263.
46 O. M. Grace, A. Dzajic, A. K. Jäger, N. T. Nyberg, A. Onder and N. Rønsted, *Phytochemistry*, 2013, **93**, 79–87.
47 G. Palmieri, M. Balestrieri, J. Peter-Katalinic, G. Pohlentz, M. Rossi, I. Fiume and G. Pocsfalvi, *J. Proteome Res.*, 2013, **12**, 2779–2790.
48 T. Toida, I. R. Vlahov, A. E. Smith, R. E. Hileman and R. J. Linhardt, *J. Carbohydr. Chem.*, 1996, **15**, 351–360.
49 S. Hakomori, *J. Biochem.*, 1964, **55**, 205–208.
50 R. Rosenfeld, H. Bangio, G. J. Gerwig, R. Rosenberg, R. Aloni, Y. Cohen, Y. Amor, I. Plaschkes, J. P. Kamerling and R. B. Maya, *J. Biochem. Biophys. Methods*, 2007, **70**, 415–426.
51 J. Hirabayashi, M. Yamada, A. Kuno and H. Tateno, *Chem. Soc. Rev.*, 2013, **42**, 4443–4458.
52 A. Kuno, A. Matsuda, Y. Ikehara, H. Narimatsu and J. Hirabayashi, *Methods Enzymol.*, 2010, **478**, 165–179.
53 A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada and J. Hirabayashi, *Nat. Methods*, 2005, **2**, 851–856.
54 H. Tateno, A. Kuno, Y. Itakura and J. Hirabayashi, *Methods Enzymol.*, 2010, **478**, 181–195.
55 H. Narimatsu, H. Sawaki, A. Kuno, H. Kaji, H. Ito and Y. Ikehara, *FEBS J.*, 2010, **277**, 95–105.
56 T. Yue, I. J. Goldstein, M. A. Hollingsworth, K. Kaul, R. E. Brand and B. B. Haab, *Mol. Cell. Proteomics*, 2009, **8**, 1697–1707.
57 B. B. Haab, *Expert Rev. Proteomics*, 2010, **7**, 9–11.
58 K. T. Pilobello, D. E. Slawek and L. K. Mahal, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11534–11539.
59 J. L. Jiménez Blanco, M. C. Ortiz and J. M. García Fernández, *Chem. Soc. Rev.*, 2013, **42**, 4518–4531.
60 W. Jian, R. W. Edom, D. Wang, N. Weng and S. W. Zhang, *Anal. Chem.*, 2013, **85**, 2867–2874.
61 N. G. Karlsson and N. H. Packer, *Anal. Biochem.*, 2002, **305**, 173–185.
62 R. P. Kozak, L. Royle, R. A. Gardner, D. L. Fernandes and M. Wuhrer, *Anal. Biochem.*, 2012, **423**, 119–128.
63 A. H. Merry, D. C. Neville, L. Royle, B. Matthews, D. J. Harvey, R. A. Dwek and P. M. Rudd, *Anal. Biochem.*, 2002, **304**, 91–99.
64 L. A. Omtvedt, L. Royle, G. Husby, K. Sletten, C. Radcliffe, R. A. Dwek, P. M. Rudd and D. J. Harvey, *Rapid Commun. Mass Spectrom.*, 2004, **18**, 2357–2359.
65 R. B. Trimble, P. H. Atkinson, A. L. Tarentino, T. H. Plummer, Jr, F. Maley and K. B. Tomer, *J. Biol. Chem.*, 1986, **261**, 12000–12005.
66 D. J. Harvey and P. M. Rudd, *J. Mass Spectrom.*, 2010, **45**, 815–819.
67 L. V. Johnson and G. S. Hageman, *Exp. Eye Res.*, 1987, **44**, 553–565.
68 J. Schwarzer, E. Rapp, R. Hennig, Y. Genzel, I. Jordan, V. Sandig and U. Reichl, *Vaccine*, 2009, **27**, 4325–4336.
69 I. Ritamo, J. Räbinä, S. Natunen and L. Valmu, *Anal. Bioanal. Chem.*, 2013, **405**, 2469–2480.
70 Y. Hu and Y. Mechref, *Electrophoresis*, 2012, **33**, 1768–1777.
71 P. M. Rudd, G. R. Guile, B. Küster, D. J. Harvey, G. Opdenakker and R. A. Dwek, *Nature*, 1997, **388**, 205–207.
72 G. R. Guile, P. M. Rudd, D. R. Wing, S. B. Prime and R. A. Dwek, *Anal. Biochem.*, 1996, **240**, 210–226.
73 M. P. Campbell, L. Royle, C. M. Radcliffe, R. A. Dwek and P. M. Rudd, *Bioinformatics*, 2008, **24**, 1214–1216.
74 N. V. Artemenko, M. P. Campbell and P. M. Rudd, *J. Proteome Res.*, 2010, **9**, 2037–2041.
75 C. Reichel, *Drug Test. Anal.*, 2011, **3**, 883–891.
76 K. R. Anumula, *Anal. Biochem.*, 2006, **350**, 1–23.
77 R. Gutiérrez-Gallego, S. R. Haseley, V. F. van Miegem, J. F. Vliegenthart and J. P. Kamerling, *Glycobiology*, 2004, **14**, 373–386.
78 E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. A. Pascual, *Proteomics*, 2007, **7**, 4278–4291.
79 P. H. Jensen, N. G. Karlsson, D. Kolarich and N. H. Packer, *Nat. Protoc.*, 2012, **7**, 1299–1310.
80 A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, *Glycoconjugate J.*, 2008, **25**, 75–84.
81 M. Pucic, A. Knezevic, J. Vidic, B. Adamczyk, M. Novokmet, O. Polasek, O. Gornik, S. Supraha-Goreta, M. R. Wormald, I. Redzic, H. Campbell, A. Wright, N. D. Hastie, J. F. Wilson, I. Rudan, M. Wuhrer, P. M. Rudd, D. Josic and G. Lauc, *Mol. Cell. Proteomics*, 2011, **10**, M111.
82 R. Gutiérrez-Gallego, J. L. Blanco, C. W. Thijssen-van Zuylen, C. H. Gotfredsen, H. Voshol, J. O. Duus, M. Schachner and J. F. Vliegenthart, *J. Biol. Chem.*, 2001, **276**, 30834–30844.
83 B. Xia, J. A. Royall, G. Damera, G. P. Sachdev and R. D. Cummings, *Glycobiology*, 2005, **15**, 747–775.
84 Z. Lin and D. M. Lubman, *Methods Mol. Biol.*, 2013, **1007**, 289–300.
85 F. Tousi, J. Bones, W. S. Hancock and M. Hincapie, *Anal. Chem.*, 2013, **85**, 8421–8428.
86 C. Jiménez-Castells, B. de la Torre, D. Andreu and R. Gutiérrez-Gallego, *Glycoconjugate J.*, 2008, **25**, 879–887.
87 I. Breloy, *Methods Mol. Biol.*, 2012, **842**, 165–177.
88 A. Toyama, H. Nakagawa, K. Matsuda, T. A. Sato, Y. Nakamura and K. Ueda, *Anal. Chem.*, 2012, **84**, 9655–9662.
89 H. Hahne and B. Kuster, *Mol. Cell. Proteomics*, 2012, **11**, 1063–1069.
90 S. A. Myers, S. Daou, E. B. Affar and A. L. Burlingame, *Proteomics*, 2013, **13**, 982–991.

91 K. Pagel and D. J. Harvey, *Anal. Chem.*, 2013, **85**, 5138–5145.
92 V. H. Pomin, *Anal. Bioanal. Chem.*, 2013, **405**, 3035–3048.
93 J. F. Vliegenthart and J. P. Kamerling, *^{1}H NMR Structural-Reporter-Group Concepts in Carbohydrate Analysis*, Elsevier, 2007, pp. 133–191.
94 T. Yoshida-Moriguchi, L. Yu, S. H. Stalnaker, S. Davis, S. Kunz, M. Madson, M. B. Oldstone, H. Schachter, L. Wells and K. P. Campbell, *Science*, 2010, **327**, 88–92.
95 M. Fellenberg, A. Coksezen and B. Meyer, *Angew. Chem., Int. Ed.*, 2010, **49**, 2630–2633.
96 A. Arda, P. Blasco, S. D. Varon, V. Schubert, S. Andre, M. Bruix, F. J. Canada, H. J. Gabius, C. Unverzagt and J. Jiménez-Barbero, *J. Am. Chem. Soc.*, 2013, **135**, 2667–2675.
97 R. E. Lehotzky, C. L. Partch, S. Mukherjee, H. L. Cash, W. E. Goldman, K. H. Gardner and L. V. Hooper, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 7722–7727.
98 H. Ghasriani, P. J. Belcourt, S. Sauve, D. J. Hodgson, D. Brochu, M. Gilbert and Y. Aubin, *J. Biol. Chem.*, 2013, **288**, 247–254.
99 A. W. Barb and J. H. Prestegard, *Nat. Chem. Biol.*, 2011, **7**, 147–153.
100 C. L. Woodin, M. Maxon and H. Desaire, *Analyst*, 2013, **138**, 2793–2803.
101 A. Porter, T. Yue, L. Heeringa, S. Day, E. Suh and B. B. Haab, *Glycobiology*, 2010, **20**, 369–380.
102 D. Kletter, S. Singh, M. Bern and B. B. Haab, *Mol. Cell. Proteomics*, 2013, **12**, 1026–1035.
103 S. V. Bennun, K. J. Yarema, M. J. Betenbaugh and F. J. Krambeck, *PLoS Comput. Biol.*, 2013, **9**, e1002813.
104 C. W. der Lieth, A. A. Freire, D. Blank, M. P. Campbell, A. Ceroni, D. R. Damerell, A. Dell, R. A. Dwek, B. Ernst, R. Fogh, M. Frank, H. Geyer, R. Geyer, M. J. Harrison, K. Henrick, S. Herget, W. E. Hull, J. Ionides, H. J. Joshi, J. P. Kamerling, B. R. Leeflang, T. Lütteke, M. Lundborg, K. Maass, A. Merry, R. Ranzinger, J. Rosen, L. Royle, P. M. Rudd, S. Schloissnig, R. Stenutz, W. F. Vranken, G. Widmalm and S. M. Haslam, *Glycobiology*, 2011, **21**, 493–502.
105 M. Vila-Perelló, R. Gutiérrez-Gallego and D. Andreu, *ChemBioChem*, 2005, **6**, 1831–1838.
106 M. M. Ivancic, H. S. Gadgil, H. B. Halsall and M. J. Treuheit, *Anal. Biochem.*, 2010, **400**, 25–32.
107 W. R. Alley, Jr, Y. Mechref and M. V. Novotny, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 161–170.
108 D. F. Zielinska, F. Gnad, J. R. Wisniewski and M. Mann, *Cell*, 2010, **141**, 897–907.
109 C. C. Nwosu, J. Huang, D. L. Aldredge, J. S. Strum, S. Hua, R. R. Seipert and C. B. Lebrilla, *Anal. Chem.*, 2013, **85**, 956–963.
110 D. Kolarich, P. H. Jensen, F. Altmann and N. H. Packer, *Nat. Protoc.*, 2012, **7**, 1285–1298.
111 M. B. West, Z. M. Segu, C. L. Feasley, P. Kang, I. Klouckova, C. Li, M. V. Novotny, C. M. West, Y. Mechref and M. H. Hanigan, *J. Biol. Chem.*, 2010, **285**, 29511–29524.
112 P. Pompach, Z. Brnakova, M. Sanda, J. Wu, N. Edwards and R. Goldman, *Mol. Cell. Proteomics*, 2013, **12**, 1281–1293.
113 U. M. Bailey, M. F. Jamaluddin and B. L. Schulz, *J. Proteome Res.*, 2012, **11**, 5376–5383.
114 M. M. Chen, A. I. Bartlett, P. S. Nerenberg, C. T. Friel, C. P. Hackenberger, C. M. Stultz, S. E. Radford and B. Imperiali, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 22528–22533.
115 D. D. Banks, *J. Mol. Biol.*, 2011, **412**, 536–550.
116 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
117 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720.
118 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, **77**, 521–555.
119 C. Breton, S. Fournel-Gigleux and M. M. Palcic, *Curr. Opin. Struct. Biol.*, 2012, **22**, 540–549.
120 R. D. Cummings, *Mol. BioSyst.*, 2009, **5**, 1087–1104.
121 J. C. Paulson, O. Blixt and B. E. Collins, *Nat. Chem. Biol.*, 2006, **2**, 238–248.
122 S. Margraf-Schönfeld, C. Böhm and C. Watzl, *J. Biol. Chem.*, 2011, **286**, 24142–24149.
123 Z. Sumer-Bayraktar, D. Kolarich, M. P. Campbell, S. Ali, N. H. Packer and M. Thaysen-Andersen, *Mol. Cell. Proteomics*, 2011, **10**, M111.
124 P. Stanley and T. Okajima, *Curr. Top. Dev. Biol.*, 2010, **92**, 131–164.
125 N. A. Rana and R. S. Haltiwanger, *Curr. Opin. Struct. Biol.*, 2011, **21**, 583–589.
126 Y. C. Liu, H. Y. Yen, C. Y. Chen, C. H. Chen, P. F. Cheng, Y. H. Juan, C. H. Chen, K. H. Khoo, C. J. Yu, P. C. Yang, T. L. Hsu and C. H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11332–11337.
127 G. W. Hart, M. P. Housley and C. Slawson, *Nature*, 2007, **446**, 1017–1022.
128 J. E. Rexach, P. M. Clark, D. E. Mason, R. L. Neve, E. C. Peters and L. C. Hsieh-Wilson, *Nat. Chem. Biol.*, 2012, **8**, 253–261.
129 G. D. Holt, C. M. Snow, A. Senior, R. S. Haltiwanger, L. Gerace and G. W. Hart, *J. Cell Biol.*, 1987, **104**, 1157–1164.
130 R. Fujiki, W. Hashiba, H. Sekine, A. Yokoyama, T. Chikanishi, S. Ito, Y. Imai, J. Kim, H. H. He, K. Igarashi, J. Kanno, F. Ohtake, H. Kitagawa, R. G. Roeder, M. Brown and S. Kato, *Nature*, 2011, **480**, 557–560.
131 K. Sakabe, Z. Wang and G. W. Hart, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19915–19920.
132 W. Yi, P. M. Clark, D. E. Mason, M. C. Keenan, C. Hill, W. A. Goddard, III, E. C. Peters, E. M. Driggers and L. C. Hsieh-Wilson, *Science*, 2012, **337**, 975–980.
133 Y. Sakaidani, T. Nomura, A. Matsuura, M. Ito, E. Suzuki, K. Murakami, D. Nadano, T. Matsuda, K. Furukawa and T. Okajima, *Nat. Commun.*, 2011, **2**, 583.
134 K. T. Schjoldager and H. Clausen, *Biochim. Biophys. Acta*, 2012, **1820**, 2079–2094.
135 J. A. Hanover, M. W. Krause and D. C. Love, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 312–321.
136 G. W. Hart, C. Slawson, G. Ramirez-Correa and O. Lagerlof, *Annu. Rev. Biochem.*, 2011, **80**, 825–858.
137 W. M. Watkins and W. T. Morgan, *Nature*, 1952, **169**, 825–826.

138 W. T. Morgan and W. M. Watkins, *Glycoconjugate J.*, 2000, **17**, 501–530.
139 J. J. van Rooijen, J. P. Kamerling and J. F. Vliegenthart, *Glycobiology*, 1998, **8**, 1065–1075.
140 N. Malagolini, D. Santini, M. Chiricolo and F. Dall'Olio, *Glycobiology*, 2007, **17**, 688–697.
141 G. F. Clark, P. Grassi, P. C. Pang, M. Panico, D. Lafrenz, E. Z. Drobnis, M. R. Baldwin, H. R. Morris, S. M. Haslam, S. Schedin-Weiss, W. Sun and A. Dell, *Mol. Cell. Proteomics*, 2012, **11**, M111.
142 T. Pouyani and B. Seed, *Cell*, 1995, **83**, 333–343.
143 D. W. Scott and R. P. Patel, *Glycobiology*, 2013, **23**, 622–633.
144 K. J. Colley, *Adv. Exp. Med. Biol.*, 2010, **663**, 111–126.
145 M. Gómez-Garcia, J. M. Benito, R. Gutiérrez-Gallego, A. Maestre, C. O. Mellet, J. M. Fernández and J. L. Blanco, *Org. Biomol. Chem.*, 2010, **8**, 1849–1860.
146 P. V. Murphy, S. André and H. J. Gabius, *Molecules*, 2013, **18**, 4026–4053.
147 S. Velasco, N. Díez-Revuelta, T. Hernández-Iglesias, H. Kaltner, S. André, H. J. Gabius and J. Abad-Rodríguez, *J. Neurochem.*, 2013, **125**, 49–62.
148 M. Bence and M. Sahin-Tóth, *FEBS J.*, 2011, **278**, 4338–4350.
149 J. Yao, W. Hong, J. Huang, K. Zhan, H. Huang and M. Hong, *PLoS One*, 2012, **7**, e52563.
150 M. Tsachaki, D. Serlidaki, A. Fetani, V. Zarkou, I. Rozani, J. Ghiso and S. Efthimiopoulos, *Glycobiology*, 2011, **21**, 1382–1388.
151 R. Chen, J. Li, C. Feng, S. Chen, Y. Liu, C. Duan, H. Li, X. Xia, T. He, M. Wei and R. Dai, *J. Cell. Biochem.*, 2012, **114**, 816–822.
152 C. P. Chan, T. Y. Mak, K. T. Chin, I. O. Ng and D. Y. Jin, *J. Cell Sci.*, 2010, **123**, 1438–1448.
153 T. Y. Weng, W. T. Chiu, H. S. Liu, H. C. Cheng, M. R. Shen, D. B. Mount and C. Y. Chou, *Biochim. Biophys. Acta*, 2013, **1833**, 1133–1146.
154 V. P. Chen, R. C. Choi, W. K. Chan, K. W. Leung, A. J. Guo, G. K. Chan, W. K. Luk and K. W. Tsim, *J. Biol. Chem.*, 2011, **286**, 32948–32961.
155 E. Riedl, H. Koeppel, F. Pfister, V. Peters, S. Sauerhoefer, P. Sternik, P. Brinkkoetter, H. Zentgraf, G. Navis, R. H. Henning, B. J. Van Den, S. J. Bakker, B. Janssen, F. J. van der Woude and B. A. Yard, *Diabetes*, 2010, **59**, 1984–1990.
156 P. Van Antwerpen, M. C. Slomianny, K. Z. Boudjeltia, C. Delporte, V. Faid, D. Calay, A. Rousseau, N. Moguilevsky, M. Raes, L. Vanhamme, P. G. Furtmüller, C. Obinger, M. Vanhaeverbeek, J. Nève and J. C. Michalski, *J. Biol. Chem.*, 2010, **285**, 16351–16359.
157 M. L. Montpetit, P. J. Stocker, T. A. Schwetz, J. M. Harper, S. A. Norring, L. Schaffer, S. J. North, J. Jang-Lee, T. Gilmartin, S. R. Head, S. M. Haslam, A. Dell, J. D. Marth and E. S. Bennett, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16517–16522.
158 K. D. Chandrasekhar, A. Lvov, C. Terrenoire, G. Y. Gao, R. S. Kass and W. R. Kobertz, *J. Physiol.*, 2011, **589**, 3721–3730.
159 A. Mant, S. Williams, L. Roncoroni, E. Lowry, D. Johnson and I. O'Kelly, *J. Biol. Chem.*, 2013, **288**, 3251–3264.
160 A. K. Walker, K. Y. Soo, V. Levina, G. H. Talbo and J. D. Atkin, *FEBS J.*, 2013, **280**, 233–243.
161 S. S. Pinho, R. Seruca, F. Gärtner, Y. Yamaguchi, J. Gu, N. Taniguchi and C. A. Reis, *Cell. Mol. Life Sci.*, 2011, **68**, 1011–1020.
162 C. Herraiz, B. L. Sánchez-Laorden, C. Jiménez-Cervantes and J. C. García-Borrón, *Pigm. Cell Melanoma Res.*, 2011, **24**, 479–489.
163 W. Huang, W. J. Luo, P. Zhu, J. Tang, X. L. Yu, H. Y. Cui, B. Wang, Y. Zhang, J. L. Jiang and Z. N. Chen, *Biochem. J.*, 2013, **449**, 437–448.
164 L. A. Cole, *Reprod. Biol. Endocrinol.*, 2009, **7**, 8.
165 D. L. Blithe and R. K. Iles, *Endocrinology*, 1995, **136**, 903–910.
166 J. J. van Rooijen, A. F. Voskamp, J. P. Kamerling and J. F. Vliegenthart, *Glycobiology*, 1999, **9**, 21–30.
167 V. Belalcazar, R. Gutiérrez-Gallego, E. Llop, J. Segura and J. A. Pascual, *Electrophoresis*, 2006, **27**, 4387–4395.
168 S. M. Muthana, C. T. Campbell and J. C. Gildersleeve, *ACS Chem. Biol.*, 2012, **7**, 31–43.
169 S. Otsuki, S. R. Hanson, S. Miyaki, S. P. Grogan, M. Kinoshita, H. Asahara, C. H. Wong and M. K. Lotz, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10202–10207.
170 J. M. Brown, J. Xia, B. Zhuang, K. S. Cho, C. J. Rogers, C. I. Gama, M. Rawat, S. E. Tully, N. Uetani, D. E. Mason, M. L. Tremblay, E. C. Peters, O. Habuchi, D. F. Chen and L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4768–4773.
171 I. Martínez-Martínez, J. Navarro-Fernández, A. Ostergaard, R. Gutiérrez-Gallego, J. Padilla, N. Bohdan, A. Miñano, C. Pascual, C. Martínez, M. E. de la Morena-Barrio, S. Aguila, S. Pedersen, S. R. Kristensen, V. Vicente and J. Corral, *Blood*, 2012, **120**, 900–904.
172 J. R. Bishop, M. Schuksz and J. D. Esko, *Nature*, 2007, **446**, 1030–1037.
173 A. Cariappa, H. Takematsu, H. Liu, S. Diaz, K. Haider, C. Boboila, G. Kalloo, M. Connole, H. N. Shi, N. Varki, A. Varki and S. Pillai, *J. Exp. Med.*, 2009, **206**, 125–138.
174 L. J. Olson, F. C. Peterson, A. Castonguay, R. N. Bohnsack, M. Kudo, R. R. Gotschall, W. M. Canfield, B. F. Volkman and N. M. Dahms, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12493–12498.
175 J. Tucholski, M. S. Simmons, A. L. Pinner, V. Haroutunian, R. E. McCullumsmith and J. H. Meador-Woodruff, *Schizophr. Res.*, 2013, **146**, 177–183.
176 Y. Zeng and X. Han, *J. Neurochem.*, 2008, **106**, 1275–1286.
177 X. Han, M. Holtzman, D. W. McKeel, Jr, J. Kelley and J. C. Morris, *J. Neurochem.*, 2002, **82**, 809–818.
178 K. I. Stanford, L. Wang, J. Castagnola, D. Song, J. R. Bishop, J. R. Brown, R. Lawrence, X. Bai, H. Habuchi, M. Tanaka, W. V. Cardoso, K. Kimata and J. D. Esko, *J. Biol. Chem.*, 2010, **285**, 286–294.
179 J. L. Stanta, R. Saldova, W. B. Struwe, J. C. Byrne, F. M. Leweke, M. Rothermund, H. Rahmoune, Y. Levin, P. C. Guest, S. Bahn and P. M. Rudd, *J. Proteome Res.*, 2010, **9**, 4476–4489.

180 J. Barnes, J. M. Lim, A. Godard, F. Blanchard, L. Wells and R. Steet, *J. Biol. Chem.*, 2011, **286**, 24855–24864.
181 C. Mandal, C. Mandal, S. Chandra, R. Schauer and C. Mandal, *Glycobiology*, 2012, **22**, 70–83.
182 M. Kobayashi, H. Hoshino, J. Masumoto, M. Fukushima, K. Suzawa, S. Kageyama, M. Suzuki, H. Ohtani, M. Fukuda and J. Nakayama, *Inflammatory Bowel Dis.*, 2009, **15**, 697–706.
183 B. Mann, E. Klussmann, V. Vandamme-Feldhaus, M. Iwersen, M. L. Hanski, E. O. Riecken, H. J. Buhr, R. Schauer, Y. S. Kim and C. Hanski, *Int. J. Cancer*, 1997, **72**, 258–264.
184 L. Zhao, M. Liu, Y. Gao, Y. Huang, G. Lu, Y. Gao, X. Guo and B. Shi, *Eur. J. Endocrinol.*, 2013, **168**, 585–592.
185 A. Kuga, M. Kanagawa, A. Sudo, Y. M. Chan, M. Tajiri, H. Manya, Y. Kikkawa, M. Nomizu, K. Kobayashi, T. Endo, Q. L. Lu, Y. Wada and T. Toda, *J. Biol. Chem.*, 2012, **287**, 9560–9567.
186 Y. Hara, B. Balci-Hayta, T. Yoshida-Moriguchi, M. Kanagawa, D. B. Beltrán-Valero, H. Gündesli, T. Willer, J. S. Satz, R. W. Crawford, S. J. Burden, S. Kunz, M. B. Oldstone, A. Accardi, B. Talim, F. Muntoni, H. Topaloglu, P. Dincer and K. P. Campbell, *N. Engl. J. Med.*, 2011, **364**, 939–946.
187 B. Adamczyk, T. Tharmalingam and P. M. Rudd, *Biochim. Biophys. Acta*, 2012, **1820**, 1347–1353.
188 K. B. Chandler and R. Goldman, *Mol. Cell. Proteomics*, 2013, **12**, 836–845.
189 Y. Mechref, Y. Hu, A. Garcia and A. Hussein, *Electrophoresis*, 2012, **33**, 1755–1767.
190 T. Angata, R. Fujinawa, A. Kurimoto, K. Nakajima, M. Kato, S. Takamatsu, H. Korekane, C. X. Gao, K. Ohtsubo, S. Kitazume and N. Taniguchi, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 159–169.
191 H. J. An and C. B. Lebrilla, *Methods Mol. Biol.*, 2010, **600**, 199–213.
192 H. Yu and X. Chen, *Org. Biomol. Chem.*, 2007, **5**, 865–872.
193 H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís, *Trends Biochem. Sci.*, 2011, **36**, 298–313.
194 S. A. Brooks and D. M. Hall, *Methods Mol. Biol.*, 2012, **878**, 31–50.
195 M. Lohr, H. Kaltner, R. Schwartz-Albiez, F. Sinowatz and H. J. Gabius, *Anat., Histol., Embryol.*, 2010, **39**, 481–493.
196 F. A. Habermann, S. André, H. Kaltner, D. Kübler, F. Sinowatz and H. J. Gabius, *Histochem. Cell Biol.*, 2011, **135**, 539–552.
197 C. J. Rogers, P. M. Clark, S. E. Tully, R. Abrol, K. C. Garcia, W. A. Goddard, III and L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 9747–9752.
198 X. Song, Y. Lasanajak, B. Xia, J. Heimburg-Molinaro, J. M. Rhea, H. Ju, C. Zhao, R. J. Molinaro, R. D. Cummings and D. F. Smith, *Nat. Methods*, 2011, **8**, 85–90.
199 Q. Tong, X. Wang, H. Wang, T. Kubo and M. Yan, *Anal. Chem.*, 2012, **84**, 3049–3052.
200 B. Lepenies, J. Yin and P. H. Seeberger, *Curr. Opin. Chem. Biol.*, 2010, **14**, 404–411.
201 T. K. Dam and C. F. Brewer, *Glycobiology*, 2010, **20**, 270–279.
202 T. K. Dam and C. F. Brewer, *Adv. Carbohydr. Chem. Biochem.*, 2010, **63**, 139–164.
203 G. Simone, P. Neuzil, G. Perozziello, M. Francardi, N. Malara, E. Di Fabrizio and A. Manz, *Lab Chip*, 2012, **12**, 1500–1507.
204 S. G. Lee, J. M. Brown, C. J. Rogers, J. B. Matson, C. Krishnamurthy, M. Rawat and L. C. Hsieh-Wilson, *Chem. Sci.*, 2010, **1**, 322–325.
205 A. Moise, S. André, F. Eggers, M. Krzeminski, M. Przybylski and H. J. Gabius, *J. Am. Chem. Soc.*, 2011, **133**, 14844–14847.
206 F. Real-Fernández, I. Passalacqua, E. Peroni, M. Chelli, F. Lolli, A. M. Papini and P. Rovero, *Sensors*, 2012, **12**, 5596–5607.
207 G. Bellapadrona, A. B. Tesler, D. Grünstein, L. H. Hossain, R. Kikkeri, P. H. Seeberger, A. Vaskevich and I. Rubinstein, *Anal. Chem.*, 2012, **84**, 232–240.
208 C. Jiménez-Castells, S. Defaus, A. Moise, M. Przbylski, D. Andreu and R. Gutiérrez-Gallego, *Anal. Chem.*, 2012, **84**, 6515–6520.
209 S. H. Millen, D. M. Lewallen, A. B. Herr, S. S. Iyer and A. A. Weiss, *Biochemistry*, 2010, **49**, 5954–5967.
210 J. L. Baker, E. Celik and M. P. Delisa, *Trends Biotechnol.*, 2013, **31**, 313–323.
211 C. Steentoft, S. Y. Vakhrushev, M. B. Vester-Christensen, K. T. Schjoldager, Y. Kong, E. P. Bennett, U. Mandel, H. Wandall, S. B. Levery and H. Clausen, *Nat. Methods*, 2011, **8**, 977–982.
212 K. T. Schjoldager, S. Y. Vakhrushev, Y. Kong, C. Steentoft, A. S. Nudelman, N. B. Pedersen, H. H. Wandall, U. Mandel, E. P. Bennett, S. B. Levery and H. Clausen, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 9893–9898.
213 L. X. Wang and J. V. Lomino, *ACS Chem. Biol.*, 2012, **7**, 110–122.
214 S. André, T. Kozár, R. Schuberth, C. Unverzagt, S. Kojima and H. J. Gabius, *Biochemistry*, 2007, **46**, 6984–6995.
215 S. André, T. Kozár, S. Kojima, C. Unverzagt and H. J. Gabius, *Biol. Chem.*, 2009, **390**, 557–565.
216 J. E. Rexach, C. J. Rogers, S. H. Yu, J. Tao, Y. E. Sun and L. C. Hsieh-Wilson, *Nat. Chem. Biol.*, 2010, **6**, 645–651.
217 J. L. Chaubard, C. Krishnamurthy, W. Yi, D. F. Smith and L. C. Hsieh-Wilson, *J. Am. Chem. Soc.*, 2012, **134**, 4489–4492.
218 L. Yang, J. O. Nyalwidhe, S. Guo, R. R. Drake and O. J. Semmes, *Mol. Cell. Proteomics*, 2011, **10**, M110.
219 T. N. Ramya, E. Weerapana, B. F. Cravatt and J. C. Paulson, *Glycobiology*, 2013, **23**, 211–221.
220 R. Orlando, J. M. Lim, J. A. Atwood, III, P. M. Angel, M. Fang, K. Aoki, G. Alvarez-Manilla, K. W. Moremen, W. S. York, M. Tiemeyer, M. Pierce, S. Dalton and L. Wells, *J. Proteome Res.*, 2009, **8**, 3816–3823.
221 K. Tanaka, K. Moriwaki, S. Yokoi, K. Koyama, E. Miyoshi and K. Fukase, *Bioorg. Med. Chem.*, 2013, **21**, 1074–1077.
222 V. Zoldos, T. Horvat and G. Lauc, *Curr. Opin. Chem. Biol.*, 2012, **17**, 34–40.
223 S. Neelamegham and G. Liu, *Glycobiology*, 2011, **21**, 1541–1553.

APPENDIX II:

Carmen Jiménez-Castells, Sira Defaus, Adrian Moise, Michael Przbylski, David Andreu, and Ricardo Gutiérrez-Gallego. "Surface-based and mass spectrometric approaches to deciphering sugar−protein interactions in a galactose-specific agglutinin". *Anal. Chem.* **84**, 6515-6520 (2012).

analytical chemistry

Article

pubs.acs.org/ac

Surface-Based and Mass Spectrometric Approaches to Deciphering Sugar−Protein Interactions in a Galactose-Specific Agglutinin

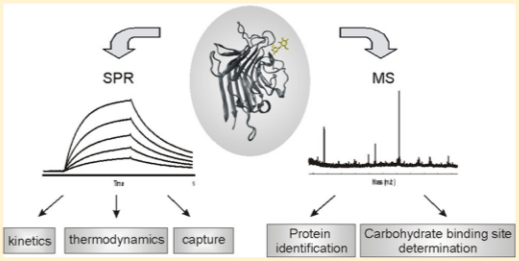
Carmen Jiménez-Castells,[†] Sira Defaus,[†] Adrian Moise,[‡] Michael Przbylski,[‡,*] David Andreu,*[,†] and Ricardo Gutiérrez-Gallego*[,†,§]

[†]Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

[‡]Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany

[§]Bio-analysis group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

ABSTRACT: Interest in powerful, nanosized tools to analyze in detail glycan−protein interactions has increased significantly over recent years. Here, we report two complementary approaches to characterize such interactions with high sensitivity, low sample consumption, and without the need for sample labeling, namely, surface plasmon resonance (SPR) and an approach that combines limited proteolysis and mass spectrometry. Combination of these two approaches to investigate glycan−protein interactions allows (1) to characterize interactions through kinetic and thermodynamic parameters, (2) to capture efficiently the carbohydrate-binding protein, and (3) to identify the interacted protein and its carbohydrate binding site by mass spectrometry. As a proof of principle, the interaction of the galactose-specific legume lectin *Erythrina cristagalli* agglutinin with several sugars has been characterized in-depth by means of these two approaches.



In recent years the interest in exploring carbohydrate−protein interactions has soared as their decisive role in many biological processes, for example, pathogen−host cell adhesion,[1] metastasis,[2] or fertilization,[3] became evident. Although it is estimated that over 50% of all proteins are glycosylated,[4] and that these glycans play key roles in all sorts of communication processes, little is known about their binding partners. Several techniques for screening interaction partners, mostly focused on protein−protein binding, have been developed, including tandem affinity purification[5] or two-hybrid screening.[6] More recently, surface plasmon resonance (SPR) has also been employed to capture new binding partners prior to characterization by mass spectrometry[7] or HPLC profiling.[8] In this approach one of two interacting entities is immobilized onto the surface of a sensor chip and its partner is flown over. Unlike other techniques, SPR provides both quantitative and qualitative interaction data, in real time and under conditions closely mimicking physiological ones. Although originally designed for protein−protein studies,[9] soon it was applied in other contexts, for example, DNA−protein,[10] sugar−protein,[11] and lipid−protein interactions.[12] While SPR-based biosensors are mostly used to determine kinetic rate constants, thermodynamic data can also be obtained by determining equilibrium constants within a given temperature range.[13] Moreover, some instrumental designs incorporate sample recovery features that allow to combine interaction analysis with MS identification.[14] In the particular case of sugar−protein interactions, we have devised a reliable method for immobilizing glycan-displaying probes on SPR chips.[15] In this approach, the sugar moiety is immobilized through a tailor-made peptide module on the sensor surface and the interacting lectin is passed over. The oxime ligation chemistry is used to attach the glycan via the aminooxy functionality (Aoa) to the peptide module, Aoa-GFKKG-amide,[16] whereas the methylated version, *N*[Me]-O-Aoa-GFKKG-amide ensures correct exposure of the carbohydrate on the chip surface as well as the conformational integrity (e.g., as a pyranose ring) of the monosaccharide unit proximal to the surface, a particularly relevant point for short (mono- and disaccharide) epitopes.[17] The two Lys residues in the peptide module guide coupling to the carboxyl-functionalized sensor surface, previously activated as *N*-hydroxysuccinimide ester.

For detailed structural information on lectin−glycan interactions, well-established techniques, such as X-ray crystallography[18] and NMR spectroscopy,[19] require amounts of both partners in the milligram range and purity levels not often easily achievable. Recently, a novel approach combining proteolytic digestion of protein−glycan complexes and mass spectrometry (CREDEX-MS, carbohydrate recognition domain excision mass spectrometry)[20,21] has demonstrated its efficiency

Received: March 19, 2012
Accepted: July 3, 2012
Published: July 3, 2012

in the identification and structural definition of carbohydrate binding sites. In this approach, the carbohydrate is coupled to a divinylsulfone-activated Sepharose support and the lectin-containing specimen is bound to it under controlled conditions. The carbohydrate-lectin complex is then digested with specific proteases and, after washing-off non-interacting fragments, the binding peptides are eluted and subsequently identified by MS. Like in all MS techniques the sample amount requirements are substantially lower (micrograms vs milligrams) than for X-ray crystallography or NMR.

Here we describe how the combined use of SPR and CREDEX-MS, two nanosized complementary analytical methodologies, can provide detailed information on carbohydrate-lectin interactions. As an example, the interaction between the β-galactose-specific legume lectin *Erythrina cristagalli* agglutinin (ECA) and a series of related β-galactosides [Gal(β1−4)GlcNAc, Gal(β1−4)Glc, Gal(β1−3)-GlcNAc, and Gal(β1−6)GlcNAc) has been studied by both SPR and MS, and the carbohydrate-binding site of the lectin has been identified.

■ MATERIALS AND METHODS

Chemicals. Fmoc [N^{α}-(9-fluorenylmethyloxycarbonyl)]-protected amino acids were purchased from Senn Chemicals (Dielsdorf, Switzerland). The dicyclohexylamine (DCHA) salt of Boc (*tert*-butyloxycarbonyl)-methylaminooxyacetic acid was from NeoMPS (Strasbourg, France). 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) was obtained from Iris Biotech (Marktredwitz, Germany). Rink amide MBHA resin was from Novabiochem (Läufelfingen, Switzerland). *N*,*N*-Diisopropylethylamine (DIEA) was from Merck Biosciences (Darmstadt, Germany), and triisopropylsilane was from Sigma-Aldrich (Madrid, Spain). HPLC-grade acetonitrile (ACN), *N*,*N*-dimethylformamide (DMF), trifluoroacetic acid (TFA), and diethyl ether were from SDS (Peypin, France). Disaccharides (Gal(β1−3,4,6)GlcNAc) were purchased from Dextra (Reading, United Kingdom). Lactose (Gal(β1−4)Glc) and lectin from *Erythrina cristagalli* (ECA) were from Sigma-Aldrich (Madrid, Spain). CM5 sensor chips, 1-ethyl-3-(3-diethylaminopropyl)-carbodiimide (EDC), *N*-hydroxysuccimide (NHS), and ethanolamine hydrochloride pH 8.5, were from BIAcore (GE Healthcare, Uppsala, Sweden), Sepharose-4B and divinylsulfone were purchased from Sigma-Aldrich (Madrid, Spain). Microcolumn and 35-μm pore size filters were from MoBiTec (Göttingen, Germany). Sequencing-grade modified porcine trypsin was from Promega (Madison, USA). Sequencing-grade chymotrypsin and Glu-C were from Roche Diagnostics GmbH (Penzberg, Germany). Poros 20 R2 was obtained from Applied BioSystems (Foster City, USA).

Peptide and Glycopeptide Synthesis. *N*[Me]-*O*-Aoa-GFKKG-amide[17] was synthesized by Fmoc-based solid-phase synthesis on a Rink MBHA resin (0.70 mmol/g). Boc-methylaminooxyacetic acid·DCHA (500 mg) was converted to the free carboxyl form by acid extraction with 0.1 M HCl and ethyl acetate (50 mL each). Manual couplings with 3 equiv each of Boc-amino acid and HBTU and 6 equivalent of DIEA were used for 1 h, r.t., in DMF. Resin cleavage and full deprotection were done with TFA−water−triisopropylsilane (95:2.5:2.5, v/v, 90 min, r.t.). Peptides were isolated by precipitation with cold *tert*-butyl-methyl ether and centrifugation, then solubilized in water and lyophilized. The synthetic product was >95% pure by analytical HPLC and had the correct mass by MALDI-TOF MS.

Conjugation of *N*[Me]-*O*-Aoa-GFKKG-amide to disaccharides was done at 20 and 25 mM, respectively in 0.1 M sodium acetate, at pH 3.5 for *N*Ac-disaccharides and pH 4.6 for lactose. After 72 h at 37 °C, the glycopeptides were purified by semipreparative HPLC on SphereClone C18 (Phenomenex, 250 × 10 mm; 5 μm) using a 10−20% linear gradient of acetonitrile into water (both eluents with 0.1% TFA). Glycopeptide-containing fractions were neutralized with 10 mM ammonium bicarbonate to prevent acid degradation and lyophilized. All disaccharide-*N*[Me]-*O*-GFKKG-amide glycopeptides had the expected mass by MALDI-TOF MS.

SPR Measurements. Experiments were performed on carboxymethyl-functionalized CM5 sensor chips in a Biacore 3000 instrument (Biacore SA, Uppsala, Sweden). In all experimental procedures the running buffer was HBS-P buffer supplemented with 5 mM $CaCl_2$ and 1 mM $MnCl_2$. For *N*[Me]-*O*-Aoa-GFKKG-amide immobilization, the surface was activated with 60 μL of a mixture of EDC (0.2 M) and NHS (0.05 M) in water, at 5 μL/min, then glycopeptides were passed at 1 mg/mL in 10 mM sodium acetate, pH 6, for 12 min, followed by a blocking step with 1 M ethanolamine·HCl, pH 8.5. As a reference surface, *N*[Me]-*O*-Aoa-GFKKG-amide was immobilized on a different flow channel.

For determination of kinetic parameters, several concentrations in the 66 nM−2.6 μM range were prepared by a set of 2/3-fold dilutions of the most concentrated sample in running buffer (10 mM HEPES, 25 mM sodium chloride, 5 mM calcium chloride, 1 mM manganese(II) chloride,[22] pH 7.4). Binding experiments were performed at 25 °C at two flow rates (10 and 50 μL/min). After lectin injection (3-min pulse), sample solution was replaced by running buffer and the dissociation phase was monitored for 6 min. Sensor surface was regenerated with a 50 μL-injection of 10 mM lactose. Two replicates were performed for each injection.

For thermodynamic experiments, five ECA concentrations (100, 250, 400, 550, and 700 nM) in running buffer were explored in the 10 to 25 °C interval, with 2.5 °C increments controlled by a Peltier device. Two replicates of each solution were injected (3-min pulse) over the *N*-acetyllactosamine-(lacNAc)-funcionalized and the reference surface at 50 μL/min. As in the kinetic experiments, after lectin injection sample injection was replaced by running buffer and the dissociation phase was monitored for 6 min. After each cycle, the sensor surface was regenerated with 50-μL injections of 25 mM lactose. In both kinetic and thermodynamic experiments, the specific binding response was obtained by subtracting form each channel the reference channel response. Curve fitting of the sensorgram curves was done with the BIAevaluation 4.0.1 software package.

For recovery experiments, a 1 μM solution of ECA was injected for 3 min at 10 μL/min over the lacNAc surface at 25 °C. By means of the MS recover function, captured material was eluted in a 2-μL volume of 10 mM lacNAc, concentrated by vacuum centrifugation and analyzed by MALDI-TOF MS.

CREDEX-MS. For disaccharide immobilization, 50 μg of dry divinylsulfonyl-activated Sepharose were treated with a solution of 5 mg of lacNAc in 50 μL of 0.5 M potassium carbonate, pH 11, overnight at r.t. The reaction mixture was poured into a microcolumn and washed sequentially with 50 mM ammonium acetate, pH 4, and 50 mM and ammonium bicarbonate, pH 8. The microcolumn was equilibrated with SPR running buffer and stored at 4 °C. For excision experiments, 20 μg of ECA in running buffer (75 μL) were loaded onto the lacNAc-Sepharose

dx.doi.org/10.1021/ac300766z | *Anal. Chem.* 2012, 84, 6515−6520

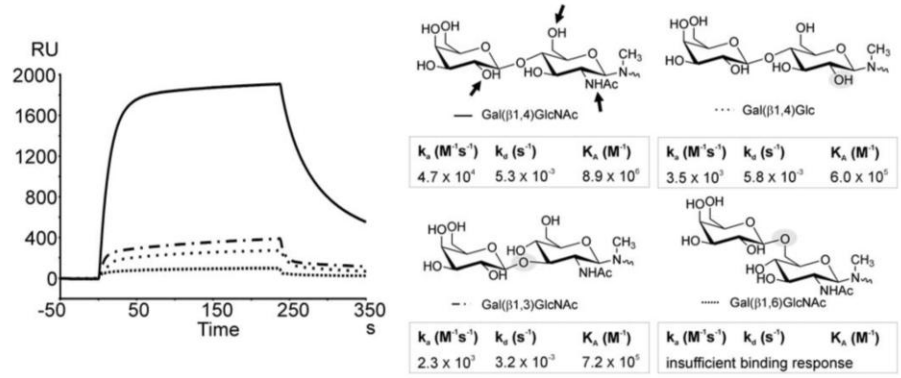
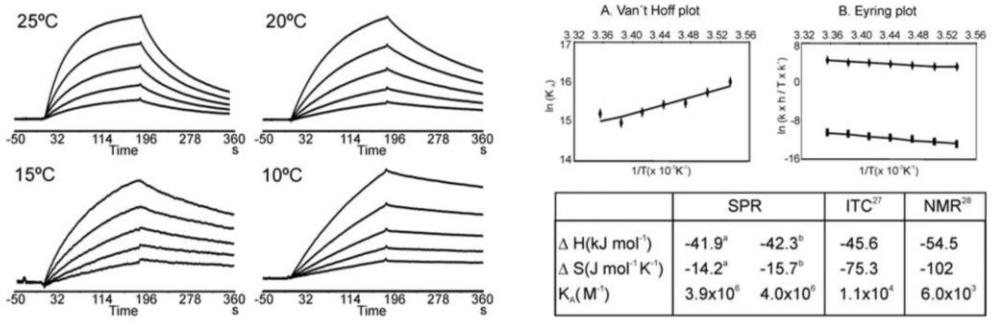


Figure 1. SPR analysis of ECA binding to four β-galactosides. Glycopeptide probes displaying the epitopes were coupled to the sensor surface at similar immobilization levels. The sensorgrams show the differential curves after subtracting a reference channel with no epitope immobilized. Gal(β1−4)GlcNAc functional groups crucially involved in ECA interaction are marked with an arrow. In the other disaccharide structures, functional groups whose modification causes loss of ECA affinity are shaded. The kinetic rate constants (k_a, k_d) and the derived affinity constant (K_A) of ECA to the four different glycoprobes exposing terminal β-galactosyl-disaccharides are provided in the boxes.



	SPR		ITC[27]	NMR[28]
Δ H(kJ mol^{-1})	-41.9[a]	-42.3[b]	-45.6	-54.5
Δ S(J mol^{-1}K^{-1})	-14.2[a]	-15.7[b]	-75.3	-102
K_A(M^{-1})	3.9x10^5	4.0x10^6	1.1x10^4	6.0x10^3

Figure 2. Left: SPR sensorgrams demonstrating the effect of temperature on the interaction between ECA and Gal-β1,4-GlcNAc. Top right: Van't Hoff plot for the binding of ECA to lacNAc. Dots showing affinity constants determined at 2.5 °C intervals in the 10−25 °C range could be linearly fitted to derive ΔH values. B. Eyring plots for the same interaction. Dots and squares correspond to association and dissociation rate constants, respectively; both series could be linearly fitted to derive ΔH values. Bottom right: Thermodynamic parameters for ECA-lacNAc interactions determined using SPR, ITC and NMR. In the SPR block the thermodynamic parameters derived from the (a) Van't Hoff or (b) Eyring equation.

microcolumn and incubated for 24 h at 37 °C, followed by washes with binding buffer. The sugar-lectin complex was then digested overnight with 1 μg trypsin in 75 μL of 25 mM ammonium bicarbonate, pH 8.5, 37 °C. Unbound digestion products were eluted and the column washed with running buffer. For chymotrypsin digestion, 1 μg enzyme in 75 μL of 100 mM Tris·HCl, 10 mM calcium chloride, pH 7.8, was added to the microcolumn and incubated for 24 h at 35 °C. After washes with running buffer, specific-bound peptides were eluted with 600 μL of acetonitrile−water (6:4; with 0.1% TFA), concentrated and lyophilized. Prior to analysis, digestion peptides were desalted on Poros R2 mini-columns packed in Geloader tips. MALDI-TOF MS measurements were carried out on a Voyager-DE STR workstation (Applied Biosystems) operating in the reflectron, positive polarity mode, with data processing by the Data Explorer Software (Applied Biosystems), or in a Bruker Biflex linear TOF mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a nitrogen UV laser (λ_{max} 337 nm) and a XMASS data system for spectra acquisition and instrument control.

RESULTS AND DISCUSSION

Structural Insights from SPR Studies. The medium-throughput screening capacity of our Biacore 3000 instrument allowed simultaneous analysis of ECA binding to four β-galactoside-containing epitopes: Gal(β1−4)GlcNAc, Gal(β1−4)Glc, Gal(β1−3)GlcNAc, and Gal(β1−6)GlcNAc. Kinetic rate constants (k_a, k_d) for both association and dissociation phases, and the derived affinity constants (K_A) could be determined for the first three disaccharides (Figure 1); for Gal(β1−6)GlcNAc the response was too low for reliable quantitative data to be derived. In addition to binding parameters, SPR results provided helpful structural insights into the binding events, particularly about the functional groups involved in each interaction. Thus, binding to ECA was strongly influenced by the type of glycosidic linkage, as well as by the nature of the monosaccharide at the reducing end, with a clear preference for Gal(β1−4)GlcNAc, in agreement with previous studies.[23] For this most favorable epitope, the 15-fold higher affinity over Gal(β1−4)Glc underlines the significant role of the *N*-acetyl group at position C2 in lectin binding. Also, by comparing the responses of the three *N*-acetyl-disaccharides,

dx.doi.org/10.1021/ac300766z | *Anal. Chem.* 2012, 84, 6515−6520

the relative role of the hydroxyls of the nonterminal sugar in the interaction can be ascertained. Thus, the decreased binding of the β1−3 isomer (8% relative to the β1−4), or the even lower affinity of the β1−6-linked disaccharide suggest that both C6 and C3 hydroxyls are significantly involved in the canonic binding of Gal(β1−4)GlcNAc to ECA, so that when either of these hydroxyls is obliged to engage in glycosidic bond formation impaired affinity ensues. In summary, straightforward inspection of SPR data highlights a key role of the *N*-acetyl group and, to a lesser extent, of the C3 and C6 hydroxyls, in sugar−ECA recognition, in good agreement with X-ray data showing the O3, N2, O6, and O_{NAc} atoms to be directly involved in the interaction.[24]

SPR-Derived Thermodynamic Parameters. In addition to the kinetic and structural information discussed above, thermodynamic parameters for the preferential Gal(β1−4)GlcNAc (lacNAc)−ECA interaction could also be determined in real time by monitoring the SPR response at various temperatures. Figure 2 (left) shows ECA binding profiles in the 10 to 25 °C range, and how temperature rise affected (i.e., accelerated) both association and dissociation steps. Both k_a and k_d rate constants were determined at each temperature by locally fitting the sensorgrams to the five concentrations used. The derived association constants (K_A) at each temperature were then used to calculate thermodynamic parameters by means of the Van't Hoff equation:

$$\ln K_A = -\Delta H^\circ/RT + \Delta S^\circ/R$$

where R = 8.314 J K^{-1} mol^{-1}.

Alternatively, thermodynamic parameters could also be determined from each rate constant (k_a, k_d), independently, by means of the Eyring equation

$$\ln(k/T) = -\Delta H/RT + -\Delta S/R + \ln(k'/h)$$

where k is the appropriate rate constant and k' and h are the Boltzmann (k' = 1.380 × 10^{-23} J K^{-1}) and the Planck (h = 6.626 × 10^{-34} J s) constants, respectively.

Both Van't Hoff and Eyring plots (Figure 2 Top right A, B) could be fitted to a linear model. The ΔH values derived from each analysis (−41.9 and −42.3 kJ mol^{-1}, respectively) were in good agreement with the −45.6 and −54.5 kJ mol^{-1} values from ITC and NMR studies, respectively.[25,26]

In all these approaches, entropic values are calculated from the Gibbs free energy equation employing the equilibrium association constant and experimental enthalpy as recommended.[27] The SPR-derived equilibrium constant for lacNAc ($K_A \approx 10^6$ M^{-1} at 25 °C, Figure 2) is about 2 orders of magnitude higher than previously reported ITC- and NMR-derived values for the same disaccharide in free form ($K_A \approx 10^4$ M^{-1}, Figure 2 bottom right). Similar differences between surface- and solution-based methods have been already observed for other carbohydrate−lectin interactions,[28] probably because lectin multivalency in addition to the rather high glycan surface density required for SPR experiments may increase the apparent affinity through secondary interactions. In any event, such differences in equilibrium constants, compounded with the slightly different ΔH values from SPR and ITC/NMR methods, can explain the discrepancy in entropic values found in Figure 2 (bottom right).

SPR Lectin Capture and MALDI-TOF MS Identification. In addition to quantitative kinetic data, SPR can serve as an affinity capture/purification platform allowing subsequent MS identification. This is shown here for the interaction of ECA with Gal(β1−4)GlcNAc, the glycoprobe with the highest affinity. The lectin, a 55 kDa dimer, gave a 790 RU readout, that is, an estimated 790 pg of surface-bound material. This was then specifically eluted with excess disaccharide by means of the instrument's recovery function, devised to retrieve affinity-captured material in a very small (2 μL) volume, separated by air bubbles to avoid sample diffusion or cross-contamination. Concentration by vacuum centrifugation furnished enough material (ca. 15 fmol) for molecular weight determination (Figure 3). In this particular case, the fast dissociation of ECA

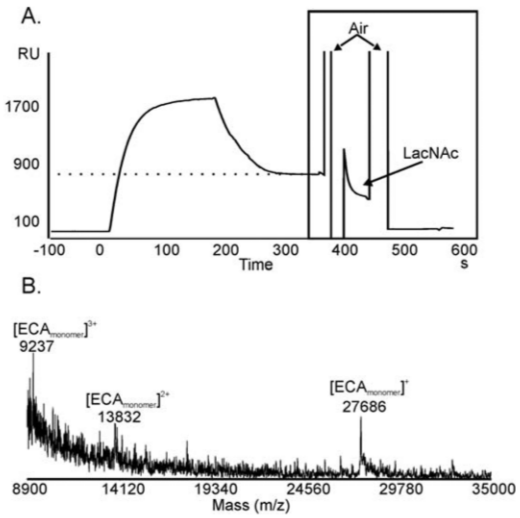


Figure 3. (A) SPR sensorgram of the recovery experiment over lacNAc-glycoprobe. ECA at 1 μM was passed on the glycosylated surface and the captured material (~15 fmol) was recovered with 2 μL lacNAc. (B) MALDI-TOF MS spectrum of the recovered protein.

(k_d = 5.3 × 10^{-3} s^{-1}) complicated recovery and subsequent analysis, compared to other lectins with lower dissociation rates and molecular weights (e.g., wheat-germ agglutinin, data not shown).

In conclusion, the combination of SPR and MALDI-TOF MS molecular weight determination allows successful characterization of sugar−lectin interactions, provided the dissociation rate of the complex and the desorption ability of the lectin are favorable enough.

Carbohydrate-Binding Site Determination by CREDEX-MS. Practically all the current approaches to protein characterization rely on the combined use of proteolysis and MS methods. Direct tryptic digestion of lectin−glycan−*N*[Me]-*O*-Aoa-GFKKG complexes on the SPR chip did not seem advisable, because the Lys residue not used for anchoring the glycoprobe to the chip surface is trypsin-susceptible and cleavage at this site would cause the affinity-bound lectin material to be lost. We therefore opted for off-line proteolysis in a divinylsulfone-based carbohydrate affinity column.[29]

To identify a carbohydrate-binding site by CREDEX-MS, a protease with optimal sequence coverage of the carbohydrate-lectin complex (lacNAc-ECA in our case) is a must. Among three standard proteases, trypsin [1:20 (w:w) ratio] was chosen because it gave 86% coverage (13 peptides, Figure 4A) vs 30% and 26% for chymotrypsin and Glu-C, respectively. Compar-

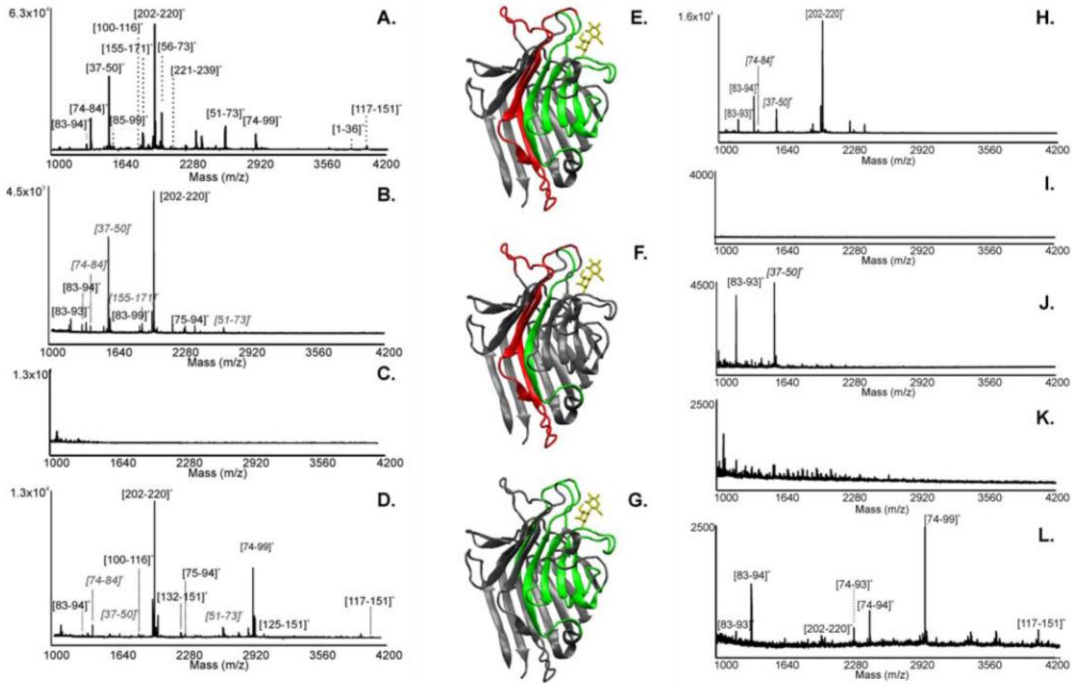


Figure 4. (A) Peptide mass fingerprint of ECA (digestion in solution). (B–D) MALDI-TOF MS spectra corresponding to different fractions of excision experiment. (B) On column digestion with trypsin of the complex lacNAc-ECA. (C) Supernatant after washing. (D) Elution fraction. Peaks in gray show the peptides with no direct contact with the sugar. (Center panel E–G) X-ray crystal structure (PDB 1GZC) of ECA in complex with lactose (in yellow). (E) Peptides identified in the elution fraction of an excision experiment with trypsin are shown in green (sugar-peptide interaction) and red (peptide–peptide interaction) in the ribbon representation. (F) Peptides [37–50], [51–73] and [74–84] (in red), noncovalently bound with spatially close [202–220] (in green), "ride" with this peptide in the elution fraction. (G) In an excision experiment with two consecutive digestions, only peptides involved in sugar-peptide interactions (in green) are detected.[22] (Right panel H–L) MALDI-TOF MS spectra corresponding to different fractions of excision experiment with two consecutive proteolytic digestions. (H) First on column digestion with trypsin of the complex lacNAc-ECA. (I) Supernatant after washing. (J) Second on column digestion with chymotrypsin. (K) Supernatant after washing. (L) Elution fraction.

ison of the standard (Figure 4A) with the flow-through from on-column digest (Figure 4B) revealed essentially identical mass fingerprint peaks with only minor differences in intensity. The column was next washed until no peptide signals were observed (Figure 4C) and then the glycan-interacting peptides were eluted with 60% acetonitrile in water and analyzed (Figure 4D). In this fraction, several peptides {[74–99], [75–94], [83–94], [100–116], [117–151], and [202–220]} (Figure 4D) that contain amino acids displaying direct contact with the carbohydrate in the X-ray structure of the Lac-ECA complex (Figure 4E) could be unequivocally assigned to the binding site. In addition to these specific peptides, three peptides found in low abundance {[37–50], [51–73], and [74–84]} showed no contacts with the sugar in the X-ray structure (Figure 4F); hence no relation to the binding site. Their presence in the elution fraction was explained as a case of "riding" (via β-strand interaction) with the spatially close, glycan-binding peptide [202–220].

To confirm the binding site identification, an additional chymotrypsin digestion subsequent to the trypsin excision was performed. Figure 4 (right panel H–L) shows that, after sequential trypsin-chymotrypsin digestions, the longest peptide [202–220] is split into shorter fragments and the "riding peptides", [37–50], [51–73], and [74–84], are no longer observed in the elution fraction, while the specific binding peptides [74–99], [117–151], as well as nondigested [202–220] remain present (Figure 4L). This result proves that the Gal(β1–4)GlcNAc-ECA interaction can withstand prolonged, sequential digestion with two proteases, despite its relatively low affinity (see Figure 1), hence making possible accurate molecular definition of the interaction partners and removal of unspecific peptides. This is valuable when proteolytic excision results in long peptides enhancing the likelihood of peptide–peptide interactions that may hamper the identification of the carbohydrate binding site.

Additional unequivocal identification of the binding site came from an extraction MS experiment,[20] where ECA was first digested with trypsin, the digest was passed through the affinity column, and only peptides [202–220] and [74–99] representing the binding site were observed (data not shown).

■ CONCLUSION

Detailed molecular description of carbohydrate–protein interactions is feasible by the combination of analytical

techniques described here, which use low amounts of both lectin and carbohydrate compatible with extraction from natural sources, in contrast with more sample-demanding techniques such as NMR or X-ray crystallography. The agglutinin ECA has been chosen as a case study to test the applicability of these techniques. SPR-based experiments showed a higher affinity of ECA for lacNAc relative to other β-galactosides. Additional structural information on the interaction was also provided by SPR, by comparing the differential binding responses between epitopes with subtle differences (i.e., glycosidic linkage or *N*-acetyl group at position C2). This analysis showed the hydroxyls at C3 and C6, as well as the *N*-acetyl at C2, being critical for interaction with ECA. Thermodynamic data on the interaction were also derived by SPR. While enthalpy values were equivalent to those obtained by ITC or NMR, the higher affinity constants determined by SPR translated into larger differences in entropy relative to ITC or NMR. Finally, SPR technology was successfully applied as a lectin capture platform for subsequent MS analysis. SPR-based results are shown to be an efficient combination with CREDEX-MS, which provides a molecular definition of the carbohydrate-binding site. CREDEX-MS in conjunction with the SPR approach described here constitutes a valuable set of tools for decrypting carbohydrate–protein interaction details.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rgutierrez@imim.es (R.G.-G.); david.andreu@upf.edu (D.A); michael.przybylski@uni-konstanz.de (M.P.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation (projects BIO2008-04487-CO3-02 and HA2007-0021 to D.A. and BIO2009-08983 to R.G.G., and predoctoral fellowship BES-2006-12879 to C.J.C.), by the Spanish Ministry of Economy and Competitiveness (project SAF2011-24899 to D.A.), and by the Deutsche Forschungsgemeinschaft (PR-175/14-1) and the EU (MSLife).

■ REFERENCES

(1) Imberty, A.; Varrot, A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 567–576.
(2) Rambaruth, N. D.; Dwek, M. V. *Acta Histochem.* **2011**, *113*, 591–600.
(3) Velasquez, J. G.; Canovas, S.; Barajas, P.; Marcos, J.; Jimenez-Movilla, M.; Gutiérrez-Gallego, R.; Ballesta, J.; Aviles, M.; Coy, P. *Mol. Reprod. Dev.* **2007**, *74*, 617–628.
(4) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
(5) Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. *Methods* **2001**, *24*, 218–229.
(6) Bruckner, A.; Polge, C.; Lentze, N.; Auerbach, D.; Schlattner, U. *Int. J. Mol. Sci.* **2009**, *10*, 2763–2788.
(7) Madeira, A.; Ohman, E.; Nilsson, A.; Sjogren, B.; Andren, P. E.; Svenningsson, P. *Nat. Protoc.* **2009**, *4*, 1023–1037.
(8) Gutiérrez-Gallego, R.; Haseley, S. R.; van Miegem, V. F.; Vliegenthart, J. F.; Kamerling, J. P. *Glycobiology* **2004**, *14*, 373–386.
(9) Slepak, V. Z. *J. Mol. Recognit.* **2000**, *13*, 20–26.
(10) Ahmed, F. E.; Wiley, J. E.; Weidner, D. A.; Bonnerup, C.; Mota, H. *Cancer Genomics Proteomics* **2010**, *7*, 303–309.
(11) Duverger, E.; Frison, N.; Roche, A. C.; Monsigny, M. *Biochimie* **2003**, *85*, 167–79.
(12) Besenicar, M.; Macek, P.; Lakey, J. H.; Anderluh, G. *Chem. Phys. Lipids* **2006**, *141*, 169–178.
(13) Roos, H.; Karlsson, R.; Nilshans, H.; Persson, A. *J. Mol. Recognit.* **1998**, *11*, 204–210.
(14) Zhukov, A.; Schurenberg, M.; Jansson, O.; Areskoug, D.; Buijs, J. *J. Biomol. Tech.* **2004**, *15*, 112–119.
(15) Vila-Perello, M.; Gutiérrez-Gallego, R.; Andreu, D. *ChemBioChem* **2005**, *6*, 1831–1838.
(16) Jimenez-Castells, C.; de la Torre, B. G.; Gutiérrez-Gallego, R.; Andreu, D. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5155–5158.
(17) Jimenez-Castells, C.; de la Torre, B. G.; Andreu, D.; Gutiérrez-Gallego, R. *Glycoconj. J.* **2008**, *25*, 879–887.
(18) Weis, W. I.; Drickamer, K. *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
(19) Asensio, J. L.; Cañada, F. J.; Bruix, M.; Gonzalez, C.; Khiar, N.; Rodriguez-Romero, A.; Jimenez-Barbero, J. *Glycobiology* **1998**, *8*, 569–577.
(20) Moise, A.; Andre, S.; Eggers, F.; Krzeminski, M.; Przybylski, M.; Gabius, H. J. *J. Am. Chem. Soc.* **2011**, *133*, 14844–14847.
(21) Przybylski, M., Moise, A., Gabius, H. J. Identification Of Ligand Recognition Domains. Patent [PCT/EP2009/003495], 2012, 2010.
(22) Svensson, C.; Teneberg, S.; Nilsson, C. L.; Kjellberg, A.; Schwarz, F. P.; Sharon, N.; Krengel, U. *J. Mol. Biol.* **2002**, *321*, 69–83.
(23) Itakura, Y.; Nakamura-Tsuruta, S.; Kominami, J.; Sharon, N.; Kasai, K.; Hirabayashi, J. *J. Biochem.* **2007**, *142*, 459–469.
(24) Elgavish, S.; Shaanan, B. *J. Mol. Biol.* **1998**, *277*, 917–932.
(25) Gupta, D.; Cho, M.; Cummings, R. D.; Brewer, C. F. *Biochemistry* **1996**, *35*, 15236–15243.
(26) De Boeck, H.; Loontiens, F. G.; Lis, H.; Sharon, N. *Arch. Biochem. Biophys.* **1984**, *234*, 297–304.
(27) Roos, H.; Karlsson, R.; Nilshans, H.; Persson, A. *J. Mol. Recognit.* **1998**, *11*, 204–210.
(28) Critchley, P.; Clarkson, G. J. *Org. Biomol. Chem.* **2003**, *1*, 4148–4159.
(29) Stefanescu, R.; Born, R.; Moise, A.; Ernst, B.; Przybylski, M. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 148–157.

APPENDIX III:

Carmen Jiménez-Castells, Sira Defaus, David Andreu, and Ricardo Gutiérrez-Gallego. "Recent progress in the field of neoglycoconjugate chemistry". *BioMol Concepts*. **1**, 85-96 (2010).

BioMol Concepts, Vol. 1 (2010), pp. 85–96 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BMC.2010.007

Review

Recent progress in the field of neoglycoconjugate chemistry

Carmen Jiménez-Castells[1], Sira Defaus[1], David Andreu[1] and Ricardo Gutiérrez-Gallego[1,2,*]

[1]Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, Dr. Aiguader 88, 08003 Barcelona, Spain
[2]Pharmacology Research Unit, Bio-analysis group, Neuropsychopharmacology program, Municipal Institute of Medical Research IMIM-Hospital del Mar, Barcelona Biomedical Research Park, Dr. Aiguader 88, 08003 Barcelona, Spain

*Corresponding author
e-mail: rgutierrez@imim.es; ricardo.gutierrez@upf.edu

Abstract

Glycosylation is probably the most complex secondary gene event that affects the vast majority of proteins in nature resulting in the occurrence of a heterogeneous mixture of glycoforms for a single protein. Many functions are exerted by single monosaccharides, well-defined oligosaccharides, or larger glycans present in these glycoproteins. To unravel these functions it is of the utmost importance to prepare well-defined single glycans conjugated to the underlying aglycon. In this review, the most recent developments are described to address the preparation of carbohydrate-amino acid (glycoconjugates). Naturally occurring N- and O-linked glycosylation are described and the preparation of non-natural sugar-amino acid linkages are also included.

Keywords: biomolecular interactions; glycosylation; neoglycopeptides; synthesis.

Introduction

With the sequencing of an ever-increasing number of genomes, and the number of genes encountered, it has become very clear that the primary gene products are mostly unique structural scaffolds and that post-translational modifications (PTMs) of the encoded proteins actually provide the structural diversity required for function. In this context, glycosylation represents the most extended and complex type of PTM, ranging from the attachment of a single monosaccharide [e.g., fucose to Thr61 in human tissue plasminogen activator (1)] to far more complex structural arrangements (e.g., tetrasialylated tetraantennary core-fucosylated complex type N-glycans) in recombinant human erythropoietin (EPO) (2) (Figure 1).

Carbohydrate attachment to the backbone, usually occurring at the protein surface, entails not only modest-to-substantial structure alteration but also often the generation of differently glycosylated variants of a single gene product. Glycosylation has been extensively studied in eukaryotes (3–6) and evidence is growing that in prokaryotes it is also more common than hitherto supposed (7, 8). Glycans have been associated with many biological events such as fertilization (9, 10), cell growth (11), tumor growth/metastasis (12), immune reactions (13, 14), cell communication (15–17), or infections (18, 19). Glycosylation, as opposed to glycation, is an enzymatic process that can occur in the endoplasmic reticulum (ER) as a cotranslational event or in the Golgi apparatus as the newly synthesized protein passes through.

It should be stressed that glycosylation is a contingent cellular process, resulting from non-template-directed, secondary gene events. Different types of glycosylation exist and the two most frequently occurring are N-glycosylation and O-glycosylation (Figure 2). A particular feature of N-glycosylation is that initiation requires a consensus sequon (Asn-Xxx-Thr/Ser; Xxx≠Pro). If this motif occurs a preassembled 14-sugar precursor (Figure 2) can be transferred to Asn while the protein is still being synthesized in the ER. Upon completion of protein translation in the ER, the three terminal glucose units are removed by glucosidases and the resulting oligomannose glycan can be trimmed by mannosidases before the glycoprotein is passed onto the Golgi. This trimming is not always quantitative and thus gives rise to structural heterogeneity [ranging from, e.g., Man_3 (20) to Man_9Glc sugars (21)]. In the Golgi several glycosyltransferases are involved in the elongation, and the nature of the resulting glycotope largely depends on the protein migration rate, the availability of donor/acceptor substrate and appropriate enzymes, environmental factors and, last but not least, the preceding enzymatic reaction (22, 23). It is also in the Golgi where the other most abundant type of glycosylation, i.e., O-glycosylation, is initiated. In contrast to N-glycosylation, O-glycosylation builds from the initial addition of an α-linked GalNAc to Ser or Thr and no consensus sequon is known for the initial addition, making this type of glycosylation less predictable. This feature is enhanced through the existence of at least 21 polypeptide-*N*-acetylgalactosaminetransferases (ppGalNAcT-1 to -21) (24) that are encoded by different genes, differ in their amino acid sequences, and catalyze the same reaction. It has be shown that these ppGalNAcTs act in either stand-alone or concerted reaction

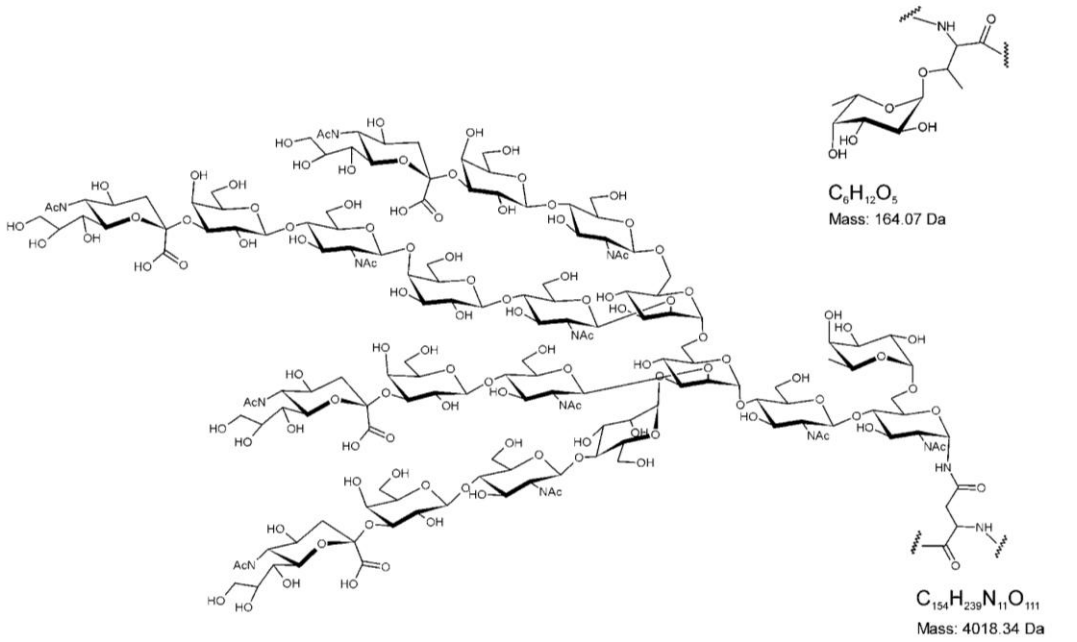


Figure 1 Glycosylation as post-translational modification ranges from a single fucose residue (top) to an *N*-acetyllactosamine repeat containing tetrasialylated tetraantennary core-fucosylated complex type N-glycan (bottom).

modes (25), which suggests that the O-glycosylation is not random at all and that its fine-tuning is not yet understood. Similarly, the other glycosylation steps can be coordinated by several homologous glycosyltransferases, some of which are truly tissue-specific, generating the inherent structural heterogeneity of glycoproteins. The way in which glycans are synthesized, added to the lack of proof-reading mechanisms, has long obscured the understanding of the crucial role played by these entities in the biological phenomena indicated above.

To further unravel the functions that particular glycotopes exert, it is of the utmost importance that chemically well-defined glycans are available, not only as free entities but also as particular glycoconjugates, so that the combined effect of carbohydrate epitope and the underlying structural entity can be evaluated. Much effort has been put into the development of different chemical and enzymatic strategies to generate such glycotope targets in sufficient amounts for structural and functional studies. Unlike DNA, proteins or peptides, where PCR amplification, recombinant DNA technology, and Merrifield solid-phase peptide synthesis (SPPS), respectively, allow for amplification and/or efficient production, carbohydrate synthesis is a technology still under development, relying on non-routine, chemically non-straightforward protocols. Several attempts are being made, particularly by the Seeberger group (26–28), to develop automated synthetic procedures, but despite encouraging results they have not yet reached the universal applicability of their peptide counterpart. In general, glycan synthetic strategies, either linear (glycan incorporated as a preformed glycosyl amino acid building block during peptide synthesis) or convergent (glycosylation carried out on a full-length peptide end product), tend to be as diverse as the structures they target, and to rely on purely chemical, purely enzymatic approaches, or on combinations of both.

In this review, an update of the field is provided in the different areas of glycopeptide and neoglycopeptide synthesis, focusing primarily on chemistries that yield the sugar-amino acid moiety. For simplicity the subject matter is divided between natural and non-natural linkages between the glycan and the non-glycan moieties, and subclassifications are made according to the type of linkage or the chemistry applied during the synthesis. Non-glycosidic linkages between sugars are not included in this review. Further recommended readings in this area can be found elsewhere (28–31).

Natural glycosidic linkages

N-linked glycosylation

Evidently, the optimal mimic of a particular glycoprotein part would be the peptide sequence containing the natural linkage to the peptide-bound monosaccharide and the subsequent glycan epitope. The most frequently occurring glycan types have been addressed by different groups. Thus, Crich et al. (32) employed *N*-benzyloxycarbonyl-L-aspartic cyclic monothioanhydrides and unprotected glycosyl primary amines and *N*-sulfonyl amino acid derivatives in a one-pot synthetic

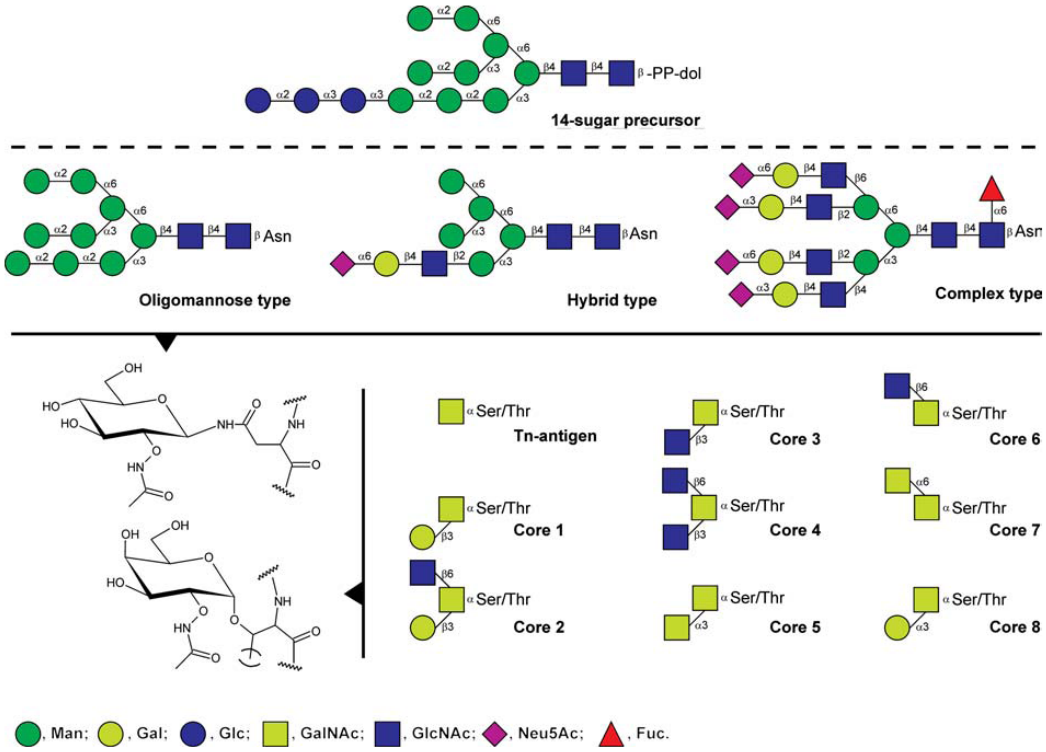


Figure 2 Short-hand notation for the two most frequently occurring types of glycosylation.
The top panel depicts N-type glycans, with the 14-residue dolichol pyrophosphate precursor above the dashed line. Below the dashed line are depicted oligomannose type (left), hybrid type (center), and complex type glycans (right). On the bottom panel on the lower-right, the eight known mucin-type O-glycan core structures are depicted. The symbols are from the Consortium for Functional Glycomics (CFG; http://www.functionalglycomics.org/static/index.shtml).

strategy to generate *N*-glycosyl asparagine derivatives. With the use of sulfonamide-functionalized amino acids the glycan-containing building block could be further elongated to a glycopeptide following a linear strategy. Unfortunately, when the commonly occurring monosaccharide in N-linked glycans, i.e., 2-acetamido-2-deoxyglucose, is a target to be used, the 2-acetamido-2-deoxy-1-glucosamine derivative was found to be a poor nucleophile and limits the yield of this strategy. Also using a glycosylamine but in a convergent strategy, De Bona et al. (33) coupled 6-amino-6-deoxy trehalose to an Aβ pentapeptide in an attempt to produce non-self-aggregating Aβ(1–42) fibrillogenesis inhibitors. In this case the primary hydroxyl functionality was selectively tosylated, then converted to an azido group that upon reduction yielded a primary amine that could be coupled to the side chain carboxyls of Asp/Glu or succinyl-functionalized pentapeptides. In an attempt to combine classical carbohydrate and solid phase peptide synthesis, Swarts et al. (34) generated from glucosamine a fully protected azido-functionalized di- or trisaccharide that was subsequently conjugated to an Asp residue. This element could in turn be incorporated in a conventional SPPS strategy to generate a glycopeptide. The major contribution of this approach resided in the use of a 2-chlorotrityl resin that allowed preservation of acid-labile glycosidic linkages (e.g., the core fucose in N-glycans) upon cleavage from the solid support. Recently, Hu et al. (35) confirmed the utility of this approach in the generation of macrocyclic glycopeptide antibiotics that contained mono- and disaccharides. This synthetic strategy, however, could face limitations when more complex glycans are targeted. Piontek and colleagues worked around this difficulty by making use of unprotected naturally occurring complex type sugars (36). From egg yolk they isolated an Asn-containing biantennary building block of 9–11 monosaccharides that was desialylated and coupled to a functionalized PEGA resin using 1-benzotriazolyloxy-tris(pyrrolidino) phosphonium in the presence of *N*,*N*-diisopropylethylamine. All free amino and hydroxyl groups were subsequently acetylated to allow elongation of the peptide sequence. Eventually, the glycopeptide and the remaining peptide moiety were put together by native chemical ligation to yield for the first time a synthetic, homogeneously glycosylated glycoprotein of more than 100 amino acids with the expected enzymatic activity (37). Nagorny et al. have embarked on the synthesis of similarly glycosylated glycoproteins building from their expertise in complex glycan synthesis. Their targets include the

β-subunit of follicle stimulating hormone (38) and EPO (39, 40), the latter a glycoprotein of 165 amino acids and four (three N-linked and one O-linked) carbohydrate chains that must be clearly regarded as a major synthetic achievement. Their strategy was based on three fragments (residues 1–28, 29–77, and 78–166), each with one N-glycan and the latter also with the O-glycan at position 126, to be combined by sequential ligation. For glycopeptide generation their strategy relied on the direct attachment of a glycosylamine via Lansbury aspartylation. The outcome of this process is significantly dictated by steric factors, and for the (1–28) fragment glycosylation was only successful (i.e., yields over 50%) when small sugars (disaccharides) were employed. To achieve glycosylation with a dodecasaccharide, smaller peptide sequences and more fragment ligations had to be employed, decreasing the overall yields of the strategy. Despite the fact that the natively occurring glycosylation on EPO [i.e., ultra-complex polysialylated tetraantennary structures (2)] has not been attempted thus far, the Lansbury aspartylation can be employed to generate dense glycoclusters as demonstrated by Krauss et al. (41). No doubt this approach represents a significant step forward in understanding and harnessing the role of single glycoforms. Its future success will largely depend on the availability of C1-amine-functionalized complex type glycans.

O-linked glycosylation

This second most frequently occurring type of glycosylation includes the attachment of Gal to Lys, hydroxy-Pro, and Thr; Glc to Tyr and Thr; GlcNAc to hydroxy-Pro, Ser, and Thr; Ara to hydroxy-Pro; Man to hydroxy-Pro, Thr, and Ser; Fuc to Ser and Thr. This section, however, will focus on the most abundant, so-called mucin-type, glycosylation where a GalNAc is linked to Ser or Thr. This type of glycosylation is produced by many epithelial tissues in vertebrates and serves a variety of functions including protection of the underlying tissue or antifreezing properties by means of T-antigen containing tripeptide repeats [(Ala-Ala-Thr)$_n$; n up to 50]. To prepare the latter, Heggemann et al. (42) have converted tri-*O*-acetyl galactal into a 2-azido-1-bromoderivative which by conjugation to Fmoc-Thr and acetamidation of the azido group renders the glycoaminoacid as an α/β epimer mixture in an overall ~50% yield. This building block can then be incorporated onto any SPPS-generated sequence. In that particular study, the authors were able to produce up to five glycosylated repeats and to demonstrate their effects on the retardation of ice-structuring activity. To generate more complex mucin-type glycosylations, chemoenzymatic approaches (*vide infra*) (43) have been employed on the Tn-containing peptide. Alternatively, the Ser/Thr can be decorated with a core-1 (Gal-β-1,3-GalNAc-α-1-) disaccharide, rather than a monosaccharide, and used as an effective building block in SPPS (44). Baumann et al. used this approach to produce, first, the disaccharide-containing Thr that was subsequently conjugated to a fully protected trichloroacetimidate functionalized tetrasaccharide to produce the sialyl Lewis^x (sLe^x)-containing hexasaccharide amino acid. This building block was subsequently included in the SPPS of the N-terminal 15 amino acid peptide of PSGL-1 (45) with double coupling reactions for the two amino acids following the glycan. Overall, the yield was ~12–24% but multi-mg amounts could be produced. In a follow-up study, the same group included both enzymatic and chemical tyrosine sulfation (46) to render the true ligand for interaction with P- and E-selectin. Vohra et al. synthesized the same glycan structure on Thr following an alternative approach (47). They also produced the sLe^x-tetrasaccharide and conjugated it ''en bloc'' to the Thr-containing T antigen. To generate the latter, galactosyl trichloroacetimidate was conjugated to a 2-azido-thiogalactosyl acceptor, activated with diphenyl sulfoxide and triflic anhydride in the presence of 2,6-di-t-butyl-4-methylpyridine and threonine to yield the Thr-disaccharide in an overall yield of 61%. A different strategy to produce glycosylated amino acids, reported by Okamoto et al. (48), made use of a sialyl Tn-Thr derivative that was incorporated to a MUC4 sequence by standard SPPS protocols. The disaccharide moiety was itself prepared by coupling of C1-trimethyl- silylethyl (SE)-functionalized 2-azido galactose to a protected Neu5Ac-1-amide-2-phosphite derivative acting as sialyl donor, with catalysis by trimethylsilyl triflate. Conversion of the SE moiety into the trichloroacetamidate and subsequent conjugation to a Ser benzyl ester were very effective (~70% overall yield) rendering this approach very attractive.

Amino acids other than serine and threonine are also targets for O-glycosylation. In an effort to deliver peptide-based vaccines more efficiently into the cytoplasm of antigen-presenting cells via the mannose-binding receptor, Lee et al. (49) incorporated a mannose residue to Fmoc-hydroxyproline (Hyp) allyl ester using perbenzoylated 1-bromomannose as mannosyl donor. Treatment with $Pd^0(Ph_3P)_4$ yielded the fully protected glycoamino acid ready for SPPS. Biondi et al. (50) described conjugation of per-O-acetylated glucose to both Hyp and Tyr residues in fairly high yields (~75%).

Glycoconjugate dendrimers

In an attempt to develop tumor-associated carbohydrate antigens with multivalent display and thus enhanced presentation of the underlying peptide aglycon to the immune system, Vichier-Guerre et al. (51) used Koenigs-Knorr condensation, as Heggemann et al. (42), but with an α-1-chloro (instead of 1-bromo) derivative of *N*-acetylgalactosamine and silver carbonate/perchlorate to favor α-anomeric configuration in the end product, which was then incorporated to an Fmoc-protected homoserine acceptor, prepared according to Shiori et al. (52). Other attempts to mimic naturally occurring, dense glycocalyx structures have involved glycoclusters made by coupling trihydroxy amine compounds to trichloracetimidate-functionalized sugars, or trivalent carboxylic acids with aminoethyl-functionalized carbohydrates (53). The Danishefsky group has also made important contributions to this field. In a recent report on the synthesis of a fucosylated GM1 epitope, the pentenyl glycoside of a fully protected hexasaccharide was conjugated to Fmoc-allylglycine benzyl ester by olefin cross-metathesis (OCM) using the Hoveyda-

Grubbs catalyst, yielding after catalytic hydrogenation a five-carbon homolog of the Ser O-glycan (54). They further exploited this approach to build a pentavalent peptide platform displaying the Globo-H, GM2, STn, TF, and Tn carbohydrate epitopes on adjacent amino acids (55). Conjugated to Keyhole Limpet Hemocyanin (KLH) and given to mice, this unique structure elicited antibodies that recognized each of the individual epitopes, thus paving the way for polyvalent carbohydrate vaccines. Jiménez-Barbero et al. (56) parted from a 4,6-silylidene-protected galactal that reacted as dienophile diastereospecifically with a thione-functionalized homoglutamate heterodiene to yield the Gal-β-1,3-Gal-α-1 mimetic which was capable of binding viscumin, a galactose-binding plant lectin. Both synthetic N- and O-linked glycosylation reactions are depicted in Figure 3.

Chemoenzymatic synthetic strategies

Strategies based on incorporating preformed glycosyl amino acids as building blocks for SPPS endure the potential drawbacks of low-yield couplings inherent to large molecular entities, and of cleavage of labile glycosidic linkages under the conditions required for peptide deprotection and cleavage from the resin. To address these problems, combined strategies involving enzymatic elaboration of sugar chains on chemically synthesized polypeptides have been devised. Different enzymes allow glycan attachment to a peptide chain in aqueous solution and with no need of protecting groups. These procedures usually start from a monosaccharide-tagged polypeptide, previously made by chemical means (*vide supra*), or from the products of endoglycosidase treatment of natural glycoproteins. Further elongation of this glycan can then be done in two ways. On one hand, glycosyltransferases can be used to extend sugar chains one monosaccharide at a time. Thus, Ueda et al. synthesized glycosylated glucagon-like peptide 1 analogs from a GlcNAc-bearing precursor peptide made by SPPS and subsequently elongated with β-1,4-galactosyltransferase (β-1,4-GalT) and α-2,6-sialyltransferase (α-2,6-SiaT) (57). Similarly, Gutiérrez-Gallego et al. used this strategy to generate the sialyl Lewis X epitope on a core-2 mucin-type glycopeptide (43). Endo-β-*N*-acetylglucosaminidases (ENGases), on the other hand, are advantageous over common glycosyltransferases in that they can attach in a single step an oligosaccharide to a GlcNAc-containing polypeptide. ENGases hydrolyze β-1,4-glycosidic bonds of N-glycans in glycoproteins and transfer the released oligosaccharyl moiety to another glycosyl polypeptide acceptor. The most used ENGases are fungal Endo-M, which can act on the three major types of N-glycans, and bacterial Endo-A, specific toward high mannose type N-glycans. Transglycosylation reactions, however, tend to suffer from low efficiencies (5–20%), and the enzyme requirement for natural N-glycans as substrate donors limits their applicability. To address these problems, Wang explored sugar oxazolines as substrate donors. Di- and tetrasaccharide oxazolines corresponding to N-glycans cores were synthesized and transglycosylated via Endo-A to a GlcNAc-heptapeptide derived from HIV-1 gp120. The reaction was run under mild aqueous conditions and efficiency was improved from 5–20% to 82%, with the resulting glycopeptides proving resistant to hydrolysis (58). Li et al. made a 47-mer peptide with two N-linked pentasaccharides also by this chemoenzymatic approach. The polypeptide, carrying GlcNAc-Asn residues, was first assembled by SPPS and used to transglycosylate N-linked tetrasaccharide oxazolines by means of Endo-A in an excellent 86% yield (59). Zeng et al. studied the structural donor substrate requirements for Endo-A in more detail and established the Man-β-1,4-GlcNAc-oxazoline as the minimum structure recognized by the enzyme. Whereas changes at the C6 did not alter enzyme activity, configuration inversion at the 2′ and/or 4′-hydroxyls of the Man residue caused total loss of activity (60). In another study, Huang et al. showed that native N-linked and triazole-linked GlcNAc were tolerated by Endo-A as acceptor entities (61). Another report by Rising et al. (62) showed excellent (91%) yields for the transglycosylation of Man-α-1,3-Glc-β-1,4-GlcNAc-oxazoline using Endo-M. In the search for new glycoengineering tools, Huang et al. generated and studied several Endo-M and Endo-A mutants to select an enzyme retaining the ability to form glycosidic bonds but devoid of hydrolytic activity. They identified the Endo-M-N175A and Endo-A-N171A mutants as glycosynthases with high transglycosylation rates (around 70% for large N-glycans) and completely non-hydrolytic (63). Matsushita et al. (64) also employed a chemoenzymatic strategy to synthesize a MUC1-related glycoprotein bearing both N- and O-linked glycans. First, per-O/N-acetylated glycosyl amino acid building blocks [GlcNAc-Asn and GalNAc-β-1,6-(Gal-β-1,3-)GalNAc-Thr] were incorporated on short peptides by microwave-assisted SPPS. The more complex step was then achieved through transglycosylation of Neu5Ac-α-2,3-Gal-β-1,4-GlcNAc-β-1,6-Man-α-1,6-(Neu5Ac-α-2,3-Gal-β-1,4-GlcNAc-β-1,6-Man-α-1,3-)Man-β-1,4-GlcNAc to the GlcNAc-Asn residue using Endo-M, whereas the core 2 epitope was elongated to a sialyl hexasaccharide by sequential action of different glycosyltransferases (β-1,4-GalT, α-2,3-SiaT). Finally, complex glycopeptides bearing N and/or O glycans, and LPKTGLR and GG signal sequences, at the *C*- or *N*-termini, respectively, could be assembled employing sortase-A-mediated ligation to yield a complex, multiglycosylated entity.

The last class of enzymes with synthetic applications is oligosaccharyl transferases (OSTs), which transfer oligosaccharide precursor onto Asn residues. One example is the PgIB from *Campylobacter jejuni* used by Glover et al. to glycosylate the peptide acceptor KDFNVSKA with a synthetic disaccharide donor (GalNAc-α-1,3-bacillosamine-pyrophosphate-undecaprenyl) (65).

Non-natural glycosidic linkages

The term neoglycopeptide was coined to define a new class of glycopeptides characterized by having a non-native sugar-peptide linkage formed between two highly reactive functional groups normally not present in natural glycoproteins. In addition to facilitating glycopeptides synthesis, the non-

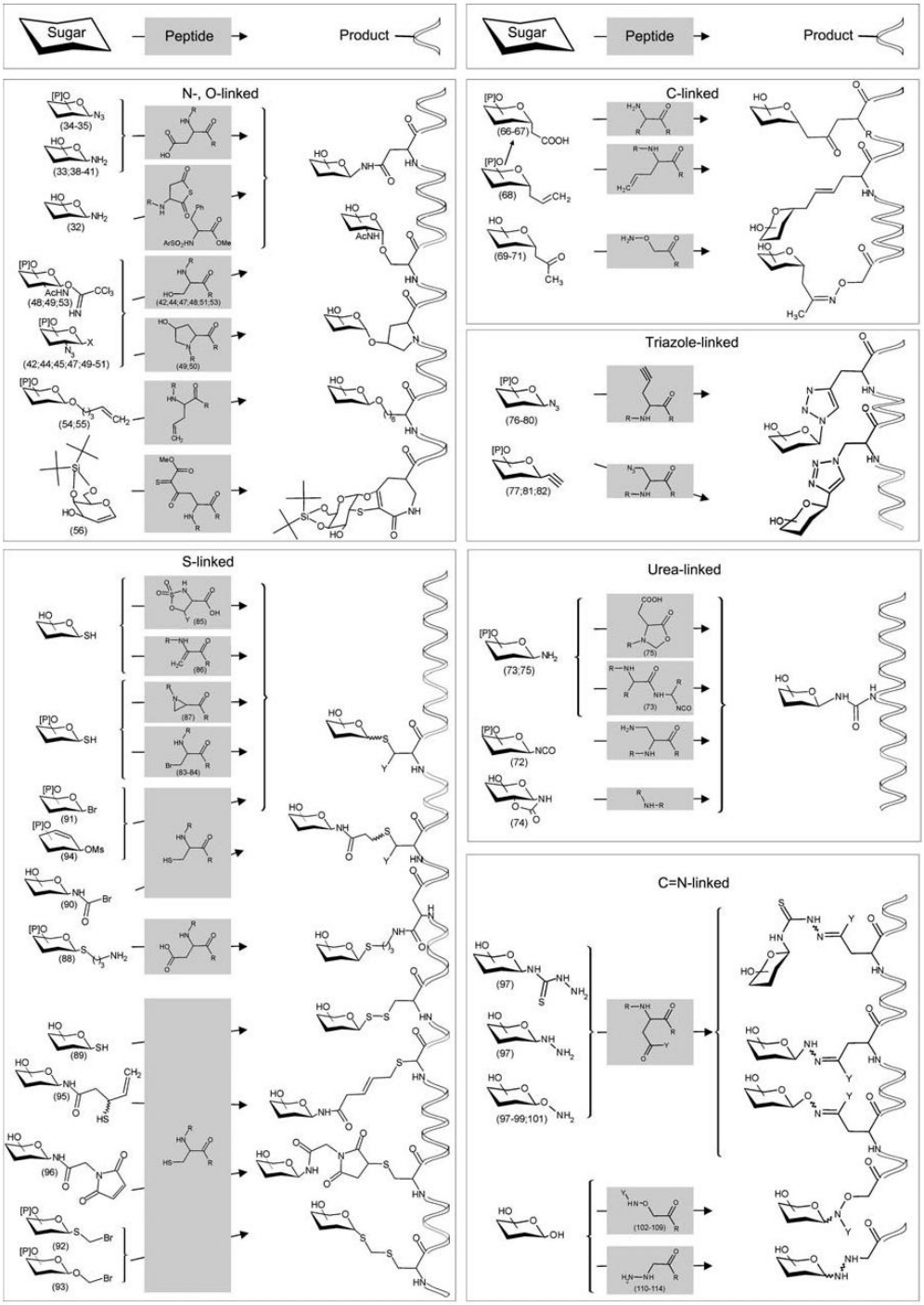


Figure 3 Synthetic chemistries for natural and non-natural glycoconjugate linkages. Chemistries are grouped according to the sugar-peptide linkage; top drawings indicate the direction of the reactions. P, protecting groups for sugar hydroxyls; R, protecting groups in the case of amino acids or remainder of the sequence in the case of peptides; X, bromine, chlorine, or thiophenyl; Y, proton or methyl. Numbers between parentheses correspond to references.

native sugar-peptide linkage can impart stability toward proteases or other adverse conditions, e.g., pH, temperature.

Among the several types of non-natural linkages between sugar and peptide, C-linked glycopeptides have been successfully synthesized and tested under different conditions (Figure 3). Thus, C-fucosyl-acetic acid building blocks were incorporated to a dendrimer library by SPPS and tested as inhibitors of biofilm formation by the pathogen *Pseudomonas aeruginosa* (66, 67). OCM was employed to incorporate C-allyl glycosides into neoglycopeptides by either linear or convergent strategies (68). Another route to C-linked glycopeptides, chemoselective ligation, has the advantage of being done in aqueous media and not requiring auxiliary coupling agents nor protecting groups on either peptide or saccharide. By this approach, C-glycopyranosyl-ketones derived from allyl-glycosides were conjugated to a peptide containing an aminooxy group on the N-terminus or the side chain to give an anticancer vaccine successfully tested against mouse mammary carcinoma (69–71).

Urea-linked glycopeptides, another class of neoglycopeptides, are more water-soluble and enzyme-stable than natural glycopeptides and thus serve as ideal analogs of peptide drugs or enzyme inhibitors. Urea-tethered glycopeptides have been prepared either in a linear strategy, using glycosyl amino acids formed by reaction between *N*-acetyl-D-glucosaminyl isocyanate and amino acid derivates (72), or in a convergent strategy by conjugation of glycosyl amines and peptidyl isocyanates in solution (73). Another approach to urea-linked glycosylated amino acids relied on the reaction between glucopyranosyl oxazolidinone and Lys ε-amino group (74), or vice versa, between glycosyl amines and Fmoc-protected Asp/Glu-5-oxazolidone (75) (Figure 3).

Another type of linkage found in neoglycopeptides is the triazole ring, formed via copper(I)-mediated 1,3-dipolar cycloaddition (Huisgen cyclization, also known as click chemistry) between an alkyne and an azide group (Figure 3). This approach is of particular interest because of the orthogonality of azide and alkyne with other functional groups, and the compatibility with aqueous medium. Several neoglycopeptides have been made from glycosyl azides and alkyne-containing peptides. The alkyne group is usually incorporated to the peptide as a non-natural amino acid (e.g., propargylglycine) (76–78); alternatively, the alkyne functionality has been generated by reaction of a Cys residue with 2-bromoacetyl propargylamide (79), or of a Lys with N-succinimidyl-4-pentynoate (80). Optimized conditions were devised for the reaction between GalNAc-azide and propargyl-containing unprotected peptides; subsequently, two of the resulting triazole-linked GalNAc-neoglycopeptides were joined by native chemical ligation, demonstrating the compatibility between the two chemistries (76). In combination with chemoenzymatic methods, click chemistry has allowed to make carbohydrate-modified cyclic peptides that could find use as antibiotics. Thus, Lin and Walsh used a linear peptide containing propargylglycine that was first cyclized by tyrocidine synthetase and afterwards coupled to 21 different azido monosaccharides (78). To synthesize more complex neoglycopeptides, an azido ''handle'' displaying either STn antigen or clustered Tn was synthesized and conjugated to an alkynyl 10-mer polypeptide. Best yields were obtained when the Cu-catalyzed Huisgen cyclization was carried out in phosphate buffered saline at pH 7.2 (80). Click chemistry can also be performed with inverted functionalities (acetylenic glycosides and azide-containing amino acids), reportedly with slightly lower yields over the original type (77). Nevertheless, GM2-derived anticancer vaccines were efficiently synthesized by ligation between a propargylated GM2 and azide-containing peptide (81) with rates improved from hours to minutes by the use of microwave irradiation (82).

Another category of chemoselective ligation is characterized by the reaction between thiol groups and a variety of electrophiles. As depicted in Figure 3, S-linked glycopeptides have been synthesized mainly by two different strategies. In the first approach, thiohexoses have been reacted with peptides bearing electrophilic moieties such as cyclic sulfamidates, dehydroalanine (Dha), aziridine or halogenated residues (83, 84). For instance, S-linked glycopeptides were made by reaction of Ser or Thr-derived sulfamidates and various unprotected 1-thio sugars, either in solution or on solid support (85). It should be noted that this strategy is limited to the incorporation of mono-, not di- or polysaccharides, at exclusively the N-terminus of the peptide, as sulfamidates do not stand the basic conditions of Fmoc deprotection, and glycosidic linkages are labile to the protic or Lewis acids required to remove the N-sulfate. Another type of chemoselective ligation, the coupling of thiohexoses and Dha-containing peptides was used by Galonic et al. to conjugate four different tumor-associated carbohydrate antigens (Tn, T, STn, and 2,6-ST) to tripeptides. The reaction generated a pair of diastereoisomers at the newly formed Cys α-carbon, but the configuration of the anomeric center was fully retained (86). The same group later used an aziridine 2-carboxylic acid (Azy)-containing peptide and explored conditions for its conjugation with α-SH-GalNAc (87). Halogenated amino acid residues such as β-bromoalanine also allow conjugation of thiosugars, generating S-linked glycosyl amino acids for solid phase or solution-based strategies (83, 84). Thiohexoses have also been incorporated to proteinogenic amino acid residues such as Asp (88) or Cys (89). Several reports have also appeared of the reverse mode of S-linked glycopeptides synthesis, involving thiol-containing peptides and sugars modified at the anomeric carbon as bromides or thiomethyl bromides (90–93), mesylates (94), thiomethylazides (90–92), allylic thiols (95), or maleimides (96).

The last group of chemoselective reactions relies on C=N linkages, such as oximes, hydrazones, and thiosemicarbazones, mimicking the natural structural motifs of N- and O-linked glycopeptides (Figure 3). One approach consists of functionalizing the reducing end of the sugar moiety as a nucleophile (aminooxy, hydrazine, thiosemicarbazide) and reacting it with a carbonyl group in a synthetic peptide (97). In this way, aminooxy carbohydrates have been prepared by reaction of *N*-hydroxysuccimide with glycosyl chlorides (98) or *N*-hydroxyphthalimide with fluoride-activated sugars (99). The electrophilic carbonyl on the peptide is generated either by ozonolysis of dehydroleucine (98) or by periodate oxi-

dation of Ser (100). Oxime ligation between these aminooxy and carbonyl components has allowed the synthesis of both linear and multivalent neoglycopeptides. Although usually carried out in solution, Renaudet and Dumy described a combined oxime ligation/solid phase approach to RAFT-derived neoglycopeptides that allowed recovery of the excess of aminooxy carbohydrate (101). These examples aside, the absence of either aldehydes or ketone groups on the side chains of natural amino acids has caused this type of chemoselective ligations to be usually performed in the reverse mode, i.e., with the aminooxy group on the peptide, either at the *N*-terminus (102) or on the side chain (103). One application of this methodology was simply analytical, as it was found that ligation of complex glycans to basic aminooxy peptides enhanced the sensitivity of their mass spectrometric detection compared to underivatized versions (102). Several neoglycopeptides prepared by this approach have been used as glycoprobes to characterize carbohydrate-lectin interactions. Thus, photoreactive glycoprobes bearing a biotin moiety were synthesized to detect lectins by chemiluminescence (104). Other biotinylated glycoprobes with two post-translational modifications (glycosylation and sulfation) were prepared by two orthogonal chemoselective ligations to study different glycosyl ligands of P-selectin (103). To study carbohydrate-lectin interactions without the requirement of labeling or high sample amounts, an oxime-linked glycopeptide was synthesized and used as immobilized ligand in surface plasmon resonance (SPR) studies (105). In this type of oxime-linked neoglycopeptides the monosaccharide closed to the peptide scaffold is expected to exist in a variety of structural forms (cylic α/β and acylic E/Z isomers), the latter (i.e., open forms) not recognizable by lectins. To overcome this problem, a variation of chemoselective ligation making use of N′-methyl derivatives of the aminooxy peptides, which ensure a closed conformation in the sugar close to the peptide backbone, has been developed. Peptides bearing this N′-methyl-aminooxy at the N-terminus (106) or side chain (106–108) were synthesized and conjugated with several different monosaccharides. A comparative study on the functionality of both types of neoglycopeptides (with versus without N′-methylation) in carbohydrate-lectin interaction studies by SPR was done; both nuclear magnetic resonance and mass spectrometry data demonstrated that for the non-methylated aminooxy function a rather unfavorable 70:30 distribution of open versus closed structures existed. SPR interaction studies with several lectins showed higher responses of the N′-methylated glycopeptides, indicative of a single binding event and in contrast with the behavior of non-methylated analogs (109). Hydrazide is another nucleophile employed to conjugate carbohydrates to peptides, with a strong α-effect that preserves a cyclic conformation of the sugar. Oligosaccharyl transferase inhibitors have been synthesized by this ligation strategy, as well as by oxime chemistry (110). In a comparison with their native-like counterparts, neoglycopeptides showed equal or better binding affinity to the enzyme (111). Hydrazine-linked glycopeptides were also used as probes in photoaffinity labeling studies (112). Possibly owing to their lower stability (compared with oxime-linked) at pH 4–6 (113), hydrazide-derived neoglycopeptides have found their main use as an enrichment handle and signal-enhancing method in N-glycan mass spectrometric analysis (114).

Outlook

The recognition that carbohydrates, despite their structural heterogeneity and non-template driven synthesis, are crucial for many biological processes has brought about considerable efforts both to analyze and describe naturally occurring glycosylation on proteins and to synthesize these glycans, both with and without aglycon, in chemically well-defined form. This, in turn, has made possible the unraveling of additional biological mechanisms where glycans play a crucial part (115–119). Recent examples include the study of the sialylation of Von Willebrand factor (VWF) (115). It was demonstrated that platelet tethering is induced by multimerization of VWF, in turn regulated in plasma by ADAMTS13. Desialylation of VWF enhanced susceptibility to ADAMTS13 proteolysis (but protected against proteolysis by serine and cysteine proteases) indicating a potential role of the absolute quantity of sialic acid on VWF in the platelet tethering. In a study on how cancer cells bind to vascular surfaces and extravasate into target organs Barthel et al. (116) described the upregulation of 3 fucosyltransferases (FT3, 6, and 7) that mediate the synthesis of α-1,3 linked fucose, in liver-metastatic PCa cells. Several membrane proteins were shown to contain the E-selectin binding determinants that subsequently bound firmly to bone marrow endothelium demonstrating that the single addition of a fucose residue dictates the fate of migrating cancer cells. A final example relates to the common flu where host glycans (principally sialic acid, SA) play an important role in the anchoring of viral hemagglutinin (HA) glycoprotein (119). Partial deglycosylation of HA increased binding to SA that was accompanied by a reduced specificity. Furthermore, the authors showed that removal of structurally irrelevant HA glycans elicited better antibody response with higher binding affinities and improved neutralization activity paving the way for next generation vaccines.

Being aware of the importance of glycosylation in nature, enormous efforts have been brought about to synthesize neoglycoconjugates mimicking the natural glycosylation but with improved biological properties (increased stability and solubility), what makes them more suitable for drug development. Thus, Cipolla et al. synthesized and tested a non-hydrolyzable vaccine against Tn-expressing carcinomas by substituting typical O-glycosidic linkage with a non-hydrolyzable C-linkage (71). Other groups centered the focus of their attention in improving the solubility in aqueous media and bioavailability and synthesized urea-linked analogs as potent enzyme inhibitors and peptide drugs analogs (73). Other stable neoglycopeptides, such as S-linked or triazole-linked glycoconjugates, were prepared and tested as analogs of the natural antibiotic tyrocidine. Whereas S-linked analogs achieved a two-fold improvement of the therapeutic index (84) in comparison with the natural component, similar tri-

azole-linked analogs showed more than six-fold increase in therapeutic index (78). Triazole-linked analogs were also considered as an alternative approach to obtain GM2 glycoconjugate vaccines, avoiding the limitations derived from the traditional hemisynthetic strategy (81). Other neoglycopeptides, e.g., S-linked glycoconjugates containing tumor-associated carbohydrate antigens, were synthesized and considered also as potential drugs in cancer vaccine therapies (86). However, well-characterized structural entities do not always lead to straightforward findings. An example at hand is the use by Joyce et al. of mono-, di-, and trivalent $Man_9GlcNAc_2$ containing cyclopeptides by Lansburry aspartylation (120). These glycopeptides had been designed to serve as HIV-1 vaccines based on the specificity of the human 2G12 mAb. Although high titers of carbohydrate-specific antisera capable of competing with 2G12 were generated, they could not neutralize viral isolates nor bind to the glycopeptide immunogen, suggesting that nature's fine-tuning of biomolecular recognition events is still not completely within our grasp. In this respect, future endeavors will certainly have to take a closer look to the non-carbohydrate part of the glycoprobes, as recent examples have shown that the specificity of the molecular communication resides in an adequate combination of carbohydrate and peptide structural information (121).

Acknowledgements

The authors would like to acknowledge the financial support received from the European Union (HEALTH-F3-2008-223414), Ministerio de Ciencia e Innovación (BIO2008-04487-CO3-02 and BIO2009-08983) and Generalitat de Catalunya (2009 SGR 492).

References

1. Borisov OV, Field M, Ling VT, Harris RJ. Characterization of oligosaccharides in recombinant tissue plasminogen activator produced in Chinese hamster ovary cells: two decades of analytical technology development. Anal Chem 2009; 81: 9744–54.
2. Llop E, Gutiérrez-Gallego R, Belalcazar V, Gerwig GJ, Kamerling JP, Segura J, Pascual JA. Evaluation of protein N-glycosylation in 2-DE: erythropoietin as a study case. Proteomics 2007; 7: 4278–91.
3. Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J. Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. Electrophoresis 2009; 30: 2957–66.
4. Krishnamoorthy L, Mahal LK. Glycomic analysis: an array of technologies. ACS Chem Biol 2009; 4: 715–32.
5. 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; 19: 498–506.
6. Varki A, Freeze HH, Manzi AE. Overview of glycoconjugate analysis. Curr Protoc Protein Sci 2009; Chapter 12: Unit 8.
7. Messner P. Prokaryotic protein glycosylation is rapidly expanding from ''curiosity'' to ''ubiquity''. Chembiochem 2009; 10: 2151–4.
8. Abu-Qarn M, Eichler J, Sharon N. Not just for eukarya anymore: protein glycosylation in bacteria and archaea. Curr Opin Struct Biol 2008; 18: 544–50.
9. Gupta SK, Bansal P, Ganguly A, Bhandari B, Chakrabarti K. Human zona pellucida glycoproteins: functional relevance during fertilization. J Reprod Immunol 2009; 83: 50–5.
10. Velasquez JG, Canovas S, Barajas P, Marcos J, Jiménez-Movilla M, Gutiérrez-Gallego R, Ballesta J, Avilés M, Coy P. Role of sialic acid in bovine sperm-zona pellucida binding. Mol Reprod Dev 2007; 74: 617–28.
11. Haltiwanger RS, Lowe JB. Role of glycosylation in development. Annu Rev Biochem 2004; 73: 491–537.
12. Zhao YY, Takahashi M, Gu JG, Miyoshi E, Matsumoto A, Kitazume S, Taniguchi N. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. Cancer Sci 2008; 99: 1304–10.
13. Marth JD, Grewal PK. Mammalian glycosylation in immunity. Nat Rev Immunol 2008; 8: 874–87.
14. van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. Nat Immunol 2008; 9: 593–601.
15. Schauer R. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol 2009; 19: 507–14.
16. Rek A, Krenn E, Kungl AJ. Therapeutically targeting protein-glycan interactions. Br J Pharmacol 2009; 157: 686–94.
17. Luther KB, Haltiwanger RS. Role of unusual O-glycans in intercellular signaling. Int J Biochem Cell Biol 2009; 41: 1011–24.
18. Stamatatos L, Morris L, Burton DR, Mascola JR. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? Nat Med 2009; 15: 866–70.
19. Reading PC, Tate MD, Pickett DL, Brooks AG. Glycosylation as a target for recognition of influenza viruses by the innate immune system. Adv Exp Med Biol 2007; 598: 279–92.
20. Li JS, Vavricka CJ, Christensen BM, Li J. Proteomic analysis of N-glycosylation in mosquito dopachrome conversion enzyme. Proteomics 2007; 7: 2557–69.
21. Di Patrizi L, Capone A, Focarelli R, Rosati F, Gutiérrez-Gallego R, Gerwig GJ, Vliegenthart JK. Structural characterization of the N-glycans of gp273, the ligand for sperm-egg interaction in the mollusc bivalve Unio elongatulus. Glycoconj J 2001; 18: 511–8.
22. Paulson JC, Colley KJ. Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J Biol Chem 1989; 264: 17615–8.
23. de Graffenried CL, Bertozzi CR. The roles of enzyme localisation and complex formation in glycan assembly within the Golgi apparatus. Curr Opin Cell Biol 2004; 16: 356–63.
24. Tarp MA, Clausen H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. Biochim Biophys Acta 2008; 1780: 546–63.
25. Hanisch FG, Reis CA, Clausen H, Paulsen H. Evidence for glycosylation-dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides. Glycobiology 2001; 11: 731–40.
26. Seeberger PH. Automated carbohydrate synthesis as platform to address fundamental aspects of glycobiology – current status and future challenges. Carbohydr Res 2008; 343: 1889–96.
27. Codee JD, Krock L, Castagner B, Seeberger PH. Automated solid-phase synthesis of protected oligosaccharides containing beta-mannosidic linkages. Chemistry 2008; 14: 3987–94.
28. Seeberger PH. Automated oligosaccharide synthesis. Chem Soc Rev 2008; 37: 19–28.

29. Bernardes GJ, Castagner B, Seeberger PH. Combined approaches to the synthesis and study of glycoproteins. ACS Chem Biol 2009; 4: 703–13.
30. Bennett CS, Wong CH. Chemoenzymatic approaches to glycoprotein synthesis. Chem Soc Rev 2007; 36: 1227–38.
31. Seeberger PH, Werz DB. Synthesis and medical applications of oligosaccharides. Nature 2007; 446: 1046–51.
32. Crich D, Sasaki K, Rahaman MY, Bowers AA. One-pot syntheses of dissymmetric diamides based on the chemistry of cyclic monothioanhydrides. Scope and limitations and application to the synthesis of glycodipeptides. J Org Chem 2009; 74: 3886–93.
33. De Bona P, Laura GM, Caraci F, Copani A, Pignataro B, Attanasio F, Cataldo S, Pappalardo G, Rizzarelli E. Design and synthesis of new trehalose-conjugated pentapeptides as inhibitors of Abeta(1-42) fibrillogenesis and toxicity. J Pept Sci 2009; 15: 220–8.
34. Swarts BM, Chang YC, Hu H, Guo Z. Synthesis and CD structural studies of CD52 peptides and glycopeptides. Carbohydr Res 2008; 343: 2894–902.
35. Hu H, Xue J, Swarts BM, Wang Q, Wu Q, Guo Z. Synthesis and antibacterial activities of N-glycosylated derivatives of tyrocidine A, a macrocyclic peptide antibiotic. J Med Chem 2009; 52: 2052–9.
36. Piontek C, Varon SD, Heinlein C, Pohner C, Mezzato S, Ring P, Martin A, Schmidt FX, Unverzagt C. Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. Angew Chem Int Ed Engl 2009; 48: 1941–5.
37. Davis BG. The linear assembly of a pure glycoenzyme. Angew Chem Int Ed Engl 2009; 48: 4674–8.
38. Nagorny P, Fasching B, Li X, Chen G, Aussedat B, Danishefsky SJ. Toward fully synthetic homogeneous beta-human follicle-stimulating hormone (beta-hFSH) with a biantennary N-linked dodecasaccharide. synthesis of beta-hFSH with chitobiose units at the natural linkage sites. J Am Chem Soc 2009; 131: 5792–9.
39. Tan Z, Shang S, Halkina T, Yuan Y, Danishefsky SJ. Toward homogeneous erythropoietin: non-NCL-based chemical synthesis of the Gln(78)-Arg(166) glycopeptide domain. J Am Chem Soc 2009; 131: 5424–31.
40. Kan C, Trzupek JD, Wu B, Wan Q, Chen G, Tan Z, Yuan Y, Danishefsky SJ. Toward homogeneous erythropoietin: chemical synthesis of the Ala(1)-Gly(28) glycopeptide domain by ''alanine'' ligation. J Am Chem Soc 2009; 131: 5438–43.
41. Krauss IJ, Joyce JG, Finnefrock AC, Song HC, Dudkin VY, Geng X, Warren JD, Chastain M, Shiver JW, Danishefsky SJ. Fully synthetic carbohydrate HIV antigens designed on the logic of the 2G12 antibody. J Am Chem Soc 2007; 129: 11052–4.
42. Heggemann C, Budke C, Schomburg B, Majer Z, Wissbrock M, Koop T, Sewald N. Antifreeze glycopeptide analogues: microwave-enhanced synthesis and functional studies. Amino Acids 2010; 38: 213–22.
43. Gutiérrez-Gallego R, Dudziak G, Kragl U, Wandrey C, Kamerling JP, Vliegenthart JF. Enzymatic synthesis of the core-2 sialyl Lewis X O-glycan on the tumor-associated MUC1a′ peptide. Biochimie 2003; 85: 275–86.
44. Brocke C, Kunz H. Synthetic tumor-associated glycopeptide antigens from the tandem repeat sequence of the epithelial mucin MUC4. Synthesis 2004; 4: 525–42.
45. Baumann K, Kowalczyk D, Kunz H. Total synthesis of the glycopeptide recognition domain of the P-selectin glycoprotein ligand 1. Angew Chem Int Ed Engl 2008; 47: 3445–9.
46. Baumann K, Kowalczyk D, Gutjahr T, Pieczyk M, Jones C, Wild MK, Vestweber D, Kunz H. Sulfated and non-sulfated glycopeptide recognition domains of P-selectin glycoprotein ligand 1 and their binding to P- and E-selectin. Angew Chem Int Ed Engl 2009; 48: 3174–8.
47. Vohra Y, Buskas T, Boons GJ. Rapid assembly of oligosaccharides: a highly convergent strategy for the assembly of a glycosylated amino acid derived from PSGL-1. J Org Chem 2009; 74: 6064–71.
48. Okamoto R, Souma S, Kajihara Y. Efficient synthesis of MUC4 sialylglycopeptide through the new sialylation using 5-acetamido-neuraminamide donors. J Org Chem 2008; 73: 3460–6.
49. Lee DJ, Kowalczyk R, Muir VJ, Rendle PM, Brimble MA. A comparative study of different glycosylation methods for the synthesis of D-mannopyranosides of Nalpha-fluorenylmethoxycarbonyl-trans-4-hydroxy-L-proline allyl ester. Carbohydr Res 2007; 342: 2628–34.
50. Biondi L, Filira F, Giannini E, Gobbo M, Lattanzi R, Negri L, Rocchi R. Novel glycosylated [Lys(7)]-dermorphin analogues: synthesis, biological activity and conformational investigations. J Pept Sci 2007; 13: 179–89.
51. Vichier-Guerre S, Lo-Man R, Huteau V, Deriaud E, Leclerc C, Bay S. Synthesis and immunological evaluation of an antitumor neoglycopeptide vaccine bearing a novel homoserine Tn antigen. Bioorg Med Chem Lett 2004; 14: 3567–70.
52. Shiori T, Irako N, Sakakibara S, Matsuura F, Hamada Y. A new efficient synthesis of nicotianamine and 2′-deoxy mugineic acid. Heterocycles 1997; 44: 519–30.
53. Shaikh HA, Sonnichsen FD, Lindhorst TK. Synthesis of glycocluster peptides. Carbohydr Res 2008; 343: 1665–74.
54. Nagorny P, Kim WH, Wan Q, Lee D, Danishefsky SJ. On the emerging role of chemistry in the fashioning of biologics: synthesis of a bidomainal fucosyl GM1-based vaccine for the treatment of small cell lung cancer. J Org Chem 2009; 74: 5157–62.
55. Zhu J, Wan Q, Lee D, Yang G, Spassova MK, Ouerfelli O, Ragupathi G, Damani P, Livingston PO, Danishefsky SJ. From synthesis to biologics: preclinical data on a chemistry derived anticancer vaccine. J Am Chem Soc 2009; 131: 9298–303.
56. Jiménez-Barbero J, Dragoni E, Venturi C, Nannucci F, Arda A, Fontanella M, Andre S, Canada FJ, Gabius HJ, Nativi C. alpha-O-linked glycopeptide mimetics: synthesis, conformation analysis, and interactions with viscumin, a galactoside-binding model lectin. Chemistry 2009; 15: 10523–31.
57. Ueda T, Tomita K, Notsu Y, Ito T, Fumoto M, Takakura T, Nagatome H, Takimoto A, Mihara S, Togame H, Kawamoto K, Iwasaki T, Asakura K, Oshima T. Chemoenzymatic synthesis of glycosylated glucagon-like peptide 1: effect of glycosylation on proteolytic resistance and in vivo blood glucose-lowering activity. J Am Chem Soc 2009; 131: 6237–45.
58. Wang LX. Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. Carbohydr Res 2008; 343: 1509–22.
59. Li B, Zeng Y, Hauser S, Song H, Wang LX. Highly efficient endoglycosidase-catalyzed synthesis of glycopeptides using oligosaccharide oxazolines as donor substrates. J Am Chem Soc 2005; 127: 9692–3.
60. Zeng Y, Wang J, Li B, Hauser S, Li H, Wang LX. Glycopeptide synthesis through endo-glycosidase-catalyzed oligosaccharide transfer of sugar oxazolines: probing substrate structural requirement. Chemistry 2006; 12: 3355–64.

61. Huang W, Groothuys S, Heredia A, Kuijpers BH, Rutjes FP, van Delft FL, Wang LX. Enzymatic glycosylation of triazole-linked GlcNAc/Glc-peptides: synthesis, stability and anti-HIV activity of triazole-linked HIV-1 gp41 glycopeptide C34 analogues. Chembiochem 2009; 10: 1234–42.
62. Rising TW, Claridge TD, Moir JW, Fairbanks AJ. Endohexosaminidase M: exploring and exploiting enzyme substrate specificity. Chembiochem 2006; 7: 1177–80.
63. Huang W, Li C, Li B, Umekawa M, Yamamoto K, Zhang X, Wang LX. Glycosynthases enable a highly efficient chemoenzymatic synthesis of N-glycoproteins carrying intact natural N-glycans. J Am Chem Soc 2009; 131: 2214–23.
64. Matsushita T, Sadamoto R, Ohyabu N, Nakata H, Fumoto M, Fujitani N, Takegawa Y, Sakamoto T, Kurogochi M, Hinou H, Shimizu H, Ito T, Naruchi K, Togame H, Takemoto H, Kondo H, Nishimura SI. Functional neoglycopeptides: synthesis and characterization of new class MUC1 glycoprotein models having core 2-based O-glycan and complex-type N-glycan chains. Biochemistry 2009; 48: 11117–33.
65. Glover KJ, Weerapana E, Numao S, Imperiali B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. Chem Biol 2005; 12: 1311–5.
66. Kolomiets E, Johansson EM, Renaudet O, Darbre T, Reymond JL. Neoglycopeptide dendrimer libraries as a source of lectin binding ligands. Org Lett 2007; 9: 1465–8.
67. Kolomiets E, Swiderska MA, Kadam RU, Johansson EM, Jaeger KE, Darbre T, Reymond JL. Glycopeptide dendrimers with high affinity for the fucose-binding lectin LecB from Pseudomonas aeruginosa. ChemMedChem 2009; 4: 562–9.
68. McGarvey GJ, Benedum TE, Schmidtmann FW. Development of co- and post-translational synthetic strategies to C-neoglycopeptides. Org Lett 2002; 4: 3591–4.
69. Peri F, Cipolla L, La Ferla B, Dumy P, Nicotra F. A highly convergent approach to O- and N-linked glycopeptide analogues. Glycoconj J 1999; 16: 399–404.
70. Peri F, Cipolla L, Rescigno M, La Ferla B, Nicotra F. Synthesis and biological evaluation of an anticancer vaccine containing the C-glycoside analogue of the Tn epitope. Bioconjug Chem 2001; 12: 325–8.
71. Cipolla L, Rescigno M, Leone A, Peri F, La Ferla B, Nicotra F. Novel Tn antigen-containing neoglycopeptides: synthesis and evaluation as anti tumor vaccines. Bioorg Med Chem 2002; 10: 1639–46.
72. Ichikawa Y, Ohara F, Kotsuki H, Nakano K. A new approach to the neoglycopeptides: synthesis of urea- and carbamate-tethered N-acetyl-D-glucosamine amino acid conjugates. Org Lett 2006; 8: 5009–12.
73. Sureshbabu VV, Venkataramanarao R, Hemantha HP. A facile synthesis of C-terminal neoglycopeptides: incorporation of urea moiety between sugars and peptides employing Curtius rearrangement. Int J Pept Res Ther 2008; 14: 34–40.
74. Ichikawa Y, Matsukawa Y, Isobe M. Synthesis of urea-tethered neoglycoconjugates and pseudooligosaccharides in water. J Am Chem Soc 2006; 128: 3934–8.
75. Nagendra G, Hemantha HP, Sureshbabu VV. Synthesis of urea tethered glycosylated amino acids and glycopeptides mediated by DPPA employing N^{α}-Fmoc-Asp/Glu-5-oxazolidinones. Indian J Chem 2009; 48B: 397–407.
76. Lee DJ, Mandal K, Harris PW, Brimble MA, Kent SB. A one-pot approach to neoglycopeptides using orthogonal native chemical ligation and click chemistry. Org Lett 2009; 11: 5270–3.
77. Kuijpers BH, Groothuys S, Keereweer AB, Quaedflieg PJ, Blaauw RH, van Delft FL, Rutjes FP. Expedient synthesis of triazole-linked glycosyl amino acids and peptides. Org Lett 2004; 6: 3123–6.
78. Lin H, Walsh CT. A chemoenzymatic approach to glycopeptide antibiotics. J Am Chem Soc 2004; 126: 13998–4003.
79. Macmillan D, Blanc J. A novel neoglycopeptide linkage compatible with native chemical ligation. Org Biomol Chem 2006; 4: 2847–50.
80. Wan Q, Chen J, Chen G, Danishefsky SJ. A potentially valuable advance in the synthesis of carbohydrate-based anticancer vaccines through extended cycloaddition chemistry. J Org Chem 2006; 71: 8244–9.
81. Bay S, Fort S, Birikaki L, Ganneau C, Samain E, Coic YM, Bonhomme F, Deriaud E, Leclerc C, Lo-Man R. Induction of a melanoma-specific antibody response by a monovalent, but not a divalent, synthetic GM2 neoglycopeptide. ChemMedChem 2009; 4: 582–7.
82. Miller N, Williams GM, Brimble MA. Synthesis of fish antifreeze neoglycopeptides using microwave-assisted ''click chemistry''. Org Lett 2009; 11: 2409–12.
83. Hossany BR, Johnston BD, Wen X, Borrelli S, Yuan Y, Johnson MA, Pinto BM. Design, synthesis, and immunochemical characterization of a chimeric glycopeptide corresponding to the Shigella flexneri Y O-polysaccharide and its peptide mimic MDWNMHAA. Carbohydr Res 2009; 344: 1412–27.
84. Thayer DA, Yu HN, Galan MC, Wong CH. A general strategy toward S-linked glycopeptides. Angew Chem Int Ed Engl 2005; 44: 4596–9.
85. Cohen SB, Halcomb RL. Application of serine- and threonine-derived cyclic sulfamidates for the preparation of S-linked glycosyl amino acids in solution- and solid-phase peptide synthesis. J Am Chem Soc 2002; 124: 2534–43.
86. Galonic DP, van der Donk WA, Gin DY. Oligosaccharide-peptide ligation of glycosyl thiolates with dehydropeptides: synthesis of S-linked mucin-related glycopeptide conjugates. Chemistry 2003; 9: 5997–6006.
87. Galonic DP, Ide ND, van der Donk WA, Gin DY. Aziridine-2-carboxylic acid-containing peptides: application to solution- and solid-phase convergent site-selective peptide modification. J Am Chem Soc 2005; 127: 7359–69.
88. Ziegler T, Schips C. An efficient Mitsunobu protocol for the one-pot synthesis of S-glycosyl amino-acid building blocks and their use in combinatorial spot synthesis of glycopeptide libraries. Nat Protoc 2006; 1: 1987–94.
89. Watt GM, Boons GJ. A convergent strategy for the preparation of N-glycan core di-, tri-, and pentasaccharide thioaldoses for the site-specific glycosylation of peptides and proteins bearing free cysteines. Carbohydr Res 2004; 339: 181–93.
90. Murase T, Tsuji T, Kajihara Y. Efficient and systematic synthesis of a small glycoconjugate library having human complex type oligosaccharides. Carbohydr Res 2009; 344: 762–70.
91. Zhu X, Haag T, Schmidt RR. Synthesis of an S-linked glycopeptide analog derived from human Tamm-Horsfall glycoprotein. Org Biomol Chem 2004; 2: 31–3.
92. Zhu X, Pachamuthu K, Schmidt RR. Synthesis of novel S-neoglycopeptides from glycosylthiomethyl derivatives. Org Lett 2004; 6: 1083–5.
93. Bengtsson M, Broddefalk J, Dahmen J, Henriksson K, Kihlberg J, Lonn H, Srinivasa BR, Stenvall K. Convergent synthesis of neoglycopeptides by coupling of 2-bromoethyl glycosides to cysteine and homocysteine residues in T cell stimulating peptides. Glycoconj J 1998; 15: 223–31.

94. Whalen LJ, Halcomb RL. Synthesis of an isostere of an O-linked glycopeptide. Org Lett 2004; 6: 3221–4.
95. Crich D, Yang F. Synthesis of neoglycoconjugates by the desulfurative rearrangement of allylic disulfides. J Org Chem 2008; 73: 7017–27.
96. Shin I, Jung H, Lee M. Chemoselective ligation of maleimidosugars to peptide/protein for the preparation of neoglycopeptides/neoglycoprotein. Tetrahedron Lett 2001; 42: 1325–8.
97. Rodriguez EC, Marcaurelle LA, Bertozzi CR. Aminooxy-, hydrazide-, and thiosemicarbazide-functionalized saccharides: versatile reagents for glycoconjugate synthesis. J Org Chem 1998; 63: 7134–5.
98. Marcaurelle LA, Shin Y, Goon S, Bertozzi CR. Synthesis of oxime-linked mucin mimics containing the tumor-related T(N) and sialyl T(N) antigens. Org Lett 2001; 3: 3691–4.
99. Grigalevicius S, Chierici S, Renaudet O, Lo-Man R, Deriaud E, Leclerc C, Dumy P. Chemoselective assembly and immunological evaluation of multiepitopic glycoconjugates bearing clustered Tn antigen as synthetic anticancer vaccines. Bioconjug Chem 2005; 16: 1149–59.
100. Wilczewski M, Van der HA, Renaudet O, Dumy P, Coche-Guerente L, Labbe P. Promotion of sugar-lectin recognition through the multiple sugar presentation offered by regioselectively addressable functionalized templates (RAFT): a QCM-D and SPR study. Org Biomol Chem 2008; 6: 1114–22.
101. Renaudet O, Dumy P. On-bead synthesis and binding assay of chemoselectively template-assembled multivalent neoglycopeptides. Org Biomol Chem 2006; 4: 2628–36.
102. Zhao Y, Kent SB, Chait BT. Rapid, sensitive structure analysis of oligosaccharides. Proc Natl Acad Sci USA 1997; 94: 1629–33.
103. Durieux P, Fernandez-Carneado J, Tuchscherer G. Synthesis of biotinylated glycosulfopeptides by chemoselective ligation. Tetrahedron Lett 2001; 42: 2297–9.
104. Hatanaka Y, Kempin U, Jong-Jip P. One-step synthesis of biotinyl photoprobes from unprotected carbohydrates. J Org Chem 2000; 65: 5639–43.
105. Vila-Perello M, Gutiérrez-Gallego R, Andreu D. A simple approach to well-defined sugar-coated surfaces for interaction studies. Chembiochem 2005; 6: 1831–8.
106. Peri F, Dumy P, Mutter M. Chemo- and stereoselective glycosylation of hydroxylamino derivatives: a versatile approach to glycoconjugates. Tetrahedron 1998; 54: 12269–78.
107. Carrasco MR, Brown RT. A versatile set of aminooxy amino acids for the synthesis of neoglycopeptides. J Org Chem 2003; 68: 8853–8.
108. Carrasco MR, Nguyen MJ, Burnell DR, MacLaren MD, Hengel SM. Synthesis of neoglycopeptides by chemoselective reaction of carbohydrates with peptides containing a novel N′-methyl-aminooxy amino acid. Tetrahedron Lett 2002; 43: 5727–9.
109. Jiménez-Castells C, de la Torre BG, Andreu D, Gutiérrez-Gallego R. Neo-glycopeptides: the importance of sugar core conformation in oxime-linked glycoprobes for interaction studies. Glycoconj J 2008; 25: 879–87.
110. Peluso S, Imperiali B. Asparagine surrogates for the assembly of N-linked glycopeptide mimetics by chemoselective ligation. Tetrahedron Lett 2001; 42: 2085–7.
111. Peluso S, de LU, O'Reilly MK, Imperiali B. Neoglycopeptides as inhibitors of oligosaccharyl transferase: insight into negotiating product inhibition. Chem Biol 2002; 9: 1323–8.
112. Muramoto K, Yamauchi F, Kamiya H. Preparation of a photoactivatable fluorescent derivative of lactose and its application to photoaffinity labeling of a conger eel lectin. Biosci Biotechnol Biochem 1994; 58: 1013–7.
113. Gudmundsdottir AV, Paul CE, Nitz M. Stability studies of hydrazide and hydroxylamine-based glycoconjugates in aqueous solution. Carbohydr Res 2009; 344: 278–84.
114. 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.
115. McGrath RT, McKinnon TA, Byrne B, O'Kennedy R, Terraube V, McRae E, Preston RJ, Laffan MA, O'Donnell JS. Expression of terminal alpha 2-6 linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. Blood 2009 (DOI 10.1182/blood-2009-09-241547).
116. Barthel SR, Wiese GK, Cho J, Opperman MJ, Hays DL, Siddiqui J, Pienta KJ, Furie B, Dimitroff CJ. Alpha 1,3 fucosyltransferases are master regulators of prostate cancer cell trafficking. Proc Natl Acad Sci USA 2009; 106: 19491–6.
117. Harfouche R, Hentschel DM, Piecewicz S, Basu S, Print C, Eavarone D, Kiziltepe T, Sasisekharan R, Sengupta S. Glycome and transcriptome regulation of vasculogenesis. Circulation 2009; 120: 1883–92.
118. Janot M, Audfray A, Loriol C, Germot A, Maftah A, Dupuy F. Glycogenome expression dynamics during mouse C2C12 myoblast differentiation suggests a sequential reorganization of membrane glycoconjugates. BMC Genomics 2009; 10: 483.
119. Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, Chen CM, Khoo KH, Cheng TJ, Cheng YS, Jan JT, Wu CY, Ma C, Wong CH. Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc Natl Acad Sci USA 2009; 106: 18137–42.
120. Joyce JG, Krauss IJ, Song HC, Opalka DW, Grimm KM, Nahas DD, Esser MT, Hrin R, Feng M, Dudkin VY, Chastain M, Shiver JW, Danishefsky SJ. An oligosaccharide-based HIV-1 2G12 mimotope vaccine induces carbohydrate-specific antibodies that fail to neutralize HIV-1 virions. Proc Natl Acad Sci USA 2008; 105: 15684–9.
121. Leppanen A, Yago T, Otto VI, McEver RP, Cummings RD. Model glycosulfopeptides from P-selectin glycoprotein ligand-1 require tyrosine sulfation and a core 2-branched O-glycan to bind to L-selectin. J Biol Chem 2003; 278: 26391–400.