

**UNIVERSITA' DEGLI STUDI DI NAPOLI
FEDERICO II**



FACOLTA' DI SCIENZE MATEMATICHE FISICHE E NATURALI

PhD IN CHEMISTRY

**ANALYSIS OF THE PRIMARY AND
SECONDARY STRUCTURE OF
GLYCOSAMINOGLYCANS FROM
ALTERNATIVE SOURCES
(NATURAL OR SYNTHETIC)**

**Tutor
Dott. Cristina. De Castro**

**PhD student
Valentina Gargiulo**

XXII CICLO

UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II



**FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E
NATURALI**

DOTTORATO IN SCIENZE CHIMICHE

**ANALISI DELLA STRUTTURA PRIMARIA
E SECONDARIA DI GLICOSAMMINOGLICANI
DA FONTE ALTERNATIVA
(NATURALE O SINTETICA)**

Tutore
Dott.ssa Cristina De Castro

Candidata
Valentina Gargiulo

XXII CICLO

RIASSUNTO

Il progetto di dottorato che ho sviluppato negli ultimi tre anni è stato incentrato fondamentalmente sull'isolamento e caratterizzazione strutturale di glicosamminoglicani (GAG) da fonte naturale.

I GAG sono polisaccaridi lineari anionici ad alto peso molecolare (10-100 KDa), caratterizzati da un'unità ripetitiva disaccaridica formata da un acido uronico (β -D-GlcA, α -L-IdoA) e da un amminozucchero (β -D-GlcNAc, β -D-GalNAc) solo in un caso il residuo acido è sostituito da un normale esoso (β -D-Gal). In base al tipo di zuccheri presenti, alla natura del legame glicosidico, e alle posizioni coinvolte nei legami inter-residuo si identificano quattro famiglie di GAG:

- 1) Acido ialuronico (HA) [4- β -D-GlcA-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow)]
- 2) Eparina (HP)/Eparan solfato (HS) [4- α -L-IdoA/ β -D-GlcA-(1 \rightarrow 4)- β -D-GlcNAc-(1 \rightarrow)]
- 3) Condroitina (CS)/Dermatan solfato (DS) [4- β -D-GlcA/ α -L-IdoA-(1 \rightarrow 3)- β -D-GalNAc-(1 \rightarrow)]
- 4) Cheratan solfato (KS) [3- β -D-Gal-(1 \rightarrow 4)- β -D-GlcNAc-(1 \rightarrow)].

La particolarità degli ultimi tre tipi di GAG è legata all'elevata eterogeneità dovuta alla possibilità di ritrovare gruppi solfato in più punti dell'unità ripetitiva. Si tratta quindi di molecole con un'elevata densità di carica e caratterizzate da più di un'unità ripetitiva. Tutte queste strutture, tranne l'acido ialuronico, sono sintetizzate nel Golgi covalentemente legate a catene di ser-glicina (proteina core) e danno origine a strutture glicoconiugate, note come proteoglicani.

I GAG sono macromolecole largamente diffuse nei tessuti connettivi e possono essere ritrovate all'esterno, all'interno e anche sulla superficie delle cellule. Svolgono ruoli strutturali, ma la loro importanza è soprattutto collegata alla capacità di interagire con enzimi, chemiochine, citochine e fattori di crescita. Di conseguenza sono molecole coinvolte in numerosi ed importanti processi biologici e quindi per tale motivo sono numerosi e continui gli studi effettuati su di essi, al fine di comprendere meglio le loro relazioni funzioni-struttura.

Esistono in letteratura diversi protocolli per l'isolamento e la caratterizzazione di glicosamminoglicani. Essi si basano su uso combinato di classiche metodologie chimiche, digestioni enzimatiche, di tecniche di purificazione cromatografica e tecniche di indagine spettroscopiche. Negli ultimi anni, a tale approccio, utile soprattutto per la definizione della struttura primaria di GAG e loro derivati, è stata affiancata anche l'utilizzo di tecniche di calcolo computazionale, che consente di acquisire informazioni sulla conformazione di tale molecole nello stato libero o addirittura quando complessate a strutture proteiche.

Utilizzando quindi queste diverse metodologie, durante i tre anni di dottorato, mi sono occupata:

- a) Dell'isolamento e caratterizzazione strutturale di
- CS e KS da cartilagine di *Scyliorhinus canicula*
 - CS da cartilagine di *Raja bochyura*
 - CS da pelle di *Raja bochyura*
 - CS da cartilagine di *Torpedo nobiliana*
- b) Di studi conformazionali in mezzo esplicito su un decasaccaride di acido ialuronico, con il supporto di dati spettroscopici ottenuti in condizioni anisotrope (RDC).
- c) Di studi conformazionali in mezzo esplicito di oligosaccaridi solfati e non di condroitina solfato.
- d) Di studi conformazionali in mezzo implicito di oligosaccaridi solfati e non di cheratan solfato

Isolamento e caratterizzazione di molecole di CS da diversi tessuti di esemplari di Scyliorhinus canicula, di Raja bochyura e di Torpedo nobiliana.

Le catene di CS isolate dai vari tessuti, hanno la composizione riportata in tabella, stabilita in base alle analisi mediante cromatografia a scambio anionico (SAX-HPLC) sulla miscela di disaccaridi insaturi ottenuti tramite digestione con ABC lyase.

	<i>Scyliorhinus canicula</i>	<i>Raja bochyura</i> cartilagine	<i>Raja bochyura</i> pelle	<i>Torpedo nobiliana</i> cartilagine
CS0S ΔHexA(1→3)GalNAc	8.3 %	9.5 %	5.7 %	0.9 %
CS-A ΔHexA(1→3)GalNAc6S	41 %	36.5 %	37.1 %	30.9 %
CS-C ΔHexA(1→3)GalNAc4S	32 %	39.0 %	38.4 %	38.2 %
CS-D ΔHexA2S(1→3)GalNAc6S	19.8 %	15 %	18.8 %	30 %

La consistente presenza di motivi mono e disolfati, rispetto a quella di specie non solfatate, indica che si tratta di molecole ad elevata densità di carica, dato supportato dall'eluizione di tali prodotti, durante i processi di purificazione cromatografica, con tamponi ad elevata forza ionica (NaCl 700 mM per la CS di *Scyliorhinus canicula* e 1M per le altre).

Per tutte le molecole di CS la composizione è quella classica (GlcA, GalNAc e tracce di Xyl e Gal), non sono stati ritrovati motivi rari contenenti fucosio, glucosio o gruppi solfato alla posizione 3 del GlcA.

Su le diverse molecole di CS, inoltre, è stata eseguita anche un'opportuna indagine per stabilire la presenza di IdoA e quindi di motivi $[-4)\text{-}\alpha\text{-L-IdoA}_{2S}\text{-(1}\rightarrow\text{3)-}\beta\text{-D-GalNAc}_{4S/6S}\text{-(1}\rightarrow\text{)]}$, mediante l'impiego combinato di enzimi specifici (AC lyase, B lyase e ABC lyase). In nessun caso si è evidenziata la presenza di una quantità di IdoA maggiore del 3% e quindi non si può parlare di catene ibride CS/DS. I dati ottenuti dalle diverse idrolisi enzimatiche, inoltre, sono stati utili alla comprensione della distribuzione dei motivi disaccaridici all'interno della catena.

Per l'analisi della CS di *Scyliorhinus canicula* sono state utilizzate anche altre tecniche per la quantificazione dei motivi presenti. Un primo metodo spettroscopico (Q-HSQC) basato sull'integrazione di una selezione di segnali dello spettro HSQC, un altro di tipo analitico (RP-HPLC). Quest'ultimo metodo è stato utile soprattutto perché ha consentito anche di separare più facilmente i prodotti di digestione enzimatica (disaccaride e glicopeptidi).

Sui campioni di CS isolati dai tessuti di razza e torpedine è stata eseguita anche un'analisi mediante l'utilizzo di AC lyase, accidentalmente durante i processi di separazione cromatografica si sono verificate reazioni di degradazione e peeling, quindi sono stati isolati diverse specie tri e trasaccaridiche.

Purificazione ed analisi di due molecole di cheratan solfato

La catena di KS isolata e caratterizzata deriva dalla cartilagine di *Scyliorhinus canicula*. Essa reca un elevato contenuto di acido sialico come zucchero di end-capping e contiene due zone, a differente grado di solfatazione, una in cui entrambi i residui dell'unità ripetitiva sono solfatati e una in cui è soltanto l'unità di GlcNAc ad essere solfatata. Questo materiale presenta un peso molecolare di circa 10 kDa e risulta essere ancorato alla proteina core mediante un legame di tipo O-glicosidico.

Studi conformazionali in mezzo esplicito su un decasaccaride di acido ialuronico, con il supporto di dati spettroscopici ottenuti in condizioni anisotropiche (RDC).

Al fine di ottenere ulteriori informazioni sull'assetto conformazionale dell'acido ialuronico, sono stati effettuati studi spettroscopici in condizioni isotropiche e anisotrope e studi di dinamica molecolare in mezzo esplicito su un decasaccaride di HA.

Gli studi in condizioni isotropiche (PBS 10 mM pH 7 in D₂O) e anisotropi che (PBS 10 mM pH 7 in D₂O, pentaetilen-glicol-mono-ottil-etero (C₈E₅), e n-ottanolo) sono stati utili per la stima delle costanti di accoppiamento dipolare residuo D_{CH}. Partendo da tali dati sperimentali e da quelli ottenuti mediante una MD in H₂O di 3,5 ns, attraverso un apposito software (Mspin) è stato possibile effettuare una selezione di conformeri (~30 % dell'intera dinamica). Analisi

condotte su tali conformeri hanno consentito di stabilire che rispetto all'insieme completo di strutture, i conformeri selezionati assumono una disposizione più elongata e non forme contratte. Pur non risultando una netta differenza tra i valori medi di Φ e Ψ dell'intero insieme di strutture rispetto a quelle della selezione, si è stimato che una proprietà (distanza testa-coda) dipendente dai valori di Φ e Ψ di ciascuna giunzione glicosidica è nettamente diversa. Partendo dai valori medi di Φ e Ψ , attraverso il software POLYS, sono state determinate le strutture elicoidali più stabili associabili alla molecola. In tal modo sono state ottenute due strutture ad elica levogire, una 3_2 e una 4_2 . Tali modelli sono stati poi paragonati ai conformeri della della selezione. Da tale comparazione è risultato che nella selezione, invece, 5.1 % alla 3_2 e il 1.2 % alla 4_2 , le strutture che non rientravano in tale percentuali, comunque presentano una conformazione più vicina ad un'elica 3_2 che 4_2 .

L'applicazione dei dati di RDC, quindi, ha consentito di stabilire che l'HA in soluzione tende ad avere una disposizione allungata riconducibile ad un'elica 3_2 .

Studi conformazionali in mezzo esplicito di ologosaccaridi solfatati e non di condroitina solfato.

Con lo scopo di ottenere informazioni conformazionali anche sul backbone di CS, sono stati condotti degli studi di MD in mezzo esplicito su esasaccaride non solfato e su un dodecasaccaride solfato rappresentativo delle unità ripetitive ritrovate nelle CS da *Scyliorhinus canicula*, da *Raja bochyura* e da *Torpedo nobiliana*. I dati di calcolo computazionale in entrambi i casi sono stati confrontati con dati spettroscopici (NMR).

Dall'analisi dei dati sull'esasaccaride non solfato è stato possibile che il legame β -(1 \rightarrow 3) è molto più flessibile rispetto a quello β -(1 \rightarrow 4). Ulteriore conferma si ha mediante l'analisi della persistenza di legami H, attraverso la quale è stato stabilito che il legame β -(1 \rightarrow 4) è stabilizzato maggiormente da tali interazioni rispetto al β -(1 \rightarrow 3).

L'analisi delle distanze interprotoniche tra i protoni dei diversi anelli ha confermato i dati sperimentali riguardo la presenza di intensi contatti NOE, solo tra i protoni a ridosso del legame glicosidico e l'assenza di NOE riconducibili a disposizioni atipiche, come disposizioni non-exoanomeriche. Sfruttando il software POLYS è stato inoltre possibile stimare il tipo di elica più stabile generabile in base ai valori medi degli angoli glicosidici ottenuti tramite la MD. In particolare si è riscontrata la presenza di un'elica sinistrorsa 3_2 .

L'analisi del dodecasaccaride solfato ha consentito di stabilire che la presenza dei gruppi solfato influenza particolarmente le giunzioni glicosidiche, rendendo meno ampie le fluttuazioni tra i valori possibili, in base al paragone con i dati relativi al prodotto non solfato. La disposizione dei gruppi acetili è sempre *trans*, il rotamero preferito è sempre il *gt* e la distribuzione dei contatti

NOE è analoga a quella riportata per la specie non solfatata. La distribuzione dei legami H è analoga a quella trovata per la molecola non solfatata quindi sembra non risentire particolarmente della presenza dei gruppi solfato.

Studi conformazionali in mezzo implicito di oligosaccaridi solfatati e non di cheratan solfato

Eseguendo calcoli di meccanica molecolare e dinamica molecolare in mezzo implicito è stato possibile stabilire che anche nel caso del cheratano la presenza di gruppi solfato influenza soprattutto le fluttuazioni degli angoli diedri Φ e Ψ , con particolare effetto su Ψ che tende ad assumere maggiormente valori positivi. In generale, comunque, si riscontra che il legame β -(1 \rightarrow 4) è più flessibile di quello β -(1 \rightarrow 3).

Summary

Chapter 1	1
The Glycosaminoglycans family	
1.1 Structural features and biological roles	1
1.1.1 Hyaluronic acid	3
1.1.2 Chondroitin and Dermatan sulfate	5
1.1.3 Heparin and Heparan sulfate	7
1.1.4 Keratan sulfate	10
1.2 The Proteoglycans	12
1.2.1 Extracellular proteoglycans	14
1.2.1.1 The Hyalectans	14
1.2.1.2 The Small leucine-rich proteoglycans	16
1.2.2 Cell surface and the basement membrane proteoglycans	18
1.3 Procedures and techniques for the isolation and structural characterization of glycosaminoglycans	20
1.3.1 Isolation and purification of the polysaccharide chains	21
1.3.2 Enzymatic degradations	21
1.3.3 Analysis of intact GAGs and the derived oligosaccharides	22
1.3.4 Spectroscopic characterization	23
1.3.5 Computational approach	25
Chapter 2	29
Isolation and structural characterization of galactosaminoglycans isolated from some cartilaginous fish specimen	
2.1 The chondroitin sulfate of small spotted dogfish (<i>Scyliorhinus canicula</i>)	29
2.1.1 Isolation and purification of the polysaccharide from fresh cartilage	30
2.1.2 NMR survey on the entire polysaccharide	30
2.1.3 Enzymatic degradation and chromatographic separations	34
2.1.3.1 Strong Anionic Exchange chromatography (SAX-HPLC) of the disaccharide mixture	34
2.1.3.2 Reverse-phase ion pairing chromatography (RPIP-HPLC) of the disaccharide mixture	35
2.1.3.3 Reverse-phase ion pairing chromatography of glycopeptides mixture belonging to the linkage region	35
2.1.4 NMR analysis of the molecules isolated through the RPIP-HPLC	36
2.2 The chondroitin sulfate of ray and skate specimen	42
2.2.1 Isolation and purification of chondroitin sulfate from the cartilage and the skin of <i>Raja brachyura</i> and from the fresh cartilage of <i>Torpedo nobiliana</i>	43
2.2.2 Enzymatic degradation and HPLC-SAX separation of the disaccharide mixture belonging to the different chains	44
2.2.3 Estimation of the content of Iduronic acid	46
2.2.4 Analysis of the AC depolymerisation products	47
2.2.4.1 NMR analysis	53

2.3	The keratan sulfate of small spotted dogfish	64
2.3.1	Isolation and purification of the polysaccharide	64
2.3.2	Molecular weight measurement	64
2.3.3	Structural characterization by NMR	65
2.4	Conclusions	68
2.5	Material and Methods	69
2.5.1	Proteolytic digestion of the fresh tissue	69
2.5.2	Anionic exchange chromatography: Q-sepharose	70
2.5.3	Chemical Analysis conditions	71
2.5.4	SDS-PAGE conditions	71
2.5.5	Enzymatic degradations	71
2.5.5.1	ABC/ AC and B lyase digestions	71
2.5.5.2	Protease digestion	72
2.5.6	Alkaline hydrolysis in reductive condition	72
2.5.7	Size Exclusion Chromatography (SEC) conditions	72
2.5.7.1	BioGel supports	72
2.5.7.2	Sephacryl supports	73
2.5.8	HPLC analyses	73
2.5.8.1	Strong Anionic Exchange conditions	73
2.5.8.2	Reverse Phase Ion-Pairing conditions	74
2.5.8.3	Size Exclusion conditions	74
2.5.9	NMR analysis	75
Chapter 3		76
	Molecular Modelling on glycosaminoglycans oligosaccharides	
3.1	Conformational studies on a deca-saccharide of hyaluronic acid with the support of Residual Dipolar Coupling	76
3.1.1	The Residual Dipolar Coupling	76
3.1.2	Isolation of the oligosaccharide and collection of spectroscopic data	78
3.1.3	Molecular mechanics calculations	80
3.1.4	Molecular Dynamics simulation in explicit water	81
3.1.5	Analysis and refinement of simulation data	83
3.1.6	Helix parameters estimation	87
3.2	MD simulation on an unsulfated hexasaccharide of chondroitin sulfate	
3.2.1	Isolation and NMR analysis of the hexasaccharide	90
3.2.2	Molecular Mechanics calculations	94
3.2.3	MD in explicit water	96
3.2.4	Analysis of the MD trajectory and comparison with experimental data	96
3.2.5	Helix parameters estimation	104
3.3	MD simulation on a representative dodecasaccharide of shark chondroitin sulfate	105
3.3.1	Molecular mechanics studies on sulfated CS disaccharides	106
3.3.2	MD simulation in explicit water	109
3.3.3	Analysis of the simulation data	109

3.3.4	Comparison with experimental data and comparison with data of unsulfate molecule	119
3.4	Molecular mechanics and dynamics on oligosaccharides of keratan sulfate	121
3.4.1	Molecular mechanics on unsulfate and sulfate disaccharides	122
3.4.2	Molecular dynamics in implicit solvent on unsulfated and sulfated oligosaccharides	124
3.5	Conclusions	129
3.6	Material and Methods	129
3.6.1	Isolation of oligosaccharides	129
3.6.2	NMR spectroscopy	130
3.6.3	Refinement of RDC data	132
3.6.4	Molecular mechanics calculations	132
3.6.5	Molecular dynamics calculation	133
3.6.5.1	MD in explicit water	133
3.6.5.2	MD in implicit solvent	134

SUMMARY

The PhD project developed during these last three years concerned the analysis of the primary and secondary structure of Glycosaminoglycans.

The aim of the work was the collection of new information (spectroscopic and conformational ones) about Hyaluronic acid, Chondroitin sulfate and Keratan sulfate, in order to amplify the literature data available about their structures-function relationship.

Glycosaminoglycans (GAGs), are a family of linear polydisperse polysaccharides that take part in several physiological phenomena, including neuronal development, cell–matrix interactions, and activation of chemokines, enzymes, and growth factors. These chains are typically linked to a protein core forming proteoglycans at the cell surface or in the extracellular matrix.

The ability of GAGs to regulate these processes is attributed to their complex structure, which arises from extensive modifications (sulfation, deacetylation and epimerization) of a nonsulfated precursor consisting of hexosamine (D-Glucosamine, D-Galactosamine) and either hexuronic acid (D-Glucuronic acid, L-iduronic acid) or galactose residues, arranged in an alternating linear sequence.

Based on of the chemical structure backbone, there are four classes of GAGs: Hyaluronan (HA), Heparin (HP)/Heparan sulfate (HS), Chondroitin (CS)/Dermatan sulfate (DS) and Keratan sulfate (KS).

Despite their simple backbone sequence, glycosaminoglycans are complex molecules characterized by a heterogeneous structure, for this reason the investigation on them is really challenging and it is performed usually combining many techniques.

In this work, analyses were carried out applying a combination of chemical, enzymatic and spectroscopic tools. In some cases the spectroscopic data collected were analyzed with the support of theoretical data obtained through molecular mechanics and dynamics simulations.

The complete work, for a easier organization, is divided into two parts:

- Structural characterization of natural GAG samples:
 - Chondroitin sulfate from the cartilage of *Scyliorhinus canicula*, *Raja brachyura*, and *Torpedo nobiliana* and also from the skin of *Raja brachyura*.
 - Keratan sulfate from the cartilage of the *Scyliorhinus canicula*.
- Conformational analyses of GAG oligosaccharides:
 - Decasaccharide of hyaluronic acid
 - Hexasaccharide of unsulfate chondroitin
 - Dodedecasaccharide of chondroitin sulfate
 - Pentasaccharides of keratan sulfate

Structural characterization of natural GAG samples

Structural analyses were performed on chondroitin sulfate and keratan sulfate chains isolated from natural source: cartilaginous fishes. Purification, quantification and separation of GAGs polysaccharides and their enzymatic-derived oligosaccharides were performed using different chromatographic approaches like anion exchange, size exclusion chromatography (SEC), strong anion exchange HPLC (SAX-HPLC) and reverse phase ion pairing HPLC (RPIP-HPLC).

Spectroscopic investigations on the intact molecules and on the different oligosaccharides isolated were performed with NMR spectroscopy.

Chondroitin sulfate from the cartilage of Scyliorhinus canicula, Raja brachyura, and Torpedo nobiliana and from the skin of Raja brachyura

Compositional analysis of all the CS molecules led to the identification of the same repeating units (CS-0S [4)-GlcA-(1→3)-GalNAc-(1-], CS-A [-4)-GlcA-(1→3)-GalNAc4S-(1-], CS-C [-4)-GlcA-(1→3)-GalNAc6S-(1-], and CS-D [-4)-GlcA2S-(1→3)-GalNAc6S-(1)], but occurring in different percentages.

In the table below a summary of the results is reported:

Specimen	CS0S	CS-A	CS-C	CS-D
<i>Scyliorhinus canicula</i> Cartilage	8.2%	41%	32%	19.8%
<i>Raja bochyura</i> Cartilage	9.5%	36.5%	39.0%	15%
<i>Raja bochyura</i> Skin	5.7%	37.1%	38.4%	18.8%
<i>Torpedo nobiliana</i> Cartilage	0.9%	30.9%	38.2%	30%

This analysis, as expected, highlighted for all the samples a high negative charge density, and among the four CS samples, that isolated from the torpedo cartilage showed the higher charge density, since it was characterized by the higher content of disulfated motif and the minor content of the unsulfated one.

For the characterization of the CS isolated from the cartilage of *Schiliorhinus canicula* a new quantitative spectroscopic method was applied (Q-HSQC) too and the data collected were confirmed by the classical analysis (SAX-HPLC) of the mixture of the unsaturated disaccharides obtained by enzymatic depolymerisation (ABC lyase digestion). This disaccharide mixture and that containing the linkage region glycopeptides, moreover, were separated by RPIP-HPLC leading to sample directly analyzed by 2D-NMR.

Characterization of the CS samples isolated from *Raja Bochyura* and *Torpedo Nobiliana* tissues was performed by a deep SAX-HPLC analysis of the oligosaccharides mixtures belonging to ABC and AC lyase depolymerisations.

The separation of the products resistant to AC lyase action has accidentally led to the isolation of oligosaccharides with a structure different from the one expected, since peeling and degradation reactions occurred during the separation steps.

In this way besides the classical compositional analysis based on the SAX-HPLC separation of the unsaturated disaccharide mixture a collection of NMR data of CS tri and tetrasaccharides was achieved as well.

Keratan sulfate from the cartilage of the Scyliorbinus canicula:

Keratan sulfate material was classified as a O-linked glycan with a molecular weight around 10 kDa. The polysaccharide chain, as expected for shark KS, was characterized by GlcNAc6S units and both Gal and Gal6S units. The structure was not fucosylated, but resulted end-capped by sialic acid units.

Conformational analyses of GAG oligosaccharides

The conformational studies on Hyaluronic acid, Chondroitin sulfate and Keratan sulfate oligosaccharides were carried out through a combination of spectroscopic and molecular modelling techniques¹.

Decasaccharide of hyaluronic acid

The conformational behaviour of a HA decasaccharide was investigated through the analysis of theoretical data, collected by a molecular dynamics simulation in explicit water, with the support of spectroscopic data collected through Residual Dipolar Coupling NMR techniques.

This particular NMR technique was used because the anisotropic conditions (liquid-crystal medium) determine a partial alignment with respect to the external magnetic field and estimation of the relative orientation of the corresponding C-H or H-H vectors with respect to the magnetic field encodes for long range information.

The combined use of these NMR and molecular modelling techniques led to establish that in solution the HA tended to assume a quite elongate conformation well described by a left-handed 3-fold helix.

Hexasaccharide of unsulfate chondroitin

The secondary structure of unsulfated and sulfated CS oligosaccharides was investigated through the analysis of simulation data collected by MD simulations in explicit water as well, but the spectroscopic data used as reference were collected through the classic NMR spectroscopy in isotropic conditions: D₂O or D₂O:H₂O mixture.

The conformational behaviour of the unsulfate CS molecule was similar to the one estimated for the HA, as expected on the basis of the structural similarity of the two chains, so, in this case a conformation compatible with a left-handed 3-fold helix geometry was found as well. Differently from

¹ All the studies performed using molecular modelling techniques in explicit solvent were performed at the CIB of Madrid with the help and supervision of the Prof. Jesus Jimenez-Barbero, during a six month exchange program for PhD students.

HA, however, a higher flexibility for the β -(1 \rightarrow 3) linkage was estimated, as indicated by the larger fluctuation range found for the torsional angle Ψ and by the weak H-bond network defined across the glycosidic junction.

Dododecasaccharide of chondroitin sulfate

Investigation on a chondroitin sulfate model sequence, supported by the comparison with the data collected for the unsulfated molecule, highlighted that the presence of sulfate groups affected first of all the behaviour of the glycosidic torsional angles, and not the *N*-Acetyl disposition, the CH₂OH preferred rotamer and the H-bond network, because they were similar to those detected for the unsulfated structure. In particular, the presence of sulfate groups led to a reduction of the fluctuation range of both Φ and Ψ torsional angles, with a higher effect on Ψ one, that shifted in many towards negative values.

Pentasaccharides of keratan sulfate

Unsulfated and sulfated KS pentasaccharides were studied with molecular mechanics and dynamics simulations in implicit solvent. In this case, the effect of the presence of sulfate groups on the conformation of polylactosamine backbone was investigated. As found for the chondroitin sulfate case, both the glycosidic linkages were affected by the presence of these bulked and negative charged groups and a shift to positive Ψ values was detected, as well.

CHAPTER 1

The Glycosaminoglycans Family

1.1 Structural features and biological roles¹

Glycosaminoglycans (GAGs) are unbranched, negative-charged, eventually sulfated polysaccharides, having a molecular weight ranging between 10-100 kDa. The repeating unit is a disaccharide made up by an hexauronic acid (D-Glucuronic acid, L-iduronic acid) linked to a *N*-Acetyl-hexosamine (D-Glucosamine, D-Galactosamine), just in one case, the acidic residue is replaced by an hexose (D-Galactose). The differences among the glycosaminoglycans chains are due to the kind of hexuronic acid and hexosamine present, to the glycosilation position and to the geometry of the glycosidic linkage, as well.

The typical feature of these polysaccharides is the high degree of sulfation due to the possibility of finding a sulfate group in different positions of the two sugars in the repeating unit. However one of the most known glycosaminoglycans, hyaluronic acid, is unsulfated.

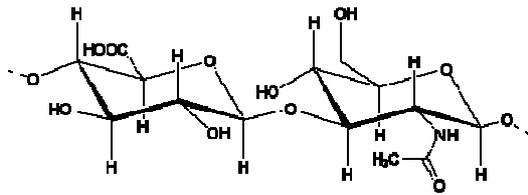
Based on the backbone chemical structure, there are four classes of GAGs: Hyaluronan (HA), Heparin (HP)/Heparan sulfate (HS), Chondroitin (CS)/Dermatan sulfate (DS) and Keratan sulfate (KS). According to the kind of hexosamine present the first two kinds are also classified as glucosaminoglycans, the last two as galactosaminoglycans. The repeating units and the possible sulfation positions are reported in Figure 1.1.

All the glycosaminoglycans chains, with the exception for the hyaluronic acid, are synthesized in the Golgi *O*-linked to a core protein, forming large polysaccharide-protein conjugates called Proteoglycans (PG). They can be found inside and around the cell, but they are first of all constituents of the extra-cellular matrix (ECM).

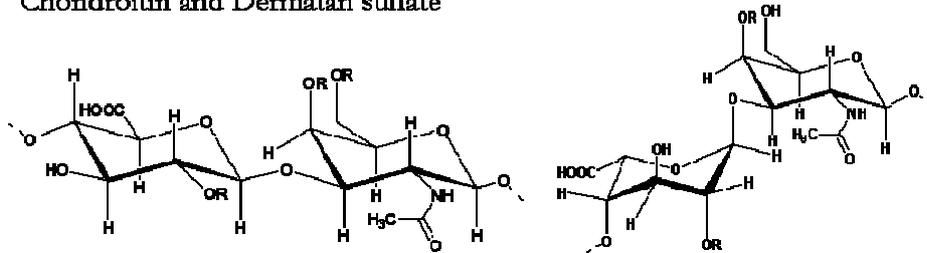
The GAGs were discovered at the beginning of the 20th century and they were classified as mucopolisaccharides, but the accumulation of data about their interaction with proteins and their participation in important biological processes (Fig 1.2) explains the growing interest about them and the large and rigorous studies on their structure and functions.

¹ Gandhi, N.S., Mancera, R.L., 2008 *Chem. Biol. Drug. Des.* 72:455-482.

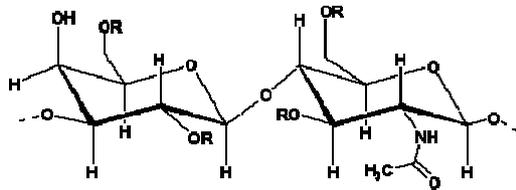
Hyaluronic Acid



Chondroitin and Dermatan sulfate



Keratan sulfate



Heparin and Heparan sulfate

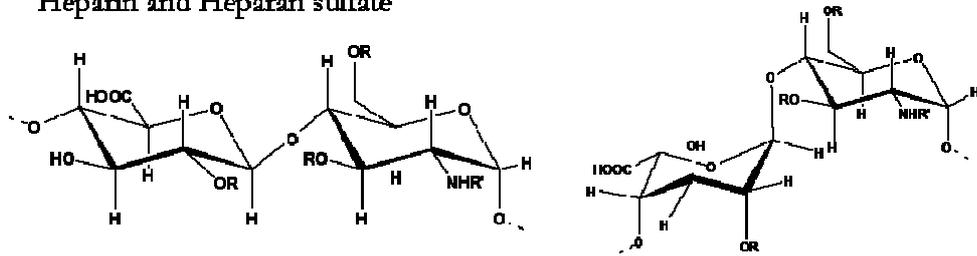


Figure 1.1: Glycosaminoglycans repeating units

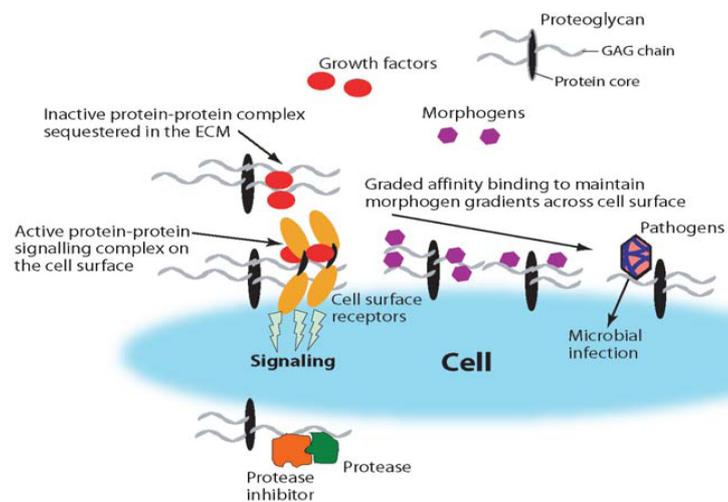


Figure 1.2: Schematic representation of the GAGs functions

1.1.1 Hyaluronic Acid

Hyaluronic acid (HA) or hyaluronan is the simplest polysaccharide belonging to the GAGs family because it is characterized by the disaccharide repeating unit [4)-D-β-GlcA-(1→3)-D-β-GalNAc-(1→] and it is unsulfated. It was discovered by Karl Meyer and John Palmer in the 1934 as an acidic component of the vitreous humor of the eye, and its name derived from the words hyaloid and uronic acid.

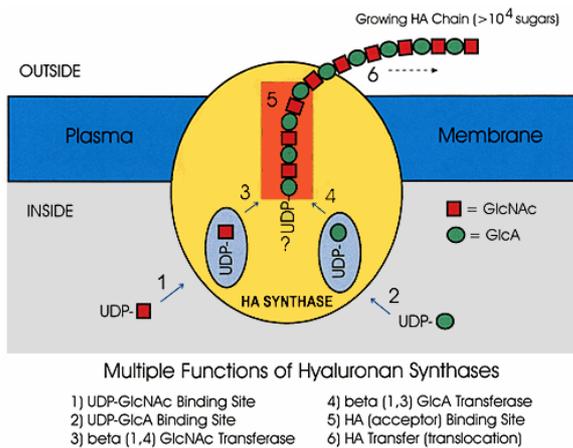


Figure 1.3: Hyaluronan syntase

Unlike other vertebrate glycoconjugates, HA is synthesized as free glycan, not attached to protein or lipids. The transmembrane hyaluronan synthases (HASs, Fig 1.3) synthesizes the chain *de novo* starting from UDP-sugars and it the growing chain is extruded out of the cell².

In the tissue, its concentration is maintained constant through a balance between its biosynthesis and catabolism, for instance it has been estimated that one the third of hyaluronan

in the human body is removed and replaced each day. It is removed through endocytic uptake within the tissue where it is made up, or in the lymph and in the liver. The internalized polysaccharide is then broken down by endosomal and lysosomal hyaluronidases³.

Hyaluronan exhibits unusual physicochemical properties in concentrated solution, for the combination of its random-coil structure, its large size, and its capacity to interact with a large number of water molecules. A molecule of HA shows a large hydrodynamic volume, and while its solutions show a high viscosity, and elasticity, these are essential features for its space filling, lubricating, and filtering properties⁴.

Different studies carried out through the combined use of spectroscopic and computational techniques suggest the occurrence of ordered domains in solution. Crystallographic analyses have revealed that HA oligomers exhibit regular helical conformation depending on the nature of the counterions, and on the pH^{5,6}. The tendency to assume a regular helix structure, namely a 4-fold left-handed structure, also in solution was demonstrated by a combined study on hyaluronan oligosaccharides with NMR and molecular dynamics calculation⁷.

² Weigel, P.H., DeAngelis, P.L. 2007 *J.B.C.* 282:3677-36781.

³ Tammi, M.I., Day, J.A., Turley, E.A. 2002 *J.B.C.* 277(7):4581-4584.

⁴ Laurent, T.C., Laurent, U.B., Fraser, J.R. 1996 *Immunol. Cell. Biol.* 74: A1-A7.

⁵ Arnott, S., Mitra, A.K., Raghunathan, S. 1983 *J. Mol. Biol.*, 169, 861-867.

⁶ Haxaire, K., braccini, I., Milas, M., Rinaudo, M., Perez, S. 2000 *Glycobiology* 10(6):587-594.

⁷ Almond, A., DeAngelis, P.L., Blundell, C.D. 2006 *J. Mol. Biol.* 358:1256-1269.

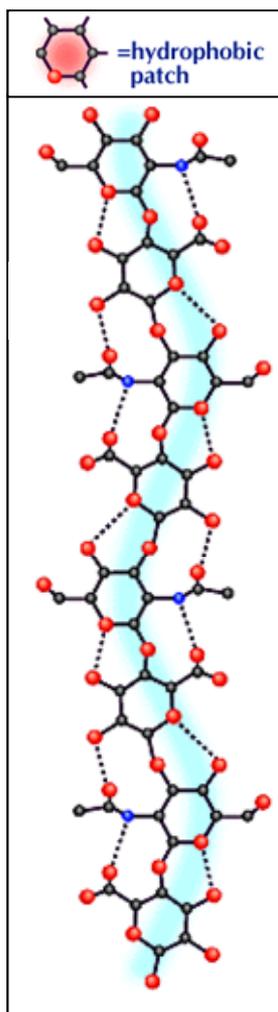


Figure 1.4: HA chain conformation

Polysaccharide chain in solution is stabilized by a solid network of H-bonds determining the occurrence of hydrophobic, and hydrophilic regions important in the intra-chains interactions⁸(Fig 1.4). In particular the conformation of the HA chain in solution allowed the disposition of each disaccharide twisted at a 180° angle compared to those ahead and behind in the chain. This particular disposition of the repeating units enhances the formation of weak interactions, namely H-bonds, across both the glycosidic linkages. Computational studies revealed that the persistence of H-bonds is high and the formation of H-bond is preferred than the formation of water bridges⁹. The presence of strong H-bonds across both glycosidic linkages is also demonstrated by the resistance of HA to be oxidated by periodate¹⁰. The network of weak interactions involving polar groups as the carboxylic and the acetoamidic ones, causes the formation of large hydrophobic regions, of about eight CH groups, on alternate sides of the HA chains. These hydrophobic patches were postulated to be important for the formation and the stabilization of duplex structures, but also to be a basis of the network-forming and laterally aggregating behaviour of the HA chains¹¹.

The original description done 70 years ago, classified HA just as an inert space-filler polysaccharide, but the discover of the large number of hyaluronan binding proteins (hyaladerins) revealed that HA takes part in

many biologically important processes¹².

Hyaluronan can be found in the extracellular matrix (ECM), on the cell surface, and inside the cell, and depending on its location it plays different roles.

In the ECM of connective tissues like cartilage, upon the interaction with the aggrecan and collagen, HA generates multimeric aggregates, maintaining, and preserving the architecture of the matrix under external events¹³.

The formation of a hyaluronan coat on the cell surface seems to be important in a variety of cell behaviours, such as cell adhesion, motility, growth, and differentiation.

The interaction with cell-surface receptors revealed also the ability of HA to activate various intracellular signalling cascades under certain physiological and pathological conditions. These

⁸ Lapčik, L., Lapčik, L., De Smedt, S., Demeester, J., Chabreček, P. 1998 *Chem. Rev.* 98(8):2663-2684.

⁹ Almond, A. 2005 *Carb. Res.*, 340: 907-920.

¹⁰ Scott, J.E., Tigwell, M.J. 1978 *Biochem. J.* 173:103-108.

¹¹ Scott, J.E., Cummings, C., Brass, A., Chen, Y., 1991 *Biochem. J.* 274:699-705.

¹² Day, A.J., Prestwich, G.D. 2002 *J.B.C.* 277(7):4585-4588.

¹³ Tozzo, R.V. 1998 *Ann. Rev. Bioche.* 67:609-652.

abilities are first of all related to the HA size: high-molecular-size hyaluronan promotes tissue integrity, and quiescence, namely HA molecules with a high-molecular weight exhibit anti-inflammatory, and immunosuppressive activities, while HA breakdown products signal that injury has occurred, and activate the recruitment of monocytes, and lymphocytes in the wound site, and the expression of inflammatory cytokines¹⁴.

1.1.2 Chondroitin and Dermatan Sulfate

Chondroitin sulfate is characterized by a repeating unit containing GlcA as acid residue and GalNAc as *N*-Acetyl hexosamine. The disaccharide unit contains alternating β -(1 \rightarrow 3), and β -(1 \rightarrow 4) linkages arranged according the following sequence [4)-GlcA-(1 \rightarrow 3)-GalNAc-(1 \rightarrow)]. Both sugars can be modified by sulfation during the biosynthesis, leading to different types of chondroitin sulfate repeating units. The commoner disaccharide units are classified as: CS-0S [4)-GlcA-(1 \rightarrow 3)-GalNAc-(1-), CS-A [-4)-GlcA-(1 \rightarrow 3)-GalNAc4S-(1-), CS-C [-4)-GlcA-(1 \rightarrow 3)-GalNAc6S-(1-), CS-E [-4)-GlcA-(1 \rightarrow 3)-GalNAc4S,6)-(1 \rightarrow)] and CS-D [-4)-GlcA2S-(1 \rightarrow 3)-GalNAc6S-(1-)]. In rare cases, the sulfate group, was detected also at position 3 of the acidic residue and the presence of α -Fuc residues or β -Glc in trisaccharidic repeating units have been identified¹⁵.

Dermatan sulfate contains iduronic acid (α -L-IdoA) in additions to GlcA and its commonest repeating unit is [-4)-IdoA(2S)-(1 \rightarrow 3)-GalNAc(4S)-(1-), normally classified as CS B.

Both chondroitin sulfate and dermatan sulfate are highly heterogeneous polysaccharides and their chains vary in size up to a hundred or more disaccharide repeating units, leading to a molecular weight averaging between 10 and 70 KDa.

They are synthesized in the Golgi and they are *O*-linked to a residue of serine. The linkage with the protein is not direct, but mediated by the tetrasaccharide: -4)- β -GlcA-(1 \rightarrow 3)- β -Gal-(1 \rightarrow 3)- β -Gal-(1 \rightarrow 4)- β -Xyl1 \rightarrow *O*-Ser. This oligosaccharide, depending on the sources, can be largely substituted by anionic groups like sulfates and phosphates¹⁶.

The biosynthetic process starts in the endoplasmatic reticulum with the synthesis of the core protein and then continues in the Golgi with the synthesis of the linkage region and the polysaccharidic chain (Fig 1.5). The first steps concern the biosynthesis of the tetrasaccharide and they are common also to heparin and heparan sulfate.

The linkage of the Xylose unit to the serine residue is formed by the action of the xylosyl-transferase (XylIT). The addition of the two units of galactose is performed, respectively, by the

¹⁴ Stern, R., Asari, A.A., Sugahara, K.N. 2006 *Eur. J. Cel. Biol.* 85:699-715.

¹⁵ Malavaki, C., Mizumoto, S., Karamanos, N., Sugahara, K. 2008 *Connective Tissue Research*, 49:133-139.

¹⁶ Krishna, N.R., Agraval, P.K. 2000 *Advanced in Carbohydrate Chemistry and Biochemistry* 56:201-234, Elsevier.

galactotransferase I (GalT-I) and galactotransferase II (GalT-II) and the unit of glucuronic acid is added by a glucuronyltransferase I (GlcAT-I). After the addition of these four sugars, the differentiation between the production of CS/DS chains or HP/HS chains takes place. The addition of a β -GalNAc unit by the β -N-acetylgalactosaminyltransferase I (GalNAcT-I) delineates the formation of a CS/DS chain.

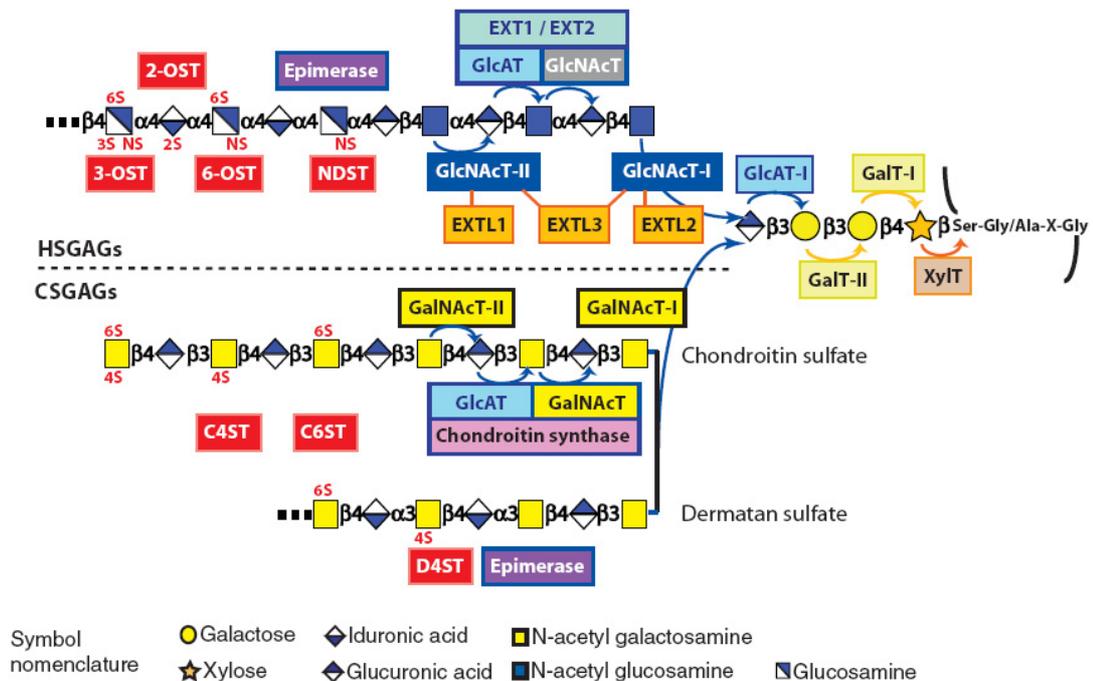


Figure 1.5: Biosynthetic pathways of CS, DS, HS and HP chains.

After the addition of the first GalNAc unit of the CS/DS chain, a multidomain enzyme, called chondroitin synthase, transfers both GalNAc and GlcA residues needed for the chain elongation. The action of sulfotransferases and the epimerase completes the biosynthetic process. In particular the action of the C-5 uronyl epimerase, that converts the D-GlcA to L-IdoA, is critical to the formation of DS chains. The addition of the sulfate groups is performed by two distinct enzymes, depending on the position of the substituent, 4-O-sulfotransferase (CS4T, DS4T) collocate the sulfate group only in the position 4 of the GalNAc, the 6-O-sulfotransferase (CS6T) only in the position 6¹⁷.

The chondroitin sulfate is found first of all in the connective tissue but also in the humour vitreous and blood vessels as component of structural proteoglycans of the extra-cellular matrix or around the cell as component of proteoglycans inserted into the plasma membrane. Similarly dermatan sulfate is a component of proteoglycans of the fibrous connective tissue and of blood vessels, but it can also be found in the skin.

¹⁷ Sasisekharan, R., Raman, R., Prabhakar, V. 2006 *Annu. Rev. Biomed. Eng.* 8:181-231.

The structure of chondroitin and dermatan sulfate, like in the case of the other sulfated GAGs, depends on the biological source: the animal species, tissues, physiological and pathological conditions.

Chondroitin and dermatan sulfate interact with a large variety of molecules as growth factors, cytokines, chemokines, adhesion molecules and lipoproteins. Consequently, they are involved in the central nervous system development (neurite outgrowth and axonal regeneration), in wound repair, infection, growth factors signalling and regulation, morphogenesis and cell division. Some diseases, like atherosclerosis and cancer, are often characterized by an imbalance in the biosynthesis of CS/DS-PGs¹⁴.

A "sulfation code" is hypothesized to explain the recognition of the different glycosaminoglycans chains by growth factors, namely it is hypothesized the precise position of sulfate groups along the chain, permits GAGs to encode information in a sequence-specific manner¹⁸.

Various biological functions of CS/DS, as happens for HP and HS, seem to be attributable to particular domain structures containing specific sulfation patterns. Particular is the case of the CS E motif that is able to inhibit the interactions with chemokines implicated in the inflammatory diseases and to favourite the binding with various heparin-binding growth factors including also the epidermal ones¹⁹. More over, CS chains containing the E disaccharide units have been found in the brain and they are implicated in the regulation of brain development due to the high affinity, comparable to the one of HS/HP chains, with midkine (MK) and pleiotrophin (PTN). After injury, CS/DS chains and the corresponding PGs can be modified, as happens, during inflammation: they are detached from the protein, became soluble and they are a major component in wound fluids. The soluble dermatan sulfate modifies several cell behaviours activating many growth factors, it can bind many molecules involved in the recruitment of leukocytes and in the early inflammatory events and it can also regulate the coagulation process. In particular, unlike HS, DS can interact only with thrombin and not also with fibrin²⁰.

1.1.3 The Heparin and Heparan sulfate

The simplest disaccharide unit of the heparin and heparan sulfate is [-4)-GlcA/IdoA-(1→4)- α -GlcNAc-(1-], but the possibility of finding an acid residue or the corresponding epimer, sulfate groups at the position 2 of the uronic acid, or at the positions 3 and /or 6 of the GlcNAc units

¹⁸ Gama, C.I., Tully, S.E., Sotogaku, N., Clark, P.M., Rawat, M., Vaidehi, N., Goddard III, W.A., Nishi, A., Hsieh-Wilson, L.C., 2006 *Nature Chemical Biology* 2(9):467-473.

¹⁹ Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., Kitagawa, H. 2003 *Current Opinion in Structural Biology* 13: 612-620.

²⁰ Taylor, K.R., Gallo, R.L. 2006 *The FASEB journal* 20:9-22.

and more over, the possible presence of a free, acetylated or sulfate amino groups leads to the identification of 48 different repeating units.

In heparin, 90% of the acidic units are IdoA residues and the trisulfated disaccharide [-4)-IdoA(2S)-(1→4)- α -GlcNS(6S)-(1-)] is the most frequent repeating unit found in a chain.

The content of sulfate groups in heparin is really high so it is classified as the molecule with the highest negative charge density of any known biological macromolecule.

Differently, heparan sulfate shows a more variable structure, which is tissue-dependent with a minor sulfate substitution and a major content of GlcA units, resulting often unsulfated.

The main difference between the two polysaccharides, however, is the localization: HS is ubiquitously distributed on both the cell surface and the extra-cellular matrix while HP has an intracellular collocation, namely it is found in the granules of the mast cells, that are found in organs such as liver, intestine and lung²¹.

Heparin was discovered, fortuitously, in the 1916 by Jay Mclean and its name is connected to the adjective hepatic, because the first it was isolated from the dog liver. Its discovery resulted important due to the anticoagulant properties showed since the beginning. This property explains the long and large use of heparin as drug to treat coagulation disorders. The heparan sulfate was discovered later and several studies have indicated its predominant role in the anticoagulant activity, differently from the heparin that it is always confined in the granules in the mast cells.

Both the polysaccharide exists as proteoglycans and are biosynthesized by a common pathway (Fig 1.5). They are attached to a serglycin protein in the same way of chondroitin and dermatan sulfate through the common tetrasaccharide segment, for which the biosynthetic steps are the same. The differentiation happens after the addition of the first α -GlcNAc unit by glycosyltransferase I (GlcNAcT-I). The chain elongation is obtained through the action of a multidomain glycosyltransferase that transfers both the GlcA and GlcNAc units. The growing chain then is modified by a multidomain enzyme (NDST) that acts as *N*-deacetylase and *N*-sulfotransferase. This modification is followed by the action of the C5 uronyl epimerase that converts the GlcA to IdoA. After epimerization, the 2-*O*-sulfation of the uronic acid takes place by the action of 2-*O*-sulfotransferase (2-OST), and the *O*-sulfation of the positions 3 and/or 6 of the GlcN units is performed by the 3-*O*-sulfotransferase (3-OST) and by the 6-*O*-sulfotransferase (6-OST), respectively.

The role of the endogenous heparin in the mast cells is not completely known: it is not involved in the blood coagulation process, but, it seems essential for the storage and coexistence of the varies positively-charged molecules occurring in the granules like proteins and histamine²².

²¹ Linhardt, R.J. 2003 *Journal of Medicinal Chemistry* 46(13):2551-2562.

²² Rabenstein, D.L. 2002 *Nat. Prod. Rep.* 19:312-331

Heparan sulfate, differently, plays different roles taking part in many biological processes. It can interact with growth factors modulating their activity, it can influence the recruitment of leukocytes to the sites of inflammation and injury, it can participate in the cytokine and chemokines regulation through the binding with P and L-selectins and it can influence the behaviour of many enzymes and enzymes inhibitors¹⁹.

The best known activity, however, is the anticoagulation one, performed through the interaction with the antithrombin III (ATIII), a serine protease inhibitor. After the binding with the sulfate polysaccharide, ATIII undergoes a conformational change and it is activated for the interaction with thrombin and other serine proteases of the coagulation cascade. The irreversible complex between the activated ATIII and the target protease is then cleared from circulation.

The smallest motif recognized in the HS and HP chains by the ATIII is a pentasaccharide containing a precise disposition of carboxyl and sulfate groups: [-4)-GlcNR-6S-(1→4)-GlcA-(1→4)-GlcNS-3,6S-(1→4)-IdoA-2S-(1→4)-GlcNS-6S-(1-)], and it is responsible of the conformational change of the protease inhibitor. Crucial roles are played by the trisulfated GlcN unit and by the 2-O-sulfated iduronic acid in the middle of the pentasaccharide.

The flexibility properties showed by the iduronic acid residues within the heparin chains seem to be responsible of most of its specific protein interactions.

IdoA residue can assume four different ring conformations (Figure 1.6), the classical 4C_1 , the other chair conformation 1C_4 and two twisted structures indicated as 2S_0 and 0S_2 . The preference of one conformation among the other possible ones depends on the

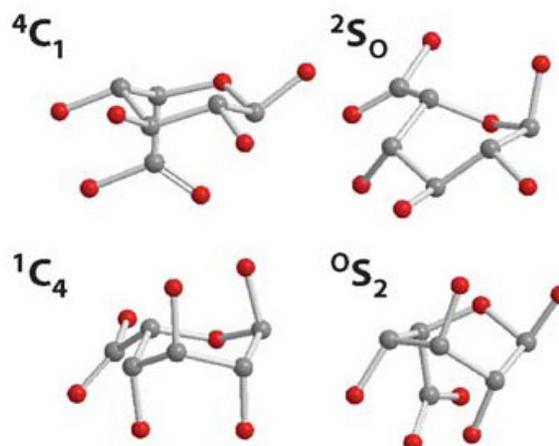


Figure 1.6: Ring conformations of iduronic acid

presence of sulfate groups and on the position of the residue in the polysaccharidic chain²³. An internal IdoA-2S unit can be in equilibrium between two conformations: the 1C_4 and the 2S_0 . Crystallographic and NMR studies performed on complex between proteins and heparin highlighted that the preference for a conformation depends on the bound site architecture, and also conformations, not classifiable at all as a chair or as twisted can be found.

²³ Ferro, D.R., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., Jacquet, J.C., Sinaÿ, Petitou, M., and Choay, 1986 *J.Am.Chem.Soc.* 108:6773-6778.

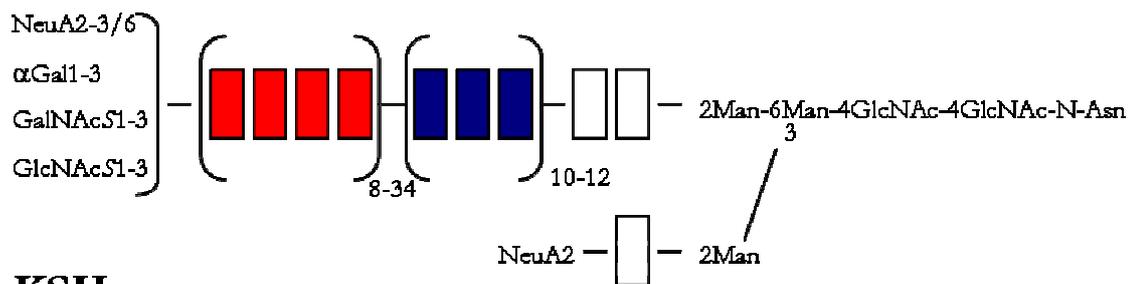
Keratan Sulfate

Keratan sulfate is characterized by a lactosamine-like repeating unit [-3)-Gal-(1→4)-GlcNAc-(1-] that can be sulfated in the position 6 of both sugars. This polysaccharide is an atypical member of the GAGs family, first of all for the absence of the acidic residue in the repeating unit, because it can be found *O*- or *N*-linked to a proteic structure and because its molecular weight usually is lesser of 30 kDa.

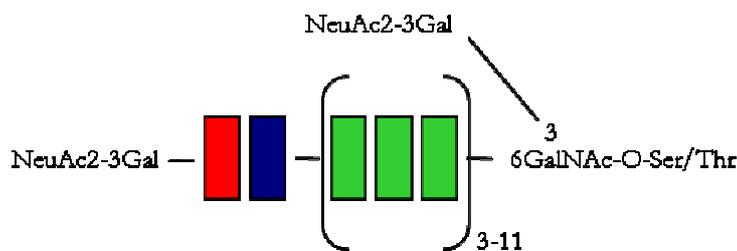
It was discovered in the 1939 by Suzuki as a component of corneal extracts, but successively was found in cartilage too. Another kind of keratan sulfate was discovered later in the brain.

Depending on the natural source, three kinds of keratan are known and classified as KSI (corneal KS), KSII (cartilaginous KS) and KSIII (brain KS). The main differences (Fig. 1.7) among the three KS species are in the nature of the linkage with the core protein, the linkage module and the sulfate pattern²⁴.

KSI



KSII



KSIII

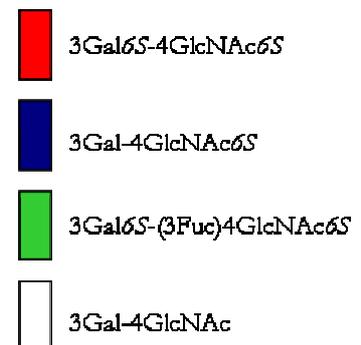
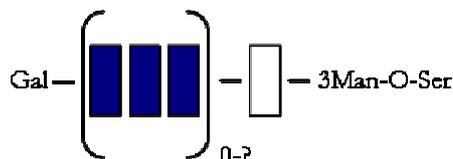


Figure 1.7: Structures of the three different types of keratan sulfate

The corneal KS (KSI) is linked to the core protein through a *N*-glycosidic linkage between a GlcNAc unit and an asparagine residue. The structures of the linkage module and of the chain

²⁴ Funderburgh, J.L., 2000 *Glycobiology* 10(10): 951-958.

are reported in Figure 1.7. Two branches start from the mannose rich linkage module, one really short made up just by a mannose unit, a lactosamine unit and capped by a sialic acid residue. The longest branch is characterized by unsulfated lactosamine units linked to the linkage module followed by 10-12 units of monosulfated units and then by 8-34 of disulfated units. The chain can terminate in different ways, with a sialic acid, with an α -Gal unit or with a sulfated GlcNAc or GalNAc unit.

The keratan sulfate of the cartilage (KSII) is O-linked to the core protein by a GalNAc unit and is richer in fucose residues. Looking at the structure reported in Figure 1.7, two branches linked to the linkage module are again present, one just made up by a sialic unit and a galactose unit and the other one characterized by a domain made up by fucosylated lactosamine units sulfated at both the sugars and by just one monosulfate and one disulfate lactosamine units. In this case the chain normally terminates with a sialic acid unit.

The structure of the brain keratan sulfate is not completely known, its main feature is the linkage to the core protein through just a mannose residue. The number of the monosulfate lactosamine units is not exactly known and the presence of fucose and sialic acid residues was detected²⁵. In Figure 1.7, the structure represented is more simplified and terminates with a unit of galactose.

The biosynthetic process for the production of KS proteoglycans is complex and not completely understood due the presence of three possible and different linkage modules. Little is known about the pool of enzymes synthesizing the linkage modules that attach the KS chains to the core protein, but probably they are the same enzymes responsible of the N/O-glycosilation of proteins with small branched oligosaccharides²⁶. The elongation of the polylactosamine chain (Figure 1.8) is performed through the combined action of a galactosyltransferase (GalT) and a N-acetylglucosaminyltransferase (GlcNAcT).

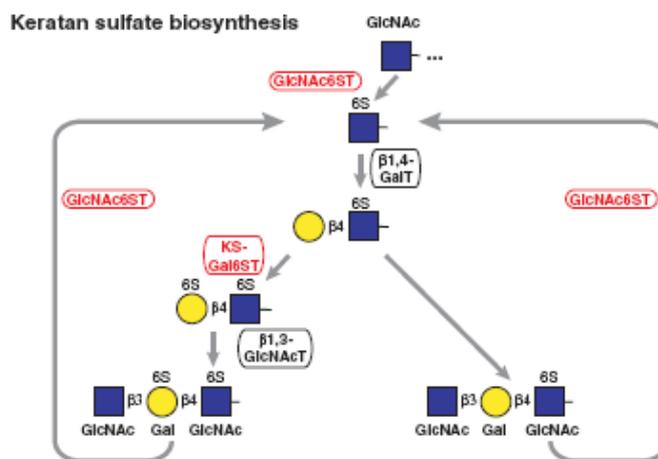


Figure 1.8: Keratan sulfate biosynthesis

The modification of the GlcNAc unit at the non-reducing end by a glucosamine-6O-sulfotransferase (GlcNAc6ST) is an essential requisite for the chain elongation. The construction of the chain, so, takes place at first with the addition of the Gal unit followed by the addition of

²⁵ Krusius, T., Finne, J., Margolis, R.K., Margolis, R.U. 1986 *J. Biol. Chem.* 261:8237-8242.

²⁶ Funderburgh, J.L., 2002 *IUBMBLife* 54:187-194.

the GlcNAc one, then a sulfate group on the GlcNAc unit is added. After this step a new unit of Gal is added and if the Gal unit has to be sulfated, before the addition of the new GlcNAc, the sulfation takes place through the action of a galactose-6O-sulfotransferase (Gal6ST)²⁷.

Keratan sulfate occurs in a wide range of tissues, it is not only found in cornea and cartilage, but also in bones, brain, pellucid zone, epidermis and endometrial epithelium²³. Keratan sulfate is a constituent of proteoglycans as aggrecan, but first of all it is a constituent of the small leucine-reach proteoglycans of the interstitial matrix.

In cornea, the high abundance of KS seems to be related to the maintenance of a level of tissue hydration critical for corneal transparency. This aspect is confirmed by the altered balance of glycosaminoglycans as consequence of macular corneal dystrophy²⁸.

KS seems to be involved in the implantation process due the higher abundance of cell-associated KS in the endometrial uterine lining during the menstrual cycle²⁹. More over, KS may play significant roles in endothelial cell migration and in other biological processes due its anti-adhesive properties³⁰.

1.2 The Proteoglycans

The proteoglycans (PG) are some of the largest and most complex molecular structures of the mammalian cells. These macromolecules are classified as glycoconjugates, due to the presence of a core protein glycosylated by glycosaminoglycan chains, and eventually by *N*- or *O*-linked oligosaccharides.

PG can be found as free components in the extra-cellular matrix or anchored to the cellular membrane.

Proteoglycans absolve many roles, taking part in many and important biological processes. Basically, proteoglycans are essential to life: they act as tissue organizers, biological filter, modulators of the growth factor activities and regulators of the collagen fibrillogenesis and of the skin tensile strength. They also influence the cell growth, the maturation of specialized tissues, the corneal transparency and neurite outgrowth. In some cases, the PGs affect tumour cell growth and cancer invasion and they take part in diseases development. The proteoglycans are largely involved in the inflammatory process and in the wound healing¹² as well.

For simplicity matters, the proteoglycans can be separated into three types: the extracellular, the cell surface (membrane-linked) and the basement membrane ones.

²⁷ Bülow, H.E., Hobert, O. 2006 *Annu. Rev. Cell. Dev. Biol.* 22: 375-407.

²⁸ Hassel, J.R., Newsome, D.A., Krachmer, J.H., Rodrigues, M.M. 1980 *Proc. Natl. Acad. Sci. USA* 77: 3705-3709.

²⁹ Graham, R.A., Li, T.C., Cooke, I.D. and Aplin, J.D. 1994 *Hum. Reprod.* 9: 9226-9230.

³⁰ Greenwood, J.A., Murphy-Ullrich, J.E. 1998 *Microsc. Res. Tech.* 43: 420-432.

The Table 1.1 contains a list of PGs belonging to the different types, the nature of GAG attached to the core protein and the corresponding tissue localization.

Table 1.1: Proteoglycans types, localization and GAGs chains present

Type	Name	GAGs and number of chains	Localization	Group
Cell surface	Glypican	HS(3-10)	Vascular Endothelium	HS proteoglycans
	Syndecan	HS(3-10)	Vascular Endothelium	
Basement membrane	Perlecan	HS (3)	Cartilage/Brain	
	Agrin	HS (3)	Cartilage/Brain	
	Bamacan	CS (3)	Skin	
Extracellular	Versican	CS/DS (~30)	Skin/Cartilage	Hyalectans or Lecticans
	Aggrecan	CS(~100)/KS	Cartilage	
	Neurocan	CS (3-7)	Brain	
	Brevican	CS (1-3)	Brain	
	Decorin	DS/CS (1)	Skin/Brain	Small Leucine-rich proteoglycans
	Biglycan	DS/CS (2)	Skin/Brain	
	Fibromodulin	KS (2-3)	Cartilage	
	Lumican	KS (3-4)	Cornea	
	Keratocan	KS (3-5)	Cornea	
	PRELP	KS (2-3)	Skin/Brain	
Osteodherin	KS (2-3)	Skin/Brain		
Epiphycan	DS/CS (2-3)	Cartilage		
Osteoglycin	KS (2-3)	Skin/Brain		

In addition to the constitutive proteoglycans, many proteins are classified as part-time proteoglycans, because their glycosilation is regulated by other factors.

One of the typical part-time proteoglycans is thrombomodulin, a critical mediator of endothelial anticoagulant defences³¹. Appican, the Alzheimer amyloid precursor protein (APP), in some cases occurs in a molecular structure bearing chondroitin sulfate side chains³². Another interesting example of a part-time proteoglycan is neuroglycan C (NGC), a transmembrane glycoprotein with a single EGF module whose expression is restricted in the central nervous system³³.

The CD-44, the best known cellular membrane receptor of hyaluronan, is sometimes classified as part-time proteoglycan due to the possibility to find one of its splicing variants glycosylated by GAGS chains on a ser-glycine sequence. This protein occurs in a wide array of alternately spliced forms, some of wich contains KS chains and some other HS chains^{34,35}.

³¹ Nadanaka, S. , Kitagawa, H., Sugahara., K. 1998 *J. Biol. Chem.* 273:33728-33734.

³² Pangalos, M.N., Efthimiopoulos, S., Shioi, J., Robakis, N.K. 1995 *J. Biol. Chem.* 270:10388-10391

³³ Aono, S., Keino, H., Ono, T. , Yasuda, Y. , Tokita, Y., Matsui, F., Taniguchi, M., Sonta, S., Oohira, A. 2000 *J. Biol. Chem.* 275:337-342 .

³⁴ Takahashi, K., Stamenkovic, I., Culter, M., Dasgupta, A., Tanabe, K.K. 1996 *J. Biol. Chem.* 271:9490-9096.

1.2.1 Extracellular proteoglycans

The family of extracellular proteoglycans comprises a large number of glycoconjugate structures. On the basis of the protein core nature, a secondary subdivision in two groups is usually performed: the Hyalactans and the Small Leucine Rich Proteoglycans (SLRP).

1.2.1.1 The Hyalactans

The hyalactans or lecticans, are large glycoconjugate molecules able to interact with hyaluronan and the different components of the extracellular matrix. Their structure can be divided into three different domains: an N-terminal domain for the HA binding, a central domain that carries the glycosaminoglycans chains and a C-terminal region containing a selectin-like domain with one or two EGF-like repeats, and a C-type lectin-like module. So these molecules interact with hyaluronic acid in the extracellular matrix or on the cell surface through the amino terminus region and they, acting as lectins, can bind oligosaccharides in the extracellular matrix or on the cell surface through the carboxyl terminus and.

Four different proteoglycans are classified as hyalactans: versican, aggrecan, neurocan and brevican (Tab 1.1).

Aggrecan (Fig. 1.10) is the main and largely studied CS-proteoglycan of the cartilage, it binds hyaluronic acid and proteins (collagens) in the extracellular matrix, forming huge aggregates leading to a hydrate gel-like matrix (Fig. 1.9). These multimeric aggregates are essential for the architecture of the cartilage, and they are responsible for the resistance to compression, and

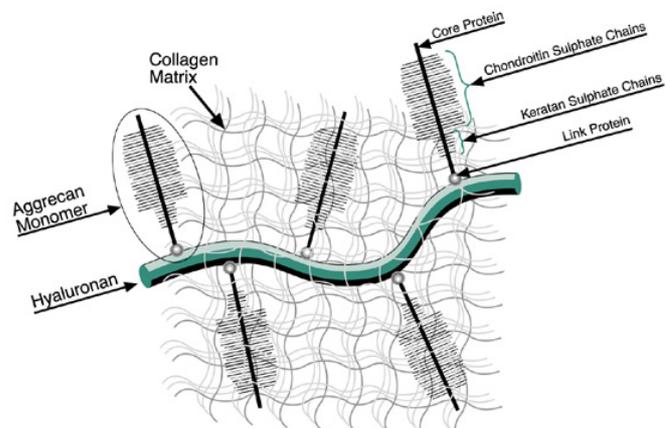


Figure 1.9: Schematic representation of cartilage ECM

deformation of joints. Aggrecan at high concentrations draws water into the tissue which swells, and expands the matrix, placing the collagen network under tension. Under this condition the tissue swells with water, resulting in a good compressive resilience.

Aggrecan is characterized by a core protein of ~220 kDa glycosylated by chains of chondroitin and keratan sulfate. In the core protein there are three different globular domains, G1, G2, and G3, an interglobular domain IGD between G1 and G2, and two sites for the glycosaminoglycans attachment located between the G2 and G3. The globular domain G1 is essential for the binding with hyaluronic acid, whereas the G3, located at the C-terminus of the core protein interacts with

³⁵ Tuhkanen, A.L., Tammi, M., Tammi, R. 1997 *J. Invest. Dermatol.* 109:213-218.

the components of the ECM. Finally, a large portion of the protein is glycosylated with up to 100 CS chains, and KS to a smaller extent³⁶. G2 presents a sequence similarity with G1, but the mutation in a few key residues has abolished its ability to link HA; actually the function of this domain is not clear.

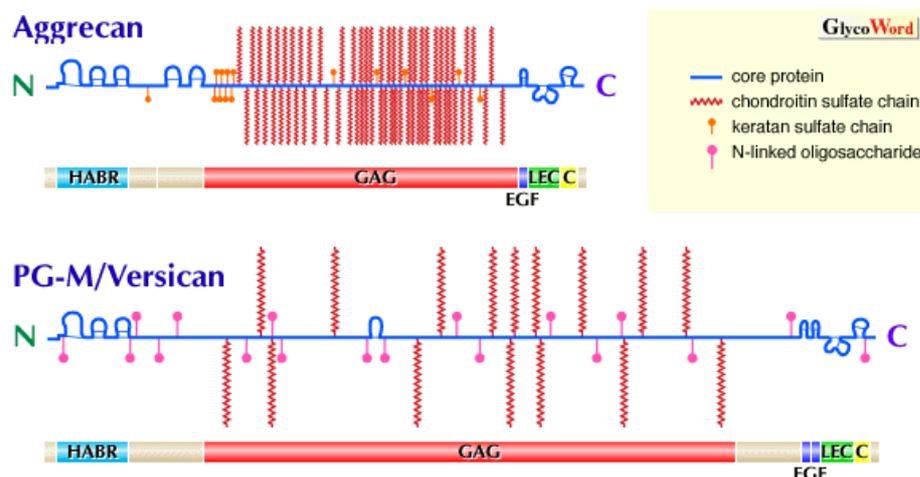


Figure 1.10: Aggrecan and Versican structures

Aggrecan undergoes molecular modifications as a consequence of catabolic and biosynthetic events leading to joint destructive diseases such as osteoarthritis, and rheumatoid arthritis. These events are regulated by many cellular, and extracellular events, and their extension is not even: factors such as species, site (which joint), zone (through the tissue depth), and region (topographical distribution), determine specific qualitative, and quantitative changes in proteoglycan³⁷. However, age appears to play the most important role on the composition of cartilage, and first of all in the sulfation pattern. In degenerative joint diseases like osteoarthritis, and rheumatoid arthritis, loss of aggrecan is observed with the consequent exposure of the collagen network to mechanical disruption, leading to the degradation of the articular cartilage. During the early stage of the development of this disease, a hypermetabolic repair response takes place with an increased synthesis of matrix components with an increased matrix turnover³⁸. In later stages the joints are characterized by extensive cartilage fibrillation, loss of PGs and bone eburnation where high contact stress take place. Cartilage breakdown products are antigenic and when released into synovial fluid may cause a synovial inflammation³⁹. Pharmacological management of the osteoarthritis is based on the use of analgesics, steroidal and nonsteroidal anti-inflammatory drugs, but also on the somministration of chondroitin sulfate as sodium salt³¹.

³⁶ Watanabe, H., Yamada, Y., Kimata, K., 1998 *J. Biochem.* 124:687-693.

³⁷ Bayliss, M., T., Osborne, D., Woodhouse, S., Davidson, C., 1999 *J. Biol. Chem.* 274:15892-15900.

³⁸ Hardingham, T., Bayliss, M., 1990 *Semin. Arthritis Rheum.* 20:12-33.

³⁹ Volpi, N., 2006 *Current Pharmaceutical Design* 12:639-658.

Versican is a proteoglycan rich in DS and CS chains, but the number of GAGs chains attached to the core protein is lower than aggrecan. It can be found in the skin, but also as a constituent of blood vessel. This last position explains the involvement of versican in the development of atherosclerotic lesions⁴⁰.

Versican can interfere with the attachment of embryonic fibroblasts at various substrate, including fibronectin, laminin and collagen.

Moreover, versican is found as CS-proteoglycan in the brain and, in particular, it is selectively expressed in embryonic tissues where it acts as a barrier to neuronal crest cell migration and to axonal outgrowth⁴¹.

Brevican and neurocan are small proteoglycans of the brain, and their functional roles are poorly understood. Differently from aggrecan and versican, the two brain PGs carry a little number of GAGs chains, as reported in Table 1.1.

Neurocan binds with a relatively high affinity neuronal cell adhesion molecules, inhibiting their homophilic interactions and blocks the neurite outgrowth.

Brevican is probably the most abundant hyalectan in the adult brain.

1.2.1.2 The small leucine-rich proteoglycans

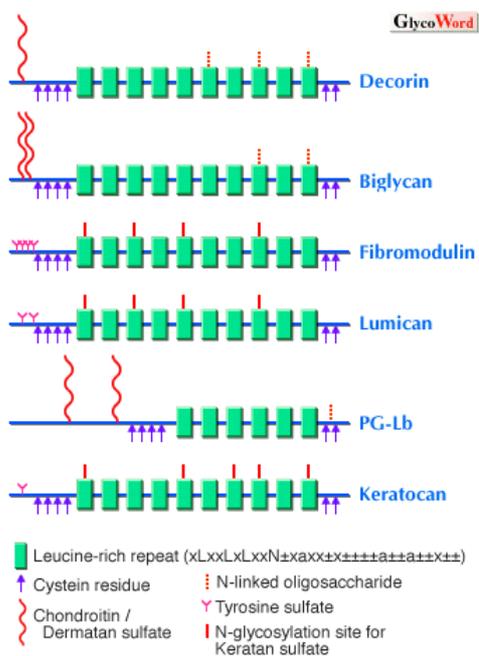


Figure 1.11: Family of SLRPs

The family of small leucine-rich proteoglycans (SLRPs) is composed by nine different PG molecules (Table 1.1, Fig. 1.11). They are CS, DS and first of all KS proteoglycans²⁴. All these PGs are characterized by a central domain containing leucine-rich repeats (LRR) flanked at either side by small cysteine clusters. The GAG attachment sites are essentially located in the cysteine rich region of the domain II. The level of complexity of these molecules increases considering the possible presence of several N-linked oligosaccharides.

Biglycan contains two chains of chondroitin sulfate (CS) or dermatan sulfate (DS) on a 38 kDa core protein. The chondroitin sulfate-containing form is commonly isolated from fetal or young bones, while the dermatan sulfate-containing form is isolated from articular cartilage.

Biglycan is found associated with the cell surface or pericellular matrix of a variety of cells including specific subsets of developing

⁴⁰ Khalil, M.F., Wagner, W.D., Golberg, I.J. 2004 *Atheroscler. Thromb. Vasc. Biol.* 24:2211-2218.

⁴¹ Galtrey, C.M., Fawcett, W. 2006 *Brain Research reviews* 54:1-18.

mesenchymal (skeletal muscles, bones and cartilage), endothelial (blood vessels) and epithelial (keratinocytes) cells.

Decorin contains one CS or DS chain on the 36 kDa core protein. The chondroitin 4-sulfate-containing molecule is isolated from the developing bones, while the DS chain-containing molecule is usually found in articular cartilage or tendons. Decorin is relatively abundant in bones, tendons, sclera, skin, aorta and cornea. In vitro experiments have shown that decorin modifies kinetics of collagen fibril formation and affects the morphology of collagen fibrils, and, interestingly, it can bind TGF-beta and can be involved in metabolic regulation of various cell functions.

Fibromodulin is a keratan sulfate proteoglycan present in many types of connective tissues, e.g. cartilage, tendons and skin. The core protein consists of 357 amino acid residues (42 kDa). Fibromodulin binds to type I and II collagen and affects collagen fibrillogenesis in vitro.

Lumican contains keratan sulfate chains and it is expressed in a number of connective tissues including cornea, muscles, intestine and cartilage. Lumican core protein consists of 338 amino acid residues (37 kDa). Lumican in cornea has highly sulfated keratan sulfate chains and may play an important role in the acquisition and maintenance of corneal transparency. However, lumican in other tissues is found as poorly sulfated or non-sulfated glycoprotein. Lumican was so named for its role in the transparency of cornea²⁴.

PG-Lb is a CS/DS proteoglycan first isolated from embryonic chick limb cartilage. Recently, occurrence in mice has been shown by the cDNA cloning and northern analysis. Epiphykan and DSPG3 have been found to be respectively the bovine and human equivalents of PG-Lb. Different from other SLRPs, PG-Lb is only expressed in the zones of flattened chondrocytes of limb cartilage. Recently, DSPG3 has been found in ligaments and placental tissues, as well as in cartilage.

Keratocan is so named because it contains keratan sulfate chains. Its distribution is rather limited compared with those of the other keratan sulfate-containing lumican or fibromodulin. It is abundant in cornea and sclera but much less in skin, ligaments and cartilage. Like lumican, keratocan in cornea is found to contain long keratan sulfate chains, but it is present in non-corneal tissues primarily as a non-sulfated glycoprotein.

Osteoglycin is one member of the LRP superfamily. It has recently been shown that this molecule exists in bovine cornea as a keratan sulfate proteoglycan and is named mimecan.

1.2.2 Cell surface and the basement membrane proteoglycans

The cell surface proteoglycans are usually divided in two groups: the syndecans and the glypicans families. The first family is made up by transmembrane heparan sulfate proteoglycans, while the second one by membrane linked HS-proteoglycans (Fig.1.12). Due to the structural variation of heparan sulfate and their physical location, these molecules play specific tasks during evolution with particular attention in the binding with many extracellular effector molecules, like growth factors. The syndecans, more over, take part in the anchorage of the cell to the ECM, maintain the endothelium morphology, and modulate the activity of many proteases and their inhibitors. The glypicans, on the other hand, can interact with molecules involved in cell adhesion and migration too⁴².

The syndecan family contains four members (syndecan-1/syndecan, syndecan-2/fibroglycan, syndecan-3/N-syndecan, syndecan-4/ryudocan (amphiglycan)), The syndecan family members are type I integral membrane proteins with homologous transmembrane and cytoplasmic domains. The combined transmembrane/cytoplasmic domains contain four well-conserved tyrosine residues, which might serve important roles for biological function⁴³.

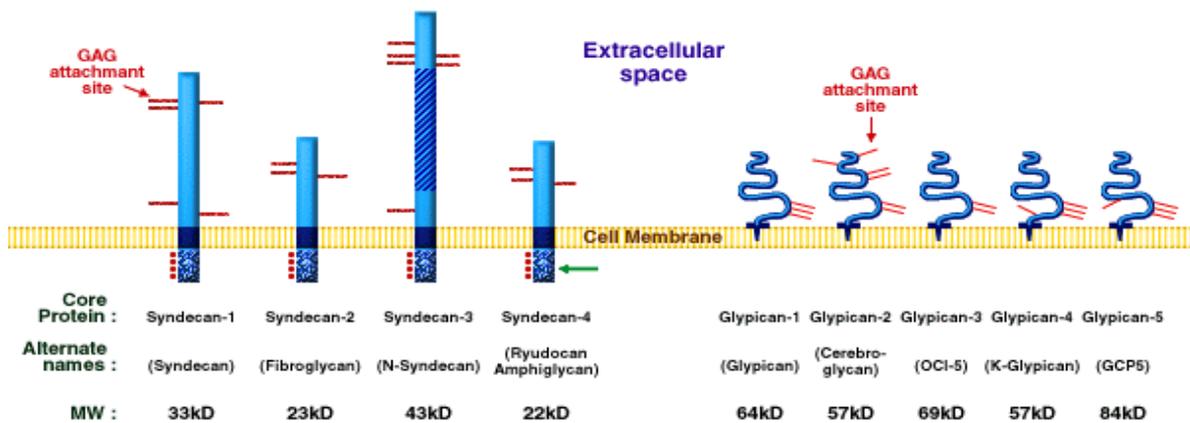


Figure 1.12: Structure, collocation and types of cell surface HSPGs

Differently from the other members of the same group, Syndecan-1 has both heparan sulfate and chondroitin sulfate glycosaminoglycan chains with a tissue-specific structural polymorphism due to distinct post translational modifications. Syndecan-1 has also been purified as an anticoagulant HSPG from endothelial cells or as a β FGF receptor molecule; this polymorphism reflects distinct HSPG functions⁴⁴.

The glypicans family is made up of five members (glypican-1/glypican, glypican-2/cerebroglycan, glypican/OCI-5, glypican-4/K-glypican, glypican-5). Glypican family members possess an

⁴² Proteoglycans: structure, biology and molecular interaction. Iozzo, R. CRC press, 2000.

⁴³ Bernfield, M., Kokenyesi, R., Kato M., Hinkes, M.T., Spring, J, Gallo, R.L, Lose, E.J.1992 *Ann. Rev. Cell Biol.*, 8:365-393.

⁴⁴ Rosenberg, R.D., Schworak, N.W., Lui, J., Schwartz, J.J., Zhang, L. 1997 *J. Clin. Invest.* 99:2062-2070.

extracellular region with GAG attachment sites, 14 invariant cysteine residues, which stabilize a highly compact tertiary structure, and a COOH-terminal GPI (glycosylphosphatidylinositol) anchor.

Glypican family members are selectively expressed on different cell types with only glypican-1 present on vascular endothelial cells. These HSPGs are mainly targeted to apical surfaces, and this process is partially dependent upon the extent of glycosilation. It is also suggested that glypican plays an important role in regulating the biological activity of fibroblast growth factors via HS GAG chains like syndecan⁴⁵.

The basement membrane proteoglycans are CS or HS proteoglycans and can be found in connective tissues, but also in brain and skin. Differently from hyalectans, the number of GAG chains carried by each PG is lower, ranging between 3-5.

Perlecan, Agrin and Bamacan are the three basement membrane proteoglycans. Among the three PGs, perlecan (Fig 1.11) is the most studied. Its name derives from the rotary shadowed electron microscopy appearance, suggesting a string of pearls. It is characterized by a big core protein (400-470 KDa) with numerous globular domains, interlinked by rod-like segments. The

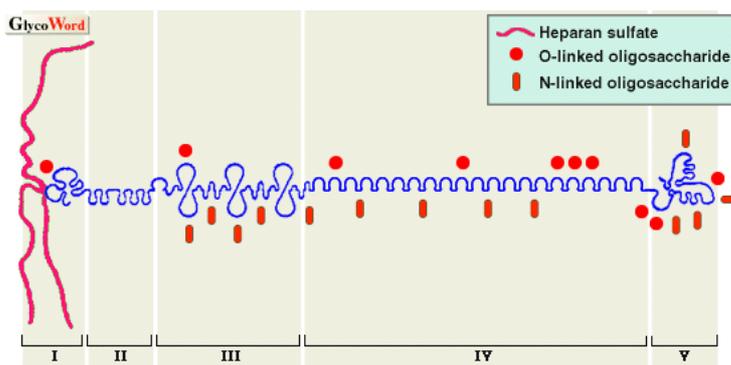


Figure 1.13: Perlecan structure

glycosaminoglycan side chains are located at the N-terminal of the protein.

Perlecan is involved in the cell differentiation and tissue morphogenesis, moreover many studies suggest that perlecan may play a role in the attachment of the

embryo to the uterine linings. Perlecan is also found as PG of cartilaginous tissue, where it persists throughout the development stages and into the adulthood. Its expression, normally, is related to tissue maturation of all vascularised organs like liver, lung, spleen, pancreas and kidney. One of the most important roles absolved by perlecan is the regulation of the access of growth factors to target cells. Perlecan, consequentially, is also involved in the growth and invasion of tumour cells.

Agrin is one of the best characterized molecules of the synaptic basement membrane, because it is the major heparan sulfate proteoglycans of neuromuscular junctions. It is characterized by a multidomain core protein and the central region has two conserved GAG attachment sites and many points for the attachment of O-linked oligosaccharides.

⁴⁵ Veugelers, M., Vermeesch, J., Reekmans, G., Steinfeld, R., Marynen, P., David, G. 1997 *Genomics* 40: 24-30.

Agrin induces the aggregation of acetylcholine receptors, and it is also a key organizer of the post synaptic apparatus.

However, agrin, is found in the renal tubules, in the basement membrane of skin, gastrointestinal tract and heart, so it has molecular functions beyond its role in synaptogenesis.

The last member of the basement membrane PGs is bamacan, a proteoglycan decorated just by chondroitin sulfate chains. Bamacan core protein is characterized by five different domains, and the II and the V contains potential binding sites for GAG chains. Bamacan is an important functional molecule within specialized basement membranes and it is implicated in the development of several disease processes.

1.3 Procedures and techniques for the isolation and structural characterization of glycosaminoglycans

Chemical characterization of glycosaminoglycans is really challenging considering the high heterogeneity within the chains, and the limited availability of the starting material. The investigation about GAGs structure and their structure-function becomes very difficult even because in many cases the presence of rare fragments with biological activity (such as ATIII-binding motif and the CS-E unit) is poor. These aspects are also amplified considering the poorness of chemical methodologies to depolymerise GAG chains. Classical hydrolyses cannot be easily performed, due to the lability in acidic conditions of the sulfate groups and the tendency of 4-linked uronic acid to undergo a β -elimination degradation in alkaline conditions⁴⁶.

Many protocols developed are based on enzymatic degradations of the polysaccharide chains and quantitative and qualitative analyses of the resulting oligosaccharide mixture through HPLC and CE.

GAGs characterization relies on spectroscopic techniques (NMR and X-Ray) and molecular modelling too. NMR is essential for the study of these molecules in solution, whereas crystallographic techniques, are important for the study of protein-GAG complex. NMR hardly can reach the same information because it does not allow a complete characterization of the binding site of the protein. Molecular modelling, when consistent with experimental data, allows the definition of the 3D shape of GAGs structure and eventually, their complexes with proteins. The following brief description is about the most important aspects of the procedures used for the isolation and structural characterization of GAGs.

⁴⁶ Monosaccharide Collins, P.M., Ferreir, R.J. Wiley, 1995.

1.3.1 Isolation and purification of the polysaccharide chains

GAGs are usually isolated from fresh tissues or biological fluids (blood, urine and saliva).

The extraction process is performed using protease (papain and pronase) in presence of metallo-proteins inhibitors (EDTA). In some cases a high temperature ($> 50^{\circ}\text{C}$) is applied to denature the sample, allowing the action of thermo-stable proteases. A pre-treatment with organic solvents can be performed, too⁴⁷. The polysaccharide components are recovered from the extraction mixture through a simple precipitation with ethanol, but ultracentrifugation or density-gradient centrifugation can be applied too. The purification of the resulting polysaccharide material is normally performed through a chromatographic separation and an anion exchange support is typically selected (DEAE, monoQ or Q-sepharose) on the basis of the polyanionic nature of the molecules⁴⁸. For little amounts of polysaccharide material, a purification with reverse-phase cartridges can be performed too⁴³.

After the extraction process, the reducing ends of GAGs chains are linked to peptides belonging to the core protein. Peptide removal is performed before the purification step, with a mild alkaline hydrolysis in reductive conditions to prevent peeling reactions.

1.3.2 Enzymatic degradations

Both bacterial and mammalian enzymes are essential tools for the depolymerisation and modification of GAG chains.

Two classes of enzymes are usually used in these processes: the lyase type and the glycosidase type.

Lyases cleave the glycosidic linkage between the hexosamine and the uronic acid through a β -elimination mechanism. This process generates a hexuronic acid characterized by an insaturation between the carbons in position 4 and 5. This particular structure is characterized by a typical absorbance at 232 nm, so the depolymerisation can be easily followed by a simple UV-detection.

The depolymerisation with lyases does not allow the differentiation between the two forms of the uronic acid (GlcA or IdoA), because the formation of the insaturation eliminates the differences between the two epimeric forms. This problem can be eliminated depolymerising GAGs with glycosidases, which hydrolyze the glycosidic linkage retaining the epimeric state of the uronic acid in the final product.

⁴⁷ Nandini, C.D., Itoh, N., Sugahara, K. 2005 *J.B.C.* 280(6):4058-4069.

⁴⁸ S.L., Carney, Cap 4, Carbohydrate analysis a practical approach II edition M.F., Chaplin, J.F., Kennedy, Eds, IRL press, 1994.

For chondroitin and dermatan sulfate, three types of lyases, named ABC lyase, AC lyase and B lyase⁴⁹ are used. The names reflect the kind of chondroitin sulfate preferred as substrate.

For heparin and heparan sulfate three types of lyases exist: heparanase I, II, III⁵⁰. The three enzymes have different substrate's selectivity.

Hyaluronidases are the most known glycosidases that can depolymerise hyaluronic acid. Both mammalian and bacterial hyaluronidases have been isolated, the differences between the two types of enzymes is related to the length of the fragment produced, and to the oligosaccharide structure used as a substrate⁵¹. Hyaluronidases can recognize and depolymerise also CS/DS and HS, but with a lower affinity.

All the enzymes able to depolymerise KS chains are glycosidases⁵². There are two types of keratanases (I and II), but β -gactosidases can be used too.

The enzymes can also be used to modify GAGs chains or the products of their depolymerisation, as reported for many types of exolytic sulfatase enzymes, which have been discovered and cloned⁵³.

1.3.3 Analysis of intact GAGs and separation of the derived oligosaccharides

Electrophoresis and especially HPLC are the techniques mostly used for the purification and analysis of intact GAGs. Compositional analysis and molecular weight determination are performed mainly with HPLC, selecting the appropriate chromatographic column.

The oligosaccharides enzymatically and/or chemically derived from GAGs depolymerisation can be separated using many analytical techniques. Many and accurate protocols for the quantitative and qualitative analysis of GAGs products have been applied and developed. These methods are based on liquid chromatography (LC, HPLC HPAEC), affinity chromatography and electrophoresis (SDS-PAGE, agarose-gel and capillary electrophoresis CE).

HPLC is the preferred technique for the separation and quantification of oligosaccharides belonging to sulfate glycosaminoglycans. The separation of the oligosaccharides can be performed by strong anion exchange HPLC (SAX-HPLC), reverse phase ion pairing HPLC (RPIP-HPLC), partition chromatography, gel permeation chromatography and reverse phase chromatography with precolumn derivatization, based on the differences in anion charge, polarity

⁴⁹ Linhardt RJ, Avci FY, Toida T, Kim YS, Cygler M. 2006 *Advances in Pharmacology* 53:187-215.

⁵⁰ Michoud, P., Da Costa, A., Courtois, B., Courtois, J. 2003 *Crit. Rev. In Biotechno* 23(4):233-266.

⁵¹ Stern, R., Jedrzejewski, M.J. (2006) *Chem. Rev.* 106(3):818-839.

⁵² Plaas, A.H.K., West, L.A., Midura, R.J. 2001 *Glycobiology* 11(10):779-790.

⁵³ Ernst, S., Langer, R., Cooney, C.L., Sasisekharan, R. 1995 *Crit. Rev. Biochem. Mol. Biol.* 30:387-444.

and molecular size. In literature can be easily found a list of HPLC columns and separation conditions developed for the different types of oligosaccharides⁵⁴.

Protocols allowing for the separation of HA oligosaccharides without the application of high pressure exist too, and they are based just on size exclusion separation⁵⁵.

CE and HPLCE are other valuable techniques for GAGs and for their derived-oligosaccharides too, and their results are comparable with those obtained through HPLC. CE and HPCE display high resolving power and great flexibility in the separation order. Moreover they are characterized by high separation efficiency, on line detection, simple operations and very low consumption of samples and buffers. Both normal polarity HPCE and reverse polarity HPCE methods have been used and developed, but the better method is the reverse polarity one, due to its capacity to separate efficiently unsaturated nonsulfated, mono-, di- and trisulfated disaccharides⁵⁶.

1.3.4 Spectroscopic characterization

The spectroscopic study of GAGs is first of all performed through NMR and X Ray diffraction⁵⁷. NMR is the most used technique for the study of glycan molecules in solution, because it is the only available technique that can determine both the anomericities and linkages of novel products. The structural characterization of oligosaccharides in aqueous solutions is based on the complete chemical shift assignment, the measurement of NOE contacts and scalar couplings. Through the application of mono and bidimensional experiments, the complete definition of the primary structure can be performed and a rough idea of the secondary structure achieved.

In general, the NMR experiments used for conformational studies of carbohydrates are similar to the ones used for other compounds. The type of NMR experiments that could be used for this purpose can span a domain from 1D ¹H and/or ¹³C spectra to 2D homo- and hetero-nuclear experiments such as DQF-COSY, TOCSY, T-ROESY, NOESY, HSQC, and HMBC⁵⁸.

Actually, all NMR parameters, are time-scale dependent and in all the cases, NMR time scale is too long with respect to that of the internal motions of saccharides. If the molecule is rigid, this feature, is not a problem, but when the molecule is flexible or converts between several conformations, NMR measures averaged values. This aspect is, in particular, relevant for NOE contacts, measured on a 50 ms to 1 s time scale⁵⁸.

⁵⁴ Imanari, T., Toida, T., Koshiishi, I., Toyoda, H. 1996 *Journal of Chromatography A*, 720:275-293.

⁵⁵ Tranchepain, F., Deschrevel, B., Courel, M.N., Levasseur, N., Le Cerf, D., Loutelier-Bourhis, C., Vincent, J.C. 2006 *Anal. Chem.*, 348 :232-242.

⁵⁶ Volpi, N. 2004 *Carb. Polym.* 55:273-281.

⁵⁷ Imberty, A., Lortat-Jacob, H., Perez, S. 2007, *Carbohydr. Res.* 342:430-439.

⁵⁸ Claridge, T.D.W. High-Resolution NMR Techniques in Organic Chemistry, Tetrahedron Organic Chemistry, Elsevier Science Ltd.

In the last years the combined use of isotope-labeled molecules and liquid-crystal technology enhanced the application of NMR spectroscopy in carbohydrate structural investigation⁵⁹. The anisotropic conditions determine a partial alignment with respect to the external magnetic field and the residual dipolar couplings can be measured⁶⁰. In this way, the relative orientation of the corresponding C-H or H-H vectors with respect to the magnetic field can be estimated and long range information can be collected too.

Classical monodimensional and bidimensional techniques have been widely applied in the study of intact GAGs and their derivatives^{61,62,63}. Particular methods have been developed for the quantification and characterization of unmodified GAGs in complex mixture, providing information on monosaccharide composition, sulfation pattern, positional linkages and amount of iduronic acid⁶⁴. The RDC technique has been applied in the structural characterization of GAGs derivatives as well^{65,66}, amplifying the conformational data available on GAGs backbone.

NMR procedures for the estimation of the presence and persistence of H-bond, in different solvents, have been applied in study of GAGs too⁶⁷. In particular, different kinds of measurements have been performed and information regarding temperature coefficients, NH proton chemical shifts and rate of exchange with solvent can be easily found⁶⁸.

The definition of the primary structure of a glycan, and so of a GAG chain, is not the unique target of NMR experiments. A lot of NMR techniques have been developed for the study of protein-sugar complex in solution⁶⁹. The Tr-NOE is a useful technique for the analysis of protein-carbohydrate complexes, but it can be applied only when the exchange between the complexed and uncomplexed states is fast enough⁴³. Saturation transfer difference (STD) is a similar technique. The saturation transfer-based experiments provide information about the ligands that are bound to the protein because the signals belonging to them are opposite in sign with respect to those free in the mixture⁷⁰.

The study of protein-GAG complex through NMR however, is not so frequent as that performed through crystallographic techniques.

⁵⁹ Jimenez-Barbero, J., Asensio, J.L., Cañada, F.J., Poveda, A. 1999 *Current Opinion in Structural Biology* 9:549-555.

⁶⁰ Lipsitz, R.S., Tjandra, N. 2004 *Annu. Rev. Biophys. Biomol. Struct.* 33:387-413.

⁶¹ Blundell, A.D., Reed, M.A.C., Almond, A. 2006 *Carbohydr. Res.* 341:2803-2815.

⁶² Mucci, A., Schenetti, L., Volpi, N. 2000 *Carbohydr. Polym.* 41:37-45.

⁶³ Huckerby, T.N. 2002 *Progress in Nuclear Magnetic Resonance Spectroscopy* 40:35-110.

⁶⁴ Guerrini, M., Naggi, A., Guglieri, S., Santarsiero, R., Torri, G., 2005 *Anal. Biochem.* 337:35-47.

⁶⁵ Silipo, A., Zhang, Z., Javier-Cañada, F., Molinaro, A., Linhart, R.J., Jimenez-Barbero, J. 2008 *ChemBioChem* 9(2): 253-260.

⁶⁶ Lan, J., Hricovini, M., Deakin, J.A., Lyon, M., Uhrin, D. 2009 *Glycobiology* 19(11):1185-1196.

⁶⁷ Bernet, B., Vasella, A. 2000 *Helvetica chimica acta* 83.

⁶⁸ Blundell, C.D., Almond A. 2007 *Magn. Reson. Chem.* 45:430-433.

⁶⁹ Meyer, B., Peters, T. 2003 *Angew. Chem. Int. Ed* 42(8):864-890.

⁷⁰ Meyer, M., Meyer, B., 2001 *J. Am. Chem. Soc.* 123:6108-6117

Glycans show a high resistance to crystallization due their internal flexibility, but when complex to protein they crystallize bound to the proteic structure and they can be analyzed also with X Ray diffraction. In the international databases a lot of structures of protein-sugar complex can be found, also belonging to protein interacting with GAGs. In particular, the complex A1III-heparin pentasaccharide is the most studied and characterized by X Ray diffraction⁷¹. In the last years, the number of structures of protein bound to GAGs oligosaccharides solved has grown considerably.

1.3.5 Computational Approach

Glycans functions depend on their bulk and three dimensional structure and these features acquire importance when they are linked to proteins, simply because of their large size and of their hydrophilicity, they can alter the behaviour of proteins, making them more soluble or resistant to proteolysis, hiding antigenic sites, altering their orientation on a membrane, providing additional recognition sites, etc.

To explore all these functions it is necessary to know the overall size, shape and charge of a glycan resulting from its detailed sequence. The techniques available for the determination of the conformation and dynamics of glycans are essentially molecular modelling techniques supported by spectroscopic data.

The individuals monosaccharides show a limited flexibility and in an oligosaccharide, they are assumed as rigid ring structures, so the conformation of a glycan can be performed just defining the behaviour of each glycosidic linkage. The connections (1→2), (1→3), and (1→4) can be fully described by the two torsion angles Φ and Ψ , whereas the connection (1→6) needs a third torsion angle ω .

The Φ torsion angle of a glycosidic linkage is determined largely by the *exo*-anomeric effect⁷². This is a stereoelectronic effect involving the lone pairs on the linkage oxygen. The Ψ torsion angle is determined largely by steric interactions and hydrogen bonding among residues and with the solvent. The ω torsion angle for 1→6 linkages can adopt three staggered rotamers⁷³ (Fig 1.14) based on steric interactions, referred to as *gauche*-*trans* (*gt*, *g* referring to the O₅-C₅-C₆-O₆ and *t* to the C₄-C₅-C₆-O₆ arrangements), *gauche*-*gauche* (*gg*), and *trans*-*gauche* (*tg*).

⁷¹ Jin, L., Pieter, J., Abrahams, I., Skinner, R., Petitous, M., Pike, R.N., Carrell, R.W. 1997 *Proc. Natl. Acad. Sci. USA* 94:14683–14688.

⁷² Lemieux, R.U., Koto, S., Voisin, D., 1979 *Anomeric Effect : Origin and Consequences*, Eds: Szarek, W.A., Horton, D. American Chemical Society Washington, DC Vol.87.

⁷³ Kirschner, N.K., Woods, R.J. 2001 *PNAS* 98(19):10541-10545.

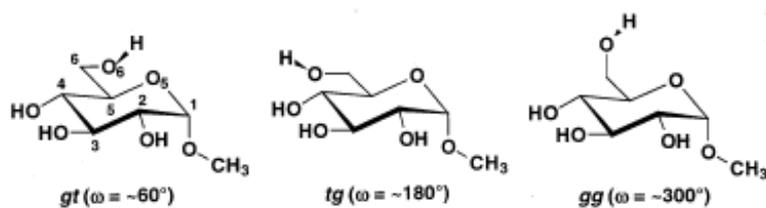


Figure 1.14: Three possible rotamers for the α -D-glucopyranoside

The determination of the torsion angles of each glycosidic bond seems simple, but is difficult to do in practise. The glycosidic linkages are flexible and for a single glycosidic junction a conformation with a large scale vibration, or two or more conformations that easily interconvert each other, can be defined. Consequently, for the complete characterization of a given glycosidic linkage, we need to know the number of the conformers adopted by the linkage, their flexibility and the time spent in each state. The study becomes harder if there are different kinds of linkages in the same oligosaccharide and it becomes mandatory to know how they behave⁷⁴.

The definition of a 3D-shape of an oligosaccharide is more challenging, considering the poorness of the experimental data available. Crystallographic data are hardly available due to the high resistance to crystallization of these molecules. The intrinsic flexibility of each glycosidic linkage prevents any homology study, largely performed in the conformational studies of proteins.

The experimental data available for glycans through NMR are first of all solution-derived data, and time average properties, making in many instances impossible the determination of the right conformation of the molecule among the many possible.

All these difficulties, related to the conformational analysis of oligosaccharides and to the interpretation of NMR data, explain the long history of application of computational methods to carbohydrates. Many modelling protocols and force field equations, suitable for carbohydrates, have been developed: MM2, MM3, AMBER, CHARMM, GROMOS and GLYCAM⁷⁵.

The force field, namely the equations containing the energy parameters required to describe the behaviour of the different kinds of atoms and bonds, is the basis of molecular mechanics (MM) and dynamics calculations (MD). The purpose of the molecular mechanics studies is to predict the energy associated with a given conformation of a molecule. The energy value estimated, depends on the number of atoms of the molecule and has no meaning on an absolute scale, so only energy differences between two or more conformations make sense.

Molecular dynamics targets, instead, are the conformational transitions and local vibrations of a molecule. The steps in a molecular dynamics simulation represent the changes in atomic position over the time and the resulting series of snapshots of structural changes over the time is called

⁷⁴ Wormald, M.R., Petrescu, A.J., Pao, Y., Glithero, A., Elliot, T., and Dwek, R.A. 2002 *Chem. Rev.* 102:371-386.

⁷⁵ Woods, R.J. 1998 *Glycoconjugate J.* 15:209-216.

trajectory. Molecular dynamics has no defined point of termination other than the amount of time that can be practically covered. Unfortunately, the current picosecond order of magnitude limit is often no long enough to follow many kinds of state to state transformations, such as large conformational transition in proteins. However, the MD simulation time, usually, is a good compromise between the complexity of the system and the computer time available.

Molecular mechanics and dynamics calculations are characterized by:

- Applicability to molecules containing thousands of atoms
- Applicability to organic molecules, oligonucleotides, peptides, proteins and saccharides
- Availability of vacuum, implicit or explicit solvent environments
- Characterization of ground state only
- Estimation of Thermodynamic and Kinetic (via molecular dynamics) properties

Molecular modelling procedures in vacuum or in implicit solvents requires the use of appropriate conditions for the simulation of the solvent bulk, for this reason solvent models have been developed.

The simulation in explicit solvents requires an appropriate setting of parameters and a previous construction of the solvent box around the solute. A realistic model of a solution requires a very large number of solvent molecules to be included along with the solute. The simple placement of the solute in a box of solvent is not enough, however, since while some solvent molecules will be at the boundary between solute and solvent, others will be within the bulk of the solvent and a large number will be at the edge of the solvent and the surrounding vacuum. This is, obviously, not a realistic picture of a bulk fluid. In order to prevent the outer solvent molecules from “boiling” off into space, and to allow a relatively small number of solvent molecules to reproduce the properties of the bulk, periodic boundary conditions are employed. According to this method the particles are enclosed in a box which is then replicated in all three dimensions to give a periodic array, a two-dimensional representation of which is shown Figure 1.15. In this periodic array a particle at position \mathbf{r} represents an infinite set of particles at:

$$\mathbf{r} + i\mathbf{x} + j\mathbf{y} + k\mathbf{z} \quad (i,j,k \rightarrow -\infty, +\infty)$$

where \mathbf{x} , \mathbf{y} and \mathbf{z} are the vectors corresponding to the box edges. During the simulation only one of the particles is represented, but the effects are reproduced over all the image particles with each particle not only interacting with the other particles but also with their images in neighbouring boxes. Particles that leave one side of the box re-enter from the opposite side as their image. In this way the total number of particles in the central box remains constant.

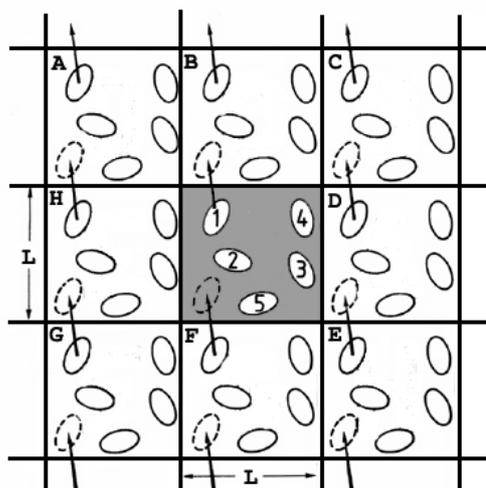


Figure 1.15: A two-dimensional array of boxes. As molecule 1 moves from the central box into box B it is replaced by its image which moves from box F into the central box. This movement is replicated across all the boxes.

Molecular mechanics and dynamics techniques are useful for predicting the solution conformations of relative rigid oligosaccharides, but also for the glycoconjugates and carbohydrate-protein complexes. Glycosaminoglycans oligosaccharides have been the subject of different theoretical studies, using both implicit and explicit solvent environments and many information can be found in the literature mainly on heparin and hyaluronic acid oligosaccharides^{7,76}.

⁷⁶ Zhang, Z., McCallum, S.A., Xie, J., Nieto, L., Corzana, F., Jimenez-Barbero, J., Chen, M., Liu, J., Linhardt, R.J. 2008 J. Am. Chem. Soc. 130(39):12998-13007.

CHAPTER 2

Isolation and structural characterization of galactosaminoglycans from some cartilaginous fish specimens

Glycosaminoglycans are polysaccharides widely distributed throughout the animal kingdom, while the marine species represent a useful source of these sulfated molecules with novel structures.

The discovery of new structures with peculiar functions is the target of many laboratories works performed on invertebrate marine specimens as ascidian¹, cucumber sea fish², squid³ and cartilaginous fishes.

Rays, skates, and sharks belong to the class *Chondrichthyes*⁴ (Greek, *chondro* meaning cartilage and *ichthos* meaning fish) and form the subclass called Elasmobranchii. Features of *Chondrichthyes* include a skeleton made up of cartilage, jaws, paired fins, and paired nostrils. Elasmobranchs are characterized by cylindrical or flattened bodies, five to seven pairs of gill slits, an upper jaw not fused to the cranium, and placoid scales. Placoid scales, also referred to as dermal denticles, have the same structure as a tooth consisting of three layers. These scales are arranged in a regular pattern on sharks and irregular patterns on rays and skates.

Different species of rays, skates, and sharks can be found in the Mediterranean sea, and some specimens, fished in the bay of Naples, have been object of these studies.

2.1 The chondroitin sulfate of lesser spotted dogfish⁵

(*Scyliorhinus canicula*)

Scyliorhinus canicula (Fig. 2.1), even known as lesser spotted dogfish, is a cartilaginous fish belonging to the *Scyliorhinidae* family of the Carcharhino-formers order⁶. This dogfish is a small, shallow-water



Figure 2.1 *Scyliorhinus canicula*

¹ Pavão, M.S., Mourão, A.S., Mulloy, B., Tollefsen, D.M. 1995 *J. B. C.* 52:31027-31036.

² Borsig, L., Wang, L., Cavalcante, M.C.M., Cardilo-Reis, L., Ferreira, P.L., Mourão, A.S.P., Esko, J.D., Pavão, M.S. 2007 *J. B. C.* 20:14984-14991.

³ Kinoshita, A., Yamada, S., Haslam S.M., Morris H.R., Dell, A., Sugahara, K. 1997 *J. B. C.* 32(8):19656-19665.

⁴ HAMLET W. 1999 *Sharks, Skates and Rays. The biology of elasmobranch fishes* The Johns Hopkins University Press Baltimore and London.

⁵ Gargiulo, V. Lanzetta, R., Parrilli, M. De Castro, R., 2009 *Glycobiology* 19(12):1485-1491.

⁶ Delabre, C., Spruyt, N., Delmarre, C., Gallut, C., Barriol, V., Janvier, P., Laudet, V., Gachelin, G., 1998 *Genetics* 150:331-344.

shark with blunt heads, and a slender, elongate body with two dorsal fins, located towards the tail end of the body. The length of the body is usually 60 to 70 cm, although it can be up to 1 m. The upper surface of the body is grayish to pale brown with small dark brown spots, the under side is cream-white. This bottom-living shark can be commonly encountered on sand or gravel at depths between 1-50 m, but it occasionally reaches greater depths.

This species is quite common all around the coast of the Mediterranean Sea, where it has no marketability because not used in the population diet; despite this, every year a large number of this specimen dies because erroneously captured during fishing expeditions. The dead animals are usually thrown back to the sea, excluding the possibility to isolate chondroitin sulfate from their tissues and to use it in commercial preparations.

2.1.1 Isolation and purification of the polysaccharide from fresh cartilage

Chondroitin sulfate was purified denaturing the cartilage tissue at 65°C: in this condition the complex formed from proteoglycan-Hyaluronic acid and other components disassembles and all the proteic components are accessible to papain digestion. CS obtained after repeated ethanol precipitations was further purified via ion exchange gel chromatography, which eluted this highly charged polymer at high ionic strength (700 mM NaCl) with a final calculated yield of 1.5% w/w, with respect to the fresh tissue.

Monosaccharide composition was compatible with that expected for chondroitin polysaccharide and its linkage oligosaccharide⁷ (glucuronic acid, galactosamine, xylose and galactose).

The identity of the polysaccharide was established on the basis of its chemical composition and via extensive Nuclear Magnetic Resonance (NMR) studies, analyzing both the intact polymer and its depolymerization products.

2.1.2 NMR survey on the entire polysaccharide

The combined use of homonuclear and heteronuclear 2D-NMR has allowed the complete assignment of proton and carbon signals of CS polysaccharide (Table 2.1).

CS proton spectrum (Figure 2.2) presented broad shaped signals (in agreement with the polysaccharide nature of the molecule) localized in the region between 4.8 and 3.3 ppm, with the exception of the methyl group of the *N*-Acetyl residues resonating around 2 ppm⁸.

The occurrence of sulfate residue and their location was indirectly deduced by the displacement observed for both proton and carbon chemical shifts: this substituent induces the low field shift of geminal (up to 1 ppm) and vicinal protons (approx. 0.2 ppm); similarly, the sulfate bearing

⁷ Sasisekharan, R., Raman, R., Prabhakar, V. 2006 *Annu. Rev. Biomed. Eng.* 8:181–231.

⁸ Mucci, A., Schenetti, L., Volpi, N. 2000 *Carbohydr. Polym.* 41:37–45.

carbon chemical shift is increased of approx. 8 ppm, whereas the one of the neighbouring carbons is reduced of ca. 1-3 ppm.

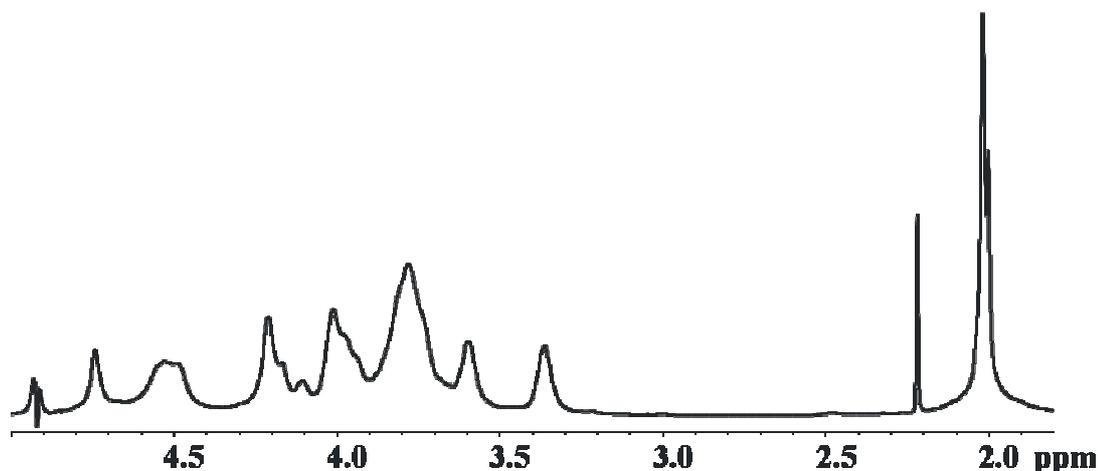


Figure 2.2: 600 MHz, 288 K, PBS 5 mM pH 7 ^1H -NMR spectrum of CS of lesser spotted dogfish

Table 2.1: (600 MHz, 288 K, PBS 5 mM pH 7) proton (plain) and carbon (italic) chemical shift attribution of the chondroitin sulfate isolate from the lesser spotted dogfish. Chemical shifts are expressed in δ relative to internal acetone (^1H at 2.225 ppm, ^{13}C at 31.5 ppm). The proton and carbon chemical shifts of methyl group of the *N*-Acetyl are 2.02 and 23.1 ppm

	1	2	3	4	5	6
4-)-β-GlcA2S-(1\rightarrow	4.74	4.11	3.81	3.86	3.78	--
U^I	<i>104.51</i>	<i>82.00</i>	<i>75.72</i>	<i>82.3</i>	<i>78.5</i>	--
4-)-β-GlcA-(1\rightarrow	4.52	3.37	3.59	3.77	3.76	--
U^{II}	<i>107.18</i>	<i>72.2</i>	<i>76.5</i>	<i>84.17</i>	<i>78.7</i>	--
4-)-β-GlcA-(1\rightarrow	4.49	3.38	3.59	3.79	3.76	--
U^{III}	<i>107.18</i>	<i>74.9</i>	<i>76.5</i>	<i>83.15</i>	<i>78.7</i>	--
3-)-β-GalNAc4S-(1\rightarrow	4.56	4.03	4.01	4.74	3.81	\sim 3.77 ⁺
N^I	<i>103.75</i>	<i>54.32</i>	<i>78.57</i>	<i>79.3</i>	<i>77.5</i>	<i>63.87</i>
3-)-β-GalNAc6S-(1\rightarrow	4.54	4.01	3.84	4.17	3.97	4.22-4.17
N^{II}	<i>103.75</i>	<i>53.63</i>	<i>82.81</i>	\sim 71*	<i>75.3</i>	<i>70.2*</i>
3-)-β-GalNAc6S-(1\rightarrow	4.54	3.94	3.94	4.22	3.97	4.22-4.17
N^{III}	<i>104.43</i>	<i>53.49</i>	<i>82.3</i>	\sim 71*	<i>75.3</i>	<i>70.2*</i>

* C-4 and C-6 signals from GalNAc6S units are partially overlapped and possess different phases in the HSQC spectrum; as a result the less intense C-4 is almost entirely cancelled from that of C-6 and its chemical shift cannot be correctly appreciated (see text and Figure 2.3).

⁺ the two H-6 signals are not resolved and merge into one signal

Therefore, CS-D motif is characterized by the GlcA H-1 and H-2 signals at 4.74 ppm and 4.11 ppm, respectively, together with the GalNAc diastereotopic H-6s at 4.22 and 4.17 ppm; it is worthy of note that these last two signals are not distinctive only of CS-D but also of CS-C motif. Finally, CS-A is identified from GalNAc H-4 proton at 4.74 ppm.

The combined use of the 2D NMR techniques, made possible the complete interpretation of the spectra recorded (Table 2.1): the occurrence of one broad peak was observed in the region between 4.49 and 4.56 ppm (Figure 2.2), its left part was composed by the different GalNAc anomeric signals, whereas the right part arose from the GlcA units.

In the case of each GalNAc residue, the COSY and TOCSY experiments led to the identification of the proton sequence from H-1 to H-4, whereas H-5 was individuated through the intra residue NOE contact with the anomeric proton; H-6s chemical shifts were deduced from the HSQC spectrum were they appeared in antiphase with respect to the other densities; additionally, each H-6 couple was assigned to a GalNAc residue on the basis of its O-4 sulfation pattern. This choice was based on Fr-2 SAX-HPLC analysis, that disclosed the occurrence of minimal amounts of CS-0S and that each GalNAc unit was sulfated either at O-4 or at O-6, but never at both positions.

Accordingly, three different types of GalNAc residues were individuated; N^I, N^{II} and N^{III} (Figure 2.2). The first one was sulfated at O-4 and belonged to the CS-A motif, the other two were sulfated at O-6 and comparison of their H-3 chemical shift with those published⁷, led their assignment to CS-C and CS-D motifs, respectively. Signals related to unsulfated GalNAc could not be detected, due to their low abundance.

Similarly, GlcA residues (except U^I unit, Figure 2.2) showed all ring proton resonances in the crowded high field region of the spectrum, in most of the cases the location of the H-5 proton was inferred thanks to the NOESY spectrum; all attributions were in agreement with literature data⁷, U^I residue was attributed to a CS-D entity whereas it was not possible to ascribe U^{II} or U^{III} to any of the chondroitin pattern, CS-A and CS-C.

Signals related to unsulfated GalNAc or IdoA could not be detected, due to their low abundance. Side by side to proton resonance attribution, HSQC spectrum assignment was performed and some difficulties were found for C-4 of N^{II} and N^{III} units, since these carbon signals coincided with those from sulfated C-6, which were very intense and with opposite phases, leading to the almost complete cancellation of the C-4 intensity.

Additionally, the HSQC spectrum was used to determine the relative proportion among disaccharide subunits. This approach is based on the assumption that the signals to be compared display similar $^1J_{\text{CH}}$ and that a difference around 5-8 Hz from the experimental set value, does not cause a substantial variation of the integrated peak volumes. More importantly, within a specific density set, such as that composed from different anomeric GlcA signals, the volumes measured scale with the $^1J_{\text{CH}}$ value used in the experiment, but their increase (or decrease) is affected to the same extent, so that the proportion between them are maintained even if their $^1J_{\text{CH}}$ is distant (i.e. 20 Hz) from the set experimental value⁹.

⁹ Guerrini, M., Naggi, A., Guglieri, S., Santarsiero, R., Torri, G. 2005 *Anal. Biochem.* 337:35–47

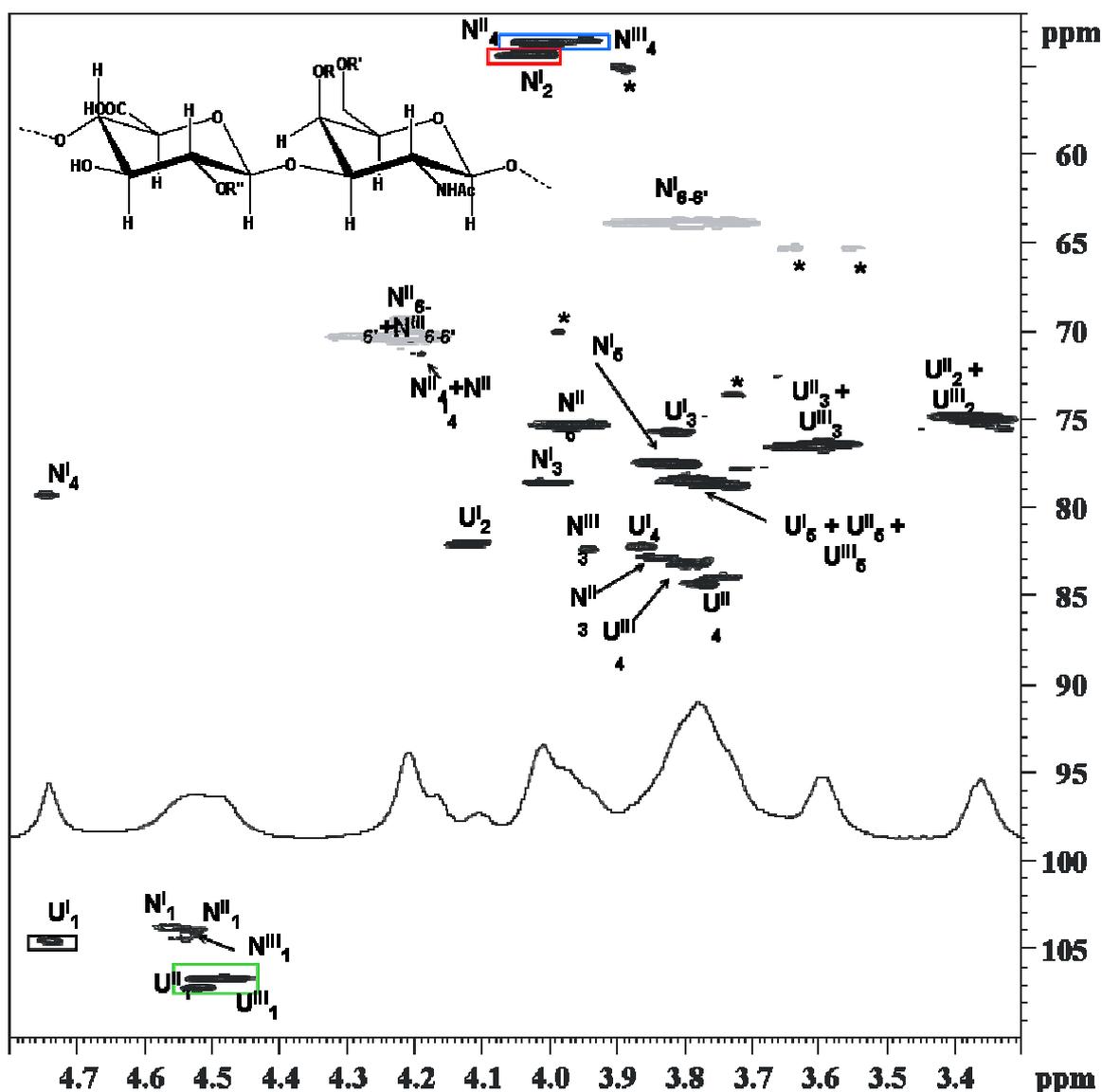


Figure 2.3: (600 MHz, 288K, PBS 5 mM pH 7) Chondroitin sulfate from lesser spotted dogfish: proton and HSQC spectra, attributions are indicated close to each density and peaks used for the integration are enclosed in a square. Density 1 (black square): GlcA H-1/C-1 cross peak, diagnostic of CS-D motif. Density 2 (green square, it arises from diverse GlcA residues): GlcA H-1/C-1 cross peak diagnostic from both CS-A and CS-C motifs. Density 3 (red square): GalNAc H-2/C-2 cross peak diagnostic of CS-A unit. Density 4 (blue square): GalNAc H-2/C-2 cross peak diagnostic of CS-C and CS-D units. Structure of the different types of chondroitin disaccharide repeating unit is shown in the inset. CS-A: R = sulfate, R' and R'' = H; CS-C: R'= sulfate, R = R''=H, CS-D: R'=R''=sulfate and R=H

Therefore, two different densities set were chosen: the first one belonging to the C-1 carbons GlcA residues, and the second one to the C-2 carbons of the GalNAc units (Figure 2.3). Integration of set 1 (Table 2.2) returned the percentage of CS-D motif (density U_1^1 in Figure 2.3) and the sum of CS-A and CS-C (density 2 in Figure 2.3); on the other hand, the percentage of CS-A was obtained analyzing the C-2 HSQC density of GalNAc N^1 residue (Table 2.2). Combining these information, the relative proportion between CS-A, CS-C and CS-D were obtained and were 45.6, 37.9 and 16.5, respectively. Signals related to the not sulfated CS-0S unit, were not detected in the spectrum and no estimation of this motif was possible at this stage.

These data indicate the strong prevalence of sulfation at position O-6 of the GalNAc residue, resulting from the sum of the CS-C and CS-D units, but the amount of the CS-A portion appears higher with respect to that reported for the shark cartilage (45.6% versus 27.7%)⁷.

Table 2.2: Percentages of CS-A, CS-C, CS-D and CS-0S motifs estimated via NMR and SAX chromatography. For the integration in the HSQC spectrum, the densities belonging to the anomeric protons of the GlcA units (U₁) and those of the protons in the position two of the GalNAc ones (N₂) were selected.

	CS-A	CS-C	CS-D	CS-0S
Density(U₁)	83.5		16.5	--
NMR Density(N₂)	45.6	54.4		--
composition	45.6	37.9*	16.5	--
HPLC	41	32	19.8	8.3

* The percentage of the CS-C motif was estimated by difference knowing the values for the CS-A and CS-D ones.

2.1.3 Enzymatic degradation and chromatographic separations

An amount of purified chondroitin sulfate of the lesser spotted dogfish underwent an enzymatic depolymerisation with ABC lyase and a treatment with trypsin to remove the residual proteic contaminants. The fragments obtained were partially purified via SEC chromatography on Bio-Gel P2, yielding to two fractions: one less abundant and less retained from the column named Fr-1 (9 mg), and Fr-2 (47 mg). This last sample was a mixture of disaccharides capped with one unit of 4-deoxy-β-L-threo-hex-4-enosyl-uronate (ΔHexA) at the not reducing end. The composition of this fraction was disclosed via both strong ion exchange chromatography (SAX) and Reverse Phase Ion-Pairing (RPIP) HPLC analysis. The other fraction, Fr-1, contained the oligomers of the linkage region, which were separated via RPIP HPLC.

2.1.3.1 Strong Anionic Exchange chromatography (SAX-HPLC) of the disaccharide mixture

The disaccharide mapping of the Fr-2 was achieved applying the conditions found in literature¹⁰. The chromatographic profile of this sample was obtained monitoring the eluate at 232 nm (Figure 2.4), and the occurrence of four, baseline separated peaks was observed. These four peaks were assigned comparing their retention time with that of commercial standards and were assigned as: ΔHexA-GalNAc (CS-0S, 12.9 min), ΔHexA-GalNAc6S (CS-C, 22.8 min), ΔHexA-GalNAc4S (24.6 min, CS-A), and ΔHexA2S-GalNAc6S (CS-D, 38 min).

Their elution order agreed with reported data⁹, and peak integration disclosed the following composition: CS-0S 8.2%, CS-A 41%, CS-C 32%, and CS-D 19.8% (Table 2.2).

¹⁰ Volpi, N., 2004 *Carbohydr. Polym.* 55:273–281.

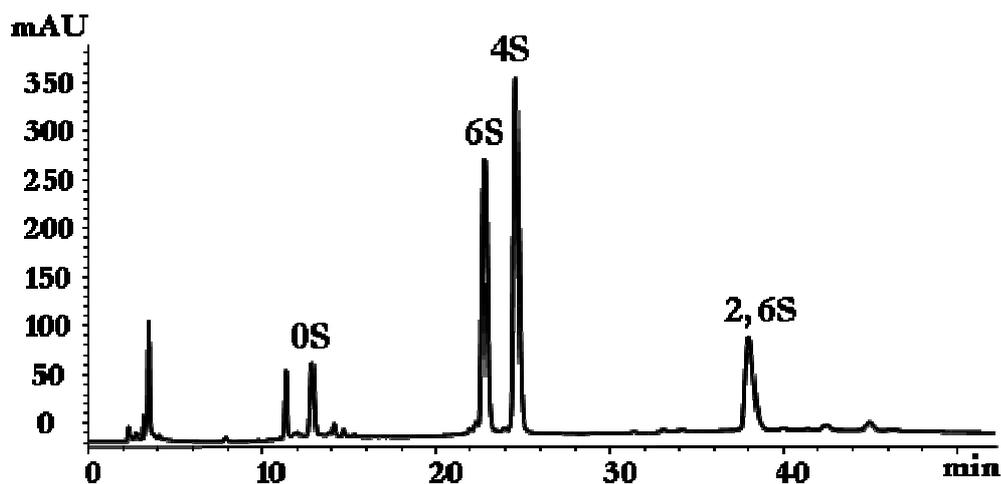


Figure 2.4: Chromatogram of the SAX-HPLC separation of the Frz.2 ($\lambda=232$ nm)

2.1.3.2 Reverse-phase ion pairing chromatography (RPIP-HPLC) of the disaccharide mixture

This separation technique has been successfully applied so far in the analysis of highly charged heparin¹¹, hyaluronic¹² and unsulfated chondroitin oligosaccharides¹³. In contrast with the eluents used in SAX chromatography, the mobile phase contains volatile ion-pairing reagents that provide a good resolution of the analytes allowing to interface the HPLC system with a MS detector.

The separation of the disaccharide mixture found in Fr-2, was performed and led to the isolation of three components: Δ HexA-GalNAc6S (RT 12.5 min), Δ HexA-GalNAc4S (RT 13.2 min) and Δ HexA2S-GalNAc6S (RT 18.8 min).

The elution order of these three oligosaccharides appeared to be similar to that found for SAX chromatography, their spectroscopical attribution is reported in Table 2.4 and, despite the presence of tributylamine (TBA) as counterion, their proton and carbon chemical shifts were similar to that from literature data¹⁴.

2.1.3.3 Reverse-phase ion pairing chromatography of glycopeptides mixture belonging to the linkage region

RPIP-HPLC separation of Fr-1 led to the isolation of two glycopeptides belonging to the linkage region of the CS chains: Fr-1A (RT 5.1 min.) and fr-1B (RT 6.8 min); the two glycopeptides differed only for the length of the chain, being the Fr-1A made up of six monosaccharides, and the other of eight residues, as shown by their proton spectra (Figure 2.5).

¹¹ Thanawiroon, C., Rice, K. G., Toida, T., Linhardt, R. J. 2004 *J. Biol. Chem.* 279:2608-2615.

¹² Volpi, N. 2007 *Anal. Chem.* 79:6390-6397.

¹³ Volpi, N., Zhang, Z., Linhardt, R.J. 2008 *Rapid Commun. Mass. Spectrom.* 22:3526-3530.

¹⁴ Yamada, S., Yoshida, K., Sugiura, M., Sugahara, K. 1992 *J. Biochem.* 112:440-447.

The eluents and the elution conditions were modified with respect to those used for the disaccharide mixture, in order to obtain an appropriate separation. The chromatographic separation was monitored again at 232 nm, due to the presence at the non-reducing end of an Δ HexA unit as in the disaccharides.

2.1.4 NMR analysis of the molecules isolated through the RPIP-HPLC

The ^1H -NMR spectra of the three disaccharides are shown in Figure 2.4. All the signals belonging to the sugar components are located between 5.8ppm, and 3.5ppm; no signals belonging to the ion-pairing reagent are found in that region.

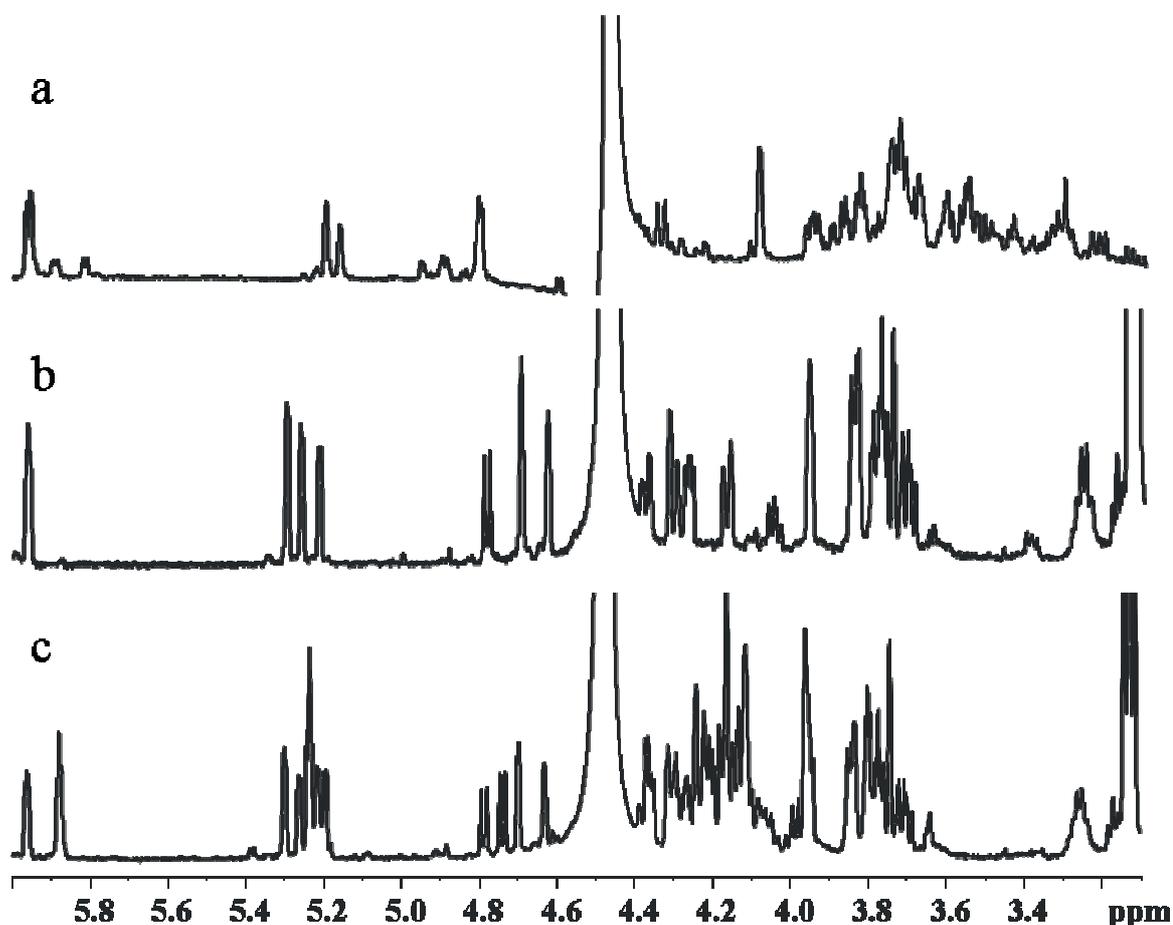


Figure 2.5: (600 MHz, D_2O) ^1H -NMR spectra of the unsaturated disaccharides of the Fr.2: a) Δ hex2S-(1 \rightarrow 3)-GalNAc6S, b) Δ hex-(1 \rightarrow 3)-GalNAc4S, c) mix of Δ hex-(1 \rightarrow 3)-GalNAc6S and Δ hex-(1 \rightarrow 3)-GalNAc4S.

The first peak component isolated by RPIP-HPLC from the Fr.2 (c, Fig 2.4) was a mix of two disaccharides Δ HexA-GalNAc6S and Δ HexA-GalNAc4S, but the high resolution of the spectra and the availability of the spectroscopic for the component Δ HexA-GalNAc4S alone (b, Fig.2.4), allowed an accurate chemical shift assignment for the component Δ HexA-GalNAc6S as well.

The spectra of the three different CS disaccharides show close similarities: in the down field region we found the protons belonging to the α reducing end, the anomeric protons of the

Δ HexA unit, and the proton of the Δ HexA units linked to the sp^2 carbon in the position 4 (two different signals can be detected for the anomeric proton of this residue due to the different behavior of the reducing end). For all the disaccharides the signal of the *N*-Acetyl group sets around 2 ppm, and the protons in the position 4 of the GalNAc residues are clearly identifiable around 4.6 ppm. The chemical shifts values for the protons of both residues, however, vary significantly depending on the sulfation pattern, and on the fashion of the reducing end. The proton chemical shifts for the Δ HexA unit, obviously, change a lot when this ring is *O*-sulfated; the presence of the sulfate ester causes the values of δ of the signals of the anomeric, and the proton in the position 2 to shift to higher values. The carbon chemical shifts show a similar behavior; both the sulfation, and the glycosilation led to the collocation of the correspondent carbons at high values of δ^{15} . The different sulfation pattern influenced the coupling constant of the unsaturated residue too¹⁶.

In all the cases the chemical shifts of the protons were assigned through the combined study of the omonuclear NMR spectra, COSY, TOCSY, and NOESY. The proton signals of the Δ HexA were easily identified through the clear COSY, and TOCSY correlations. The protons from 1 to 4 for the GalNAc residues were individuated through the scalar couplings present in the COSY, and TOCSY spectra. The protons 5, and 6 were detected on the other end using the dipolar couplings with the proton in the position 4. The carbon chemical shifts were assigned through the HSQC and HMBC spectra. The proton chemical shift assignments agree with the literature data^{8,9}, the carbon assignments are slowly different from those reported, probably due to the presence of the counterion belonging to the ion-pairing reagent. All the chemical shift assignments are reported in the Table 2.3.

¹⁵ Yamada, S., Yoshida, K., Sugiura, M., Sugahara, K. 1992 *J. Biochem.* 112:440-447.

¹⁶ Ragazzi, M., Ferro, D.R., Provasoli, A., et al. 1993 *J. Carbohydr. Chem.* 12:523-535.

Table 2.3: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the three unsaturated oligosaccharide isolated via RPIP-HPLC chromatography.

ΔHexA-(1\rightarrow3)-GalNAc4S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.3	3.84	3.96	5.96	--	--
	<i>101.1</i>	<i>69.9</i>	<i>65.9</i>	<i>107.74</i>	<i>144.3</i>	<i>169.3</i>
t-ΔHexA-(1\rightarrow	5.26	3.85	3.96	5.96	--	--
	<i>101.1</i>	<i>69.9</i>	<i>65.9</i>	<i>107.7</i>	<i>144.3</i>	<i>169.3</i>
3)-α-GalNAc4S(1\rightarrow	5.21	4.38	4.31	4.70	4.27	3.70-3.78
	<i>92.4</i>	<i>50.9</i>	<i>73.9</i>	<i>78.4</i>	<i>71.6</i>	<i>62.3</i>
3)-β-GalNAc4S(1\rightarrow	4.78	4.05	4.17	4.63	3.84	3.74-3.77
	<i>96.1</i>	<i>54.9</i>	<i>76.6</i>	<i>77.3</i>	<i>75.8</i>	<i>62.4</i>
ΔHexA-(1\rightarrow3)-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.24	3.79	4.11	5.88	--	--
	<i>102.3</i>	<i>70.8</i>	<i>67.1</i>	<i>108.4</i>	<i>144.9</i>	<i>169.40</i>
t-ΔHexA-(1\rightarrow	5.19	3.80	4.11	5.88	--	--
	<i>102.3</i>	<i>70.8</i>	<i>67.1</i>	<i>108.4</i>	<i>144.9</i>	<i>169.40</i>
3)-α-GalNAc6S(1\rightarrow	5.23	4.30	4.12	4.22	4.36	4.15-4.21
	<i>92.5</i>	<i>50.1</i>	<i>77.9</i>	<i>69.4</i>	<i>69.62</i>	<i>69.0</i>
3)-β-GalNAc6S(1\rightarrow	4.74	3.99	3.95	4.18	3.95	4.18-4.21
	<i>96.3</i>	<i>53.</i>	<i>80.7</i>	<i>69.06</i>	<i>74.1</i>	<i>69.0</i>
ΔHexA2S-(1\rightarrow3)-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA2S(1\rightarrow	5.71	4.60	4.34	6.46	--	--
	<i>98.0</i>	<i>74.1</i>	<i>61.8</i>	<i>110.8</i>	--	--
t-ΔHexA2S(1\rightarrow	5.69	4.60	4.34	6.46	--	--
	<i>98.1</i>	<i>74.1</i>	<i>61.8</i>	<i>110.8</i>	--	--
3)-α-GalNAc6S(1\rightarrow	5.31	4.40	4.25	4.2	4.46	4.20-4.25
	<i>91.2</i>	<i>48.6</i>	<i>77.4</i>	<i>67.3</i>	<i>68.2</i>	<i>67.2</i>
3)-β-GalNAc6S(1\rightarrow	4.85	4.11	4.06	4.25	4.05	4.20-4.25
	<i>94.9</i>	<i>52.2</i>	<i>80.3</i>	<i>67.3</i>	<i>72.38</i>	<i>67.2</i>

The two oligosaccharides contained the canonical sequence of the linkage region of the CS-proteoglycans, and were devoid of any inorganic substituent, like sulfate or phosphate.

Proton spectrum (Fig 2.5) of Fr-1A showed the H-4 and H-1 signals of the non-reducing unit, a Δ HexA residue (VI), at 5.90 and 5.18 ppm, respectively; the anomeric signals of the other five residues were found in the region between 4.7 and 4.4 ppm, and were assigned to: one xylose (Xyl, 4.46 ppm, I), two Galactose (Gal, 4.66 (I) and 4.53 ppm (II)), one GalNAc (4.53 ppm, V), and one GlcA (4.66 ppm, IV) residues, all in the β configuration on the basis of their chemical shift and of their the $^3J_{H1,H2}$ coupling constants (ca. 8 Hz). The only two signals belonging to a Δ HexA in the region between 6 and 5.3 ppm confirmed the presence at the reducing end of a non-sugar substituent, then identified as a serine unit¹⁷.

Fr-1A

Δ HexA-(1 \rightarrow 3)-GalNAc-(1 \rightarrow 4)-GlcA-(1 \rightarrow 3)-Gal-(1 \rightarrow 3)-Gal-(1 \rightarrow 4)-Xyl-(1 \rightarrow)Ser

VI V IV III II I

Fr-1B

Δ HexA-(1 \rightarrow 3)-GalNAc-(1 \rightarrow 4)-GlcA-(1 \rightarrow 3)-GalNAc-(1 \rightarrow 4)-GlcA-(1 \rightarrow 3)-Gal-(1 \rightarrow 3)-Gal-(1 \rightarrow 4)-Xyl-(1 \rightarrow)Ser

VIII' VII' VI' V' IV' III' II' I'

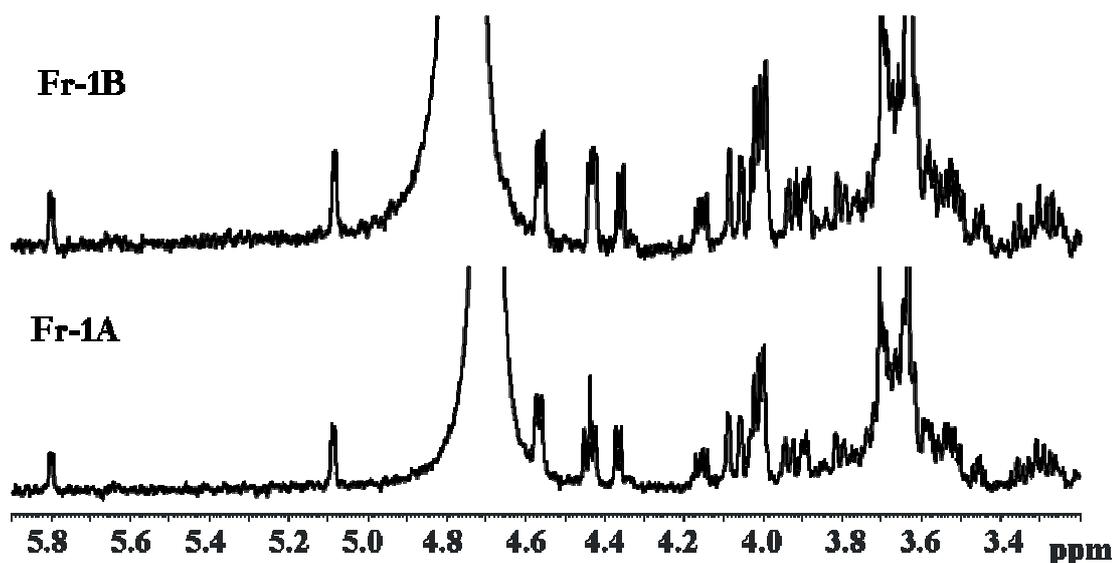


Figure 2.6: (600 MHz, 290 K, D₂O) ¹H-NMR spectra and sequences of the glycopeptides belonging to the linkage region

The 2D spectra collected were interpreted according to the general strategy adopted for CS polysaccharide, leading to the complete attribution of all the signals (Table 2.4) and to establish the sequence between the different units.

In particular, for both glycopeptides, the protons from H1 to H4 of the sugar with the *galacto* configuration were identified through scalar coupling, while the proton H5, and the two of the

¹⁷ Sugahara, K., Yamashin, I., De Waardll, P., Van HalbeekII, H., Vliegthart, J.F.G. 1988 *J. Biol. Chem.* 263:10168-10174.

CH₂OH groups, were found through the NOE contacts with the proton in the position 4. The protons belonging to the GlcA, ΔHexA, and the Xyl units, were easily detected due to the strong, and clear COSY, and TOCSY correlations. The chemical shift assignment of the three protons belonging to the serine residue was easily performed on the basis of the strong scalar couplings in the COSY spectrum. The sequence of the two molecules were determined on the basis of dipolar coupling in the NOESY spectrum, and in the case of Fr.1A also with the α-glycosylation shifts estimated for the five sugars through the study of the HSQC.

As a result, Fr-1A contained the canonical sequence of the linkage tetrasaccharide elongated with one CS repeating unit at its non-reducing end (Fig 2.5).

Although this molecule was studied as TBE salt form, its proton chemical shift agreed with the published data¹⁶, additionally, in this case it was possible to determine the carbon chemical shifts as well, so far reported only for the alditol form of the molecule.

Table 2.4: (600 MHz, 293 K, D₂O) proton (plain) and carbon (italic) chemical shift attribution of Fr-1A, analyzed in its salt form with TBA as counterion. The HMBC experiment wasn't performed, so the carbon chemical shifts of the quaternary carbons were not reported.

	1	2	3	4	5	6
t-ΔHexA-(1→	5.18	3.79	4.09	5.90	--	--
VI	<i>102.15</i>	<i>70.8</i>	<i>66.9</i>	<i>108.6</i>	--	--
4-)-β-GlcA-(1→	4.66	3.45	3.62	3.77	3.71	--
IV	<i>105.2</i>	<i>74.2</i>	<i>75.0</i>	<i>80.9</i>	<i>77.6</i>	--
3-)-β-Gal-(1→	4.66	3.74	3.80	4.15	3.68*	3.77*
III	<i>105.2</i>	<i>71.3</i>	<i>83.9</i>	<i>69.5</i>	<i>76.2*</i>	<i>62.41*</i>
3-)-β-Gal-(1→	4.53	3.67	3.82	4.19	3.73*	3.77*
II	<i>102.5</i>	<i>71.2</i>	<i>83.4</i>	<i>69.8</i>	<i>76.3*</i>	<i>62.41*</i>
3)-β-GalNAc-(1→	4.53	4.00	3.90	4.09	3.73*	3.77*
V	<i>102.5</i>	<i>52.4</i>	<i>81.2</i>	<i>68.8</i>	<i>76.3*</i>	<i>62.41*</i>
3)-β-Xyl-(1→	4.46	3.37	3.61	3.87	4.11-3.40	--
I	<i>103.9</i>	<i>73.9</i>	<i>75.0</i>	<i>77.7</i>	<i>64.3</i>	--
	Hα/Cα	Hβ/Cβ				
Ser	3.99	4.02-4.26	--	--	--	--
	<i>55.84</i>	<i>69.2</i>	--	--	--	--

* The proton and the carbon chemical shift values are the same for both the Gal units and for the GalNAc one, due to the overlapping of the corresponding H6,6' signals.

In contrast with proton spectrum of Fr-1A, that recorded for Fr-1B (Figure 2.5) showed a more crowded carbinolic region, due to the presence of two additional sugar units, identified as a GlcA and a GalNAc moiety.

Due to the low abundance of this compound, it was possible to attribute its proton chemical shift (Table 2.5) and the sequence among the residues, resulting on the structure reported in Figure 2.5.

Table 2.5: (600 MHz, 293 K, D₂O) proton chemical shift attribution of Fr-1B, analyzed in its salt form with TBA as counterion.

	1	2	3	4	5	6
<i>t</i>-ΔHexA-(1→ VIII'	5.18	3.79	4.09	5.90	--	--
4-)-β-GlcA-(1→ IV'	4.66	3.45	3.63	3.77	3.71	--
3-)-β-Gal-(1→ III'	4.66	3.74	3.80	4.12	3.69	*
3-)-β-Gal-(1→ II'	4.53	3.67	3.83	4.19	3.73	*
3)-β-GalNAc-(1→ V'	4.53	4.00	3.90	4.10	3.72	*
3)-β-GalNAc-(1→VII'	4.50	3.99	3.80	4.16	3.69	*
4-)-β-GlcA-(1→ VI'	4.48	3.37	3.57	3.74	3.69	--
3)-β-Xyl-(1→ I'	4.46	3.37	3.61	3.87	4.12-3.40	--
	Hα/Cα		Hβ/Cβ			
Ser	3.98		4.02-4.26			

* the signals of the CH₂OH protons of the Gal and GalNAc units are localized in the crowded region between 3.7 and 3.8 ppm.

2.2 The chondroitin sulfate of ray and skate specimens

Rays and skates are dorsoventrally flattened fishes that are closely related to sharks. Rays belong to three scientific orders - *Pristiiformes*, *Myliobatiformes*, and *Torpediniformes* while skates are classified in the order *Rajiformes*. There are approximately 500 different species of rays and skates that are divided into 18 families. The main difference between rays and skates lies in their reproductive strategies. Rays are live bearing (ovoviviparous) while skates are egg laying (oviparous), releasing their eggs in hard rectangular cases sometimes called "mermaid's purses". Moreover, skates typically have a prominent dorsal fin, while the dorsal fin is absent or greatly reduced in rays. Most rays are kite-shaped with whip-like tails, possessing one or two stinging spines while skates have fleshier tails and lack spines. Rays protect themselves from predators with these stinging

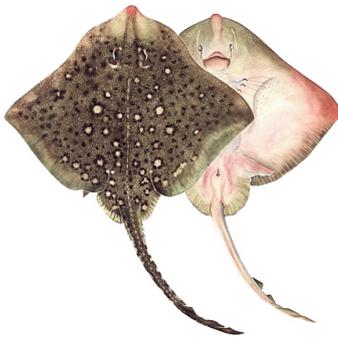


Figure 2.7: *Raja brachyura*

spines or barbs while skates rely on thorny projections on their backs and tails.

Skates have small teeth while rays have plate-like teeth adapted for crushing prey. Another difference is that rays are generally larger than skates.

The common feature is: rays and skates are not considered quality items for human consumption.

Raja brachyuran (Fig. 2.6), also called blond skate, is a shallow water species living on sandy and sand-rock bottoms². It is well distributed throughout European waters, generally at depths up to 100 metres. This fish has a relatively short snout, and corners of pectoral fins almost right-angled. In adult fishes the whole back area may be prickled, while younger fishes are less spotted. The specimens have large eyes, while the tail bears a row of large spines centrally. The back is usually brown with creamy white blotches, and dense dark spots which extend to the edge of the fins, while the ventral one is white.



Figure 2.8: *Torpedo nobiliana*

Torpedo nobiliana (Fig. 2.7) is a bottom living fish, characterized by a disk roughly subcircular, truncated in front, and somewhat broader than long. The eyes are very small and set far forward. The two dorsal fins, the first of which is larger, stand on the forward end of the tail, the second, indeed, partly above the bases of the pelvic fins. They are separated by an interspace nearly as long as the second dorsal fin. The tail fin is of ordinary fish form, triangular and nearly as long as deep. The tail is shorter than in the skates, it occupies only about two-fifths of the total length of the fish, measured from the cloaca.

The teeth are small, with sharp curved points, and are in about 60 series, with up to 7 rows

exposed and functioning at one time. The colour is dark chocolate to purplish brown above, some with a few obscure darker spots; lower surface white except that the edges of disk, fins, and tail are of the same dark tint as the upper side. Adult torpedoes are usually 2 to 5 feet long or a little longer, and heavy for their size. The most interesting thing about the torpedo is its ability to give electric shocks of considerable strength to anyone touching it.

Torpedo has of no commercial value nowadays, but it is an object of interest to the workers in biological laboratories because of its electric discharges.

2.2.1 Isolation and purification of chondroitin sulfate from the cartilage and the skin of *Raja brachyura* and from the cartilage of *Torpedo nobiliana*

Chondroitin sulfate was isolated from the skin and the cartilaginous body of an exemplar of a *Raja brachyura* (1 adult specimen of approx. 1 kg) and from the cartilage of three *Torpedo nobiliana* specimens (approx. 300 g each) with the same procedure adopted for the cartilage of lesser spotted dogfish.

After the purification with anionic exchange chromatography, the yield of the purified polysaccharide materials with respect to the fresh tissue was 3% for the CS of the cartilage of skate, 2% for the CS of the skin of skate and 3,5% for the CS isolated from the cartilage of torpedo.

In all the cases CS was recovered in the fractions eluted at high ionic strength (NaCl 1M). This aspect indicated a high charge density for all the sulfated polysaccharides.

All the CS samples underwent a compositional analysis and for all the typical composition was estimated: high content of GlcA and GalNAc and just a little amount of Xyl and Gal. Atypical components like fucose, glucose or mannose were not detected, so the presence of rare repeating units was not considered¹⁸. These data were also confirmed by a ¹H-NMR analysis.

The samples were also analyzed through SDS-PAGE, to have an idea of their molecular weight. The results of the staining with alcian blue dye (1, Fig 2.8) and through silver staining (2, Fig 2.8) are shown in Figure 2.8. The molecular weight was similar for all the skate and torpedo samples (lines a-e), but also similar to the one of the CS isolated from the lesser spotted dogfish in line f. Moreover, all the samples have a molecular weight higher than 20 kDa, as indicated by the comparison with the standard of sulfate dextrane (MW 9-20 kDa) in line g.

¹⁸ Malavaki, C., Mizumoto, S., Karamanos, N., Sugahara, K. 2008 *Connective Tissue Research* 49:133-139.

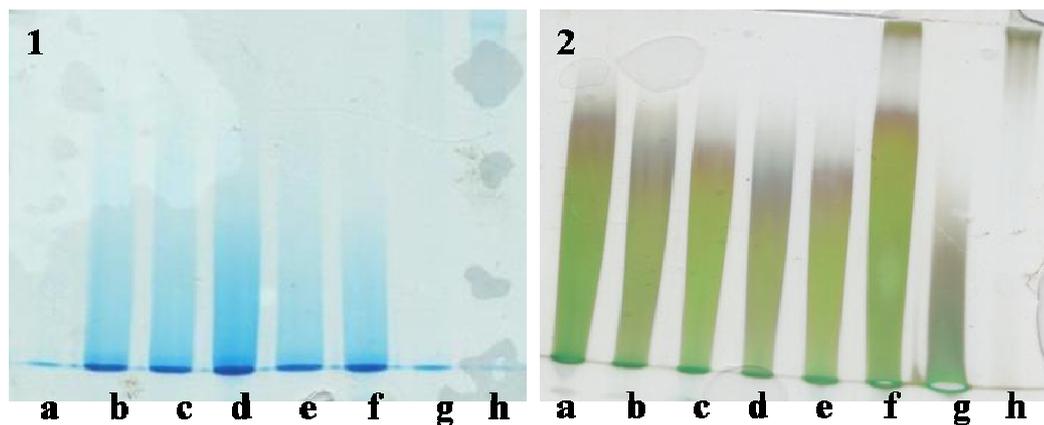


Figure 2.9: SDS-PAGE 8% stained with acian blue (1) and silver staining (2). Line a: CS cartilage skate, Lines b, d: CS skin skate, Lines c, e: CS cartilage torpedo, Line f: CS lesser spotted dogfish, Line g: dextran sulfate 9-20 kDa, Line h: dextran sulfate 500 kDa.

The CS material was then depolymerised with lyase type enzymes to perform the classical disaccharide mapping.

2.2.2 Enzymatic degradation and HPLC-SAX separation of the disaccharide mixture belonging to different chains

An amount of each CS sample (87 mg of CS from the skin of skate, 11 mg of CS from the cartilage of skate, 28 mg of the CS from the cartilage of torpedo) was depolymerised with ABC lyase and the corresponding oligosaccharide mixture were fractionated on a Biogel P2, as reported for the CS of the lesser spotted dogfish. For the CS of both the skin and the cartilage of the skate, again, two main fractions were isolated, one in the void volume of the column containing oligosaccharides larger than 1800 Da and glycopeptides belonging to the linkage region and a second fraction containing the disaccharide mixture.

The chromatographic separation of the digested CS material belonging to torpedo, instead, has led to recover three different fractions, the first like above, in the void volume of the column, and two distinct disaccharide fractions, the second containing the disulfated unsaturated disaccharides $\Delta\text{HexA}2\text{S}-(1\rightarrow3)\text{-GalNAc}6\text{S}$ and the last containing a mixture of monosulfated unsaturated disaccharides $\Delta\text{HexA}-(1\rightarrow3)\text{-GalNAc}4\text{S}$ and $\Delta\text{HexA}-(1\rightarrow3)\text{-GalNAc}6\text{S}$.

The exact percentage of each disaccharide motif in the different CS samples was determined through strong-anionic HPLC (SAX-HPLC) analyses. The chromatograms of the SAX-HPLC separation are shown in Figure 2.9.

In the case of fractions belonging to the CS of torpedo, each Biogel P2 fraction was analyzed by SAX-HPLC to establish its purity. Disaccharide composition was established mixing aliquots of the chromatographic fraction respecting their relative abundance after the SEC separation.

The results of the SAX-HPLC analyses are listed in Table 2.6.

Table 2.6: Disaccharide mapping of the three CS samples

Specimen	CS $0S$	CS- A	CS- C	CS- D
<i>Raja bochyura</i> cartilage	9.5 %	36.5 %	39.0 %	15 %
<i>Raja bochyura</i> skin	5.7 %	37.1 %	38.4 %	18.8 %
<i>Torpedo nobiliana</i> cartilage	0.9 %	30.9 %	38.2 %	30 %

Disaccharide mapping confirmed the high-density charge of the samples due the low percentage of unsulfated unsaturated disaccharide and higher percentage of disulfated unsaturated motifs with the respect to the CS of the lesser spotted dogfish (see 2.1.2).

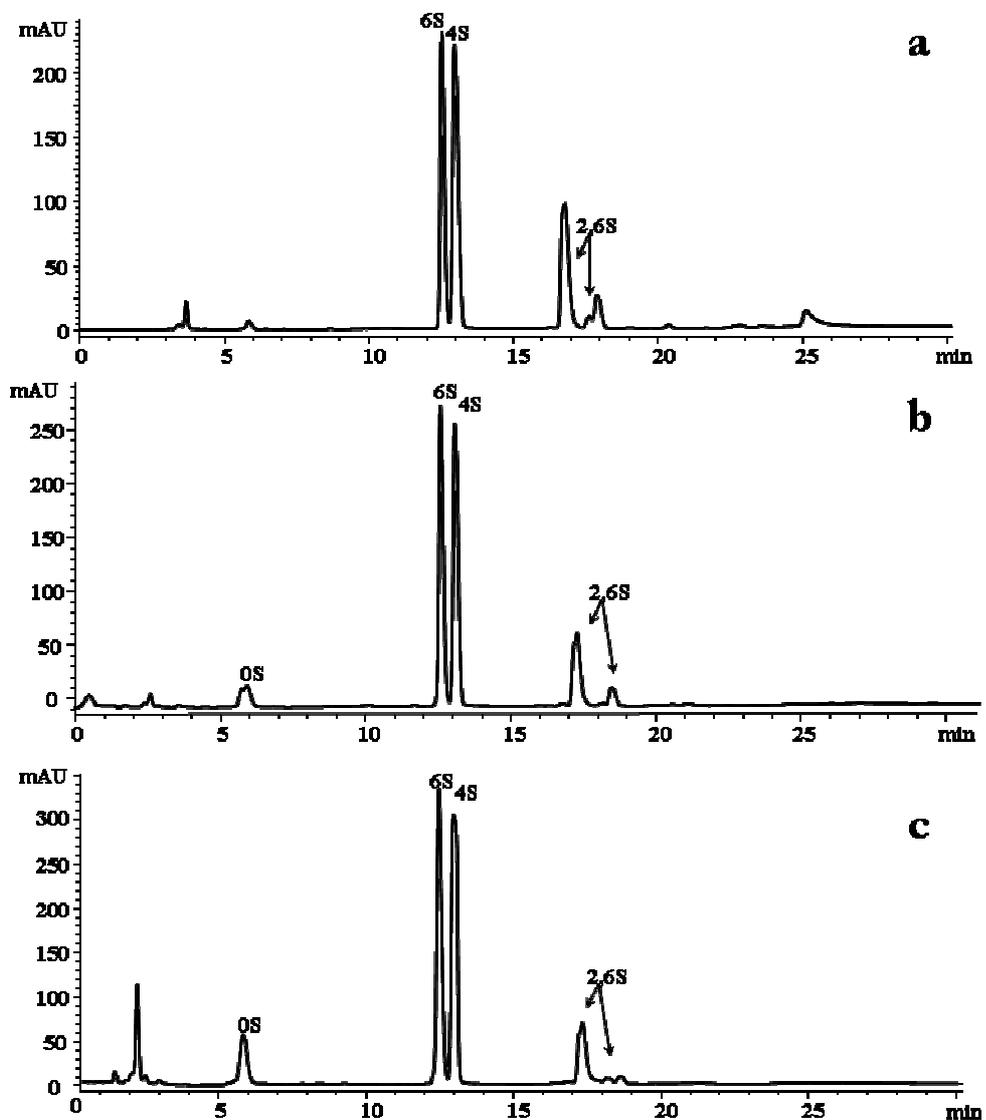


Figure 2.10: Chromatographic profiles of the SAX-HPLC separations ($\lambda=232$ nm) of disaccharide mixtures belonging to ABC lyase treatment of CS from cartilage of skate (c), of the CS from skin of skate (b), of the CS from cartilage of torpedo (a).

2.2.3 Estimation of the content of Iduronic acid

The CS samples isolated from skate and torpedo underwent also chemical and enzymatical analyses to estimate the content of iduronic acid, due to the possibility of finding hybrid chains of CS/DS¹⁹.

The determination of chemical composition with acetylated O-methyl-glycoside (MGA) method²⁰ was the first and qualitative method applied. Starting from a sample of dermatan sulfate a standard of iduronic acid was prepared²¹. Through the comparison with the standard, the presence of a little amount of IdoA was individuated just in the sample belonging to the CS of skin of skate.

A second and more accurate procedure, with the use of enzymes, was applied too¹⁵. Three different enzymatic digestions were performed at the same time on an amount of each sample. Depolymerisations were performed using three different CS lyases: ABC lyase, AC lyase and B lyase. The depolymerisations were followed spectrophotometrically reading the absorbance at 232 nm.

This approach leads to a more and efficient estimation of the iduronic content, due to the high substrate specificity of each enzyme²². In particular the quantification of DS is usually carried out on the basis of the difference in the analytical values of chondroitinases ABC and AC I digestions, since DS is not depolymerised by AC I lyase. The resulting data, are then confirmed by the analysis of the products obtained through B lyase action that cleaves just the linkage between GalNAc and IdoA²³.

For the digestions with ABC lyase and AC lyase an increment of absorption during the time was detected. In particular, the absorbance of the samples belonging to the ABC digestion were every time higher with respect to those of the AC digestion, more over the maximum of absorbance of the ABC digestions was reached after 3h, differently from the AC lyase digestions.

For the digestions with the B lyase, no variations were detected over 24 hours, indicating that in all the CS samples there were no CS-B motifs to be recognized. The comparison between the absorbance of the AC and ABC lyase depolymerised products obtained after 16 h, indicated that the amount of unsaturated species generated with the AC digestion was around the 60-70% of that produced by ABC lyase. The not complete depolymerisation confirmed the specificity of AC I lyase for CS-A and CS-C motifs and not for CS-D one.

¹⁹ Nandini, C.D., Itoh, N., Sugahara, K. 2005 *J.B.C.* 280(6):4058-4069

²⁰ Holst, O., 2000 *Methods in Molecular Biology Humana*, Totowa, 145:345–353.

²¹ Sudo, M., Sato, K., Chaidedgumjorn, A., Toyoda, H., Toida, T., Imanari, T. 2001 *Anal. Biochem.* 297:42–51.

²² Linhardt RJ, Avci FY, Toida T, Kim YS, Cygler M. 2006 *Advances in Pharmacology* 53:187-215

²³ Imanari, T., Toida, T., Koshiishi, I., Toyoda, H. 1996 *Journal of Chromatography A* 720:275-293.

After 24 hours the enzymes were denaturated and all the samples belonging to ABC and AC lyase degradations were fractionated through a SEC chromatography on the Biogel P2. The corresponding fractions were, then, analyzed by NMR and SAX-HPLC.

The samples belonging to the digestions with B lyase were directly analyzed by NMR and the presence of intact polysaccharides was detected.

2.2.4 Analysis of the AC lyase digestion products

The chromatographic purification of the AC lyase products, on a BioGel P2, has led in the three cases, to the isolation of 5 fractions: one in the void volume of the column (Fr-A), containing sulfated oligosaccharides and glycopeptides of the linkage region, a second fraction (Fr-B) made up by unsaturated monosulfated disaccharides, a third fraction (Fr-C) containing unsaturated unsulfated disaccharides, a fourth fraction (Fr-D) made up by contaminants belonging to the enzyme and a last one (Fr-E) containing a reducing monosaccharide (approx. 1 mg).

The latest fraction underwent a T.L.C analysis and the comparison with standards has led to the classification of the sugar as unsulfated GalNAc. This monosaccharide was not classified as a degradation product of unsaturated disaccharides belonging to rare motifs containing a GlcA3S²⁴ (CS-L: [4]-GlcA3S-(1→3)-GalNAc6S-(1-), CS-K: [4]-GlcA3S-(1→3)-GalNAc4S-(1-) and CS-M: [4]-GlcA3S-(1→3)-GalNAc4S,6S-(1-)]²⁵, due to the absence of sulfate group. Its production was explained later on the basis of the data collected for the oligosaccharides of the fractions A, that highlighted the presence of enzymatic contaminants (sulfatases and exoglycosidases) in the starting AC lyase preparation.

Fractions B and C were analyzed through SAX-HPLC in order to established their exact composition. The results are listed in Table 2.7.

²⁴ Fongmoon, D., Shetty, A.K., Basappa, Yamada, S., Sugiura, M., Kongtawelert, P., Sugahara, K. 2007 *J.B.C.* 282(51):36895-36904.

²⁵ Malavaki, C., Mizumoto, S., Karamanos, N., Sugahara, K. 2008 *Connective Tissue Research* 49:133-139.

Table 2.7: Composition of the Fr-B and Fr-C belonging of the AC lyase digestions of the three CS samples

	<i>CS0S</i>	<i>CS-A</i>	<i>CS-C</i>	<i>CS-D</i>
Fr-B				
<i>Raja bochyura</i> cartilage	--	7.1 %	92.0 %	--
<i>Raja bochyura</i> skin	--	13.2 %	86.8 %	--
<i>Torpedo nobiliana</i> cartilage	--	4.7 %	95.3 %	--
Fr-C				
<i>Raja bochyura</i> cartilage	92.5 %	7.5 %	7.5 %	--
<i>Raja bochyura</i> skin	88.3 %	2.4 %	9.3 %	--
<i>Torpedo nobiliana</i> cartilage	77.6 %	12.2 %	10.2 %	--

These analyses established that the AC lyase digestion liberated, as larger amounts, just Δ hex-GalNAc and Δ hex-GalNAc6S disaccharides and traces of the other two species (Δ hex-GalNAc4S and Δ hex2S-GalNAc6S).

Looking at these data and taking in account those available about the ABC lyase products (see 2.2.2), the absence of disulfated motifs was explained considering AC lyase specificity for the linkages β -GalNAc4S/6S-(1 \rightarrow 4)-GlcA.

To explain the presence of Δ Hex-GalNAc4S, just in trace, two hypothesis were formulated: one regarding the involvement of this motif in sequences resistant to the action of AC lyase and another supposing that the AC lyase used for the depolymerisation was contaminated by a 4-sulfoesterase, datum supported also by the large CS-0S amount isolated and not detected in the previous ABC digestions of the same product.

In order to validate one of the two hypotheses, a deep analysis through SAX-HPLC was performed on each Fr-A and the enzymatic activity of AC lyase was tested on reference compounds.

Analysis of fractions A through SAX-HPLC

The two fractions A obtained from the skate and torpedo cartilage had a very similar composition, and were separated applying the same NaCl gradient (Fig. 2.11). Differently, the fraction derived from the skate skin CS contained more sulfated structures, so it was resolved with a different gradient (Fig 2.12).

The main components of the different fractions A were isolated on the basis of the chromatographic profiles and analyzed by NMR, in order to establish their composition and structure.

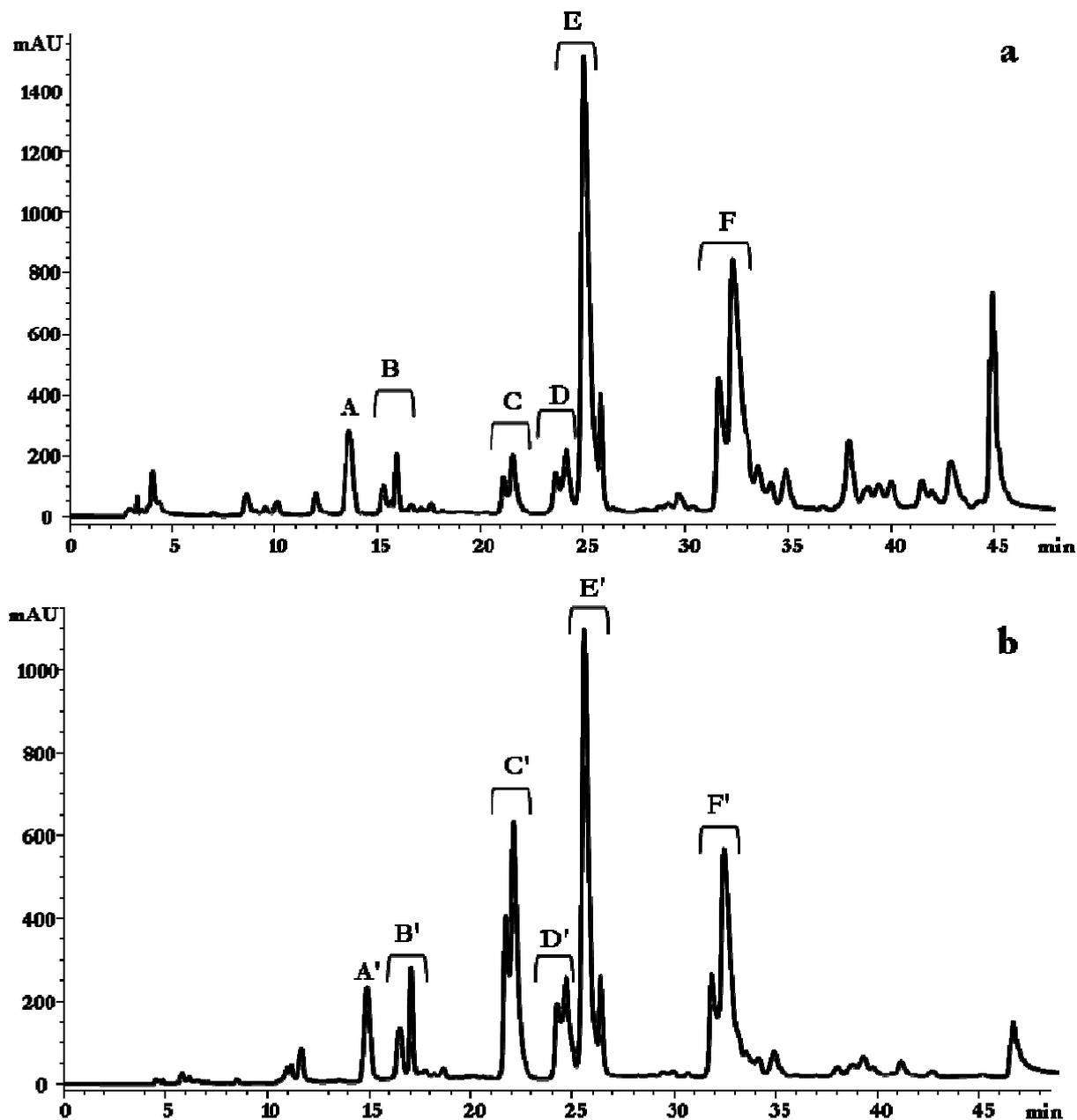


Figure 2.11: Chromatogram profiles ($\lambda=232$ nm) of the SAX-HPLC separation of the fractions A belonging to the AC digestion of skate cartilage CS (b) and of torpedo cartilage CS (a).

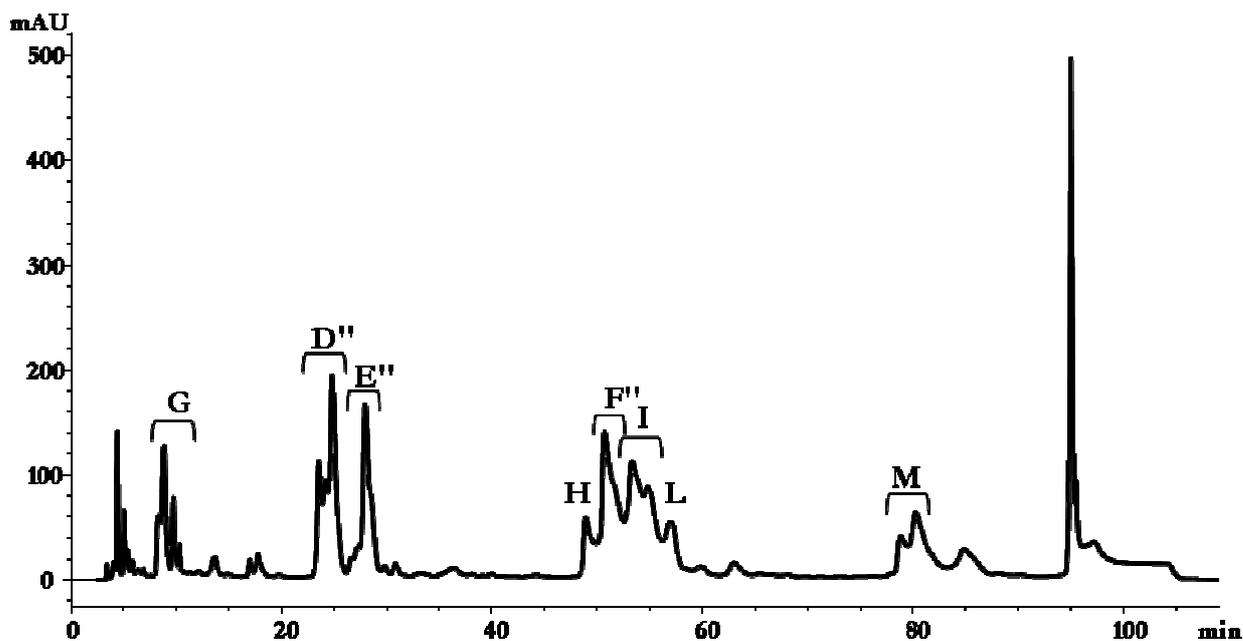


Figure 2.12: Chromatogram profile ($\lambda=232$ nm) of the SAX-HPLC separation of the Frz-A belonging to the AC digestion of skate skin CS.

The SAX-HPLC separation led to the isolation of the structures reported in Table 2.8, the relative amount of each component (Tab. 2.9) was estimated through the integration of the chromatographic profiles.

Table 2.8: Unsaturated oligosaccharides isolated through SAX-HPLC analyses from the Fr.A of the three AC lyase digestions. The components containing each oligosaccharide are reported too.

Oligosaccharides	Components
I) $\Delta\text{HexA}(1\rightarrow3)\text{GalNAc}(1\rightarrow4)\text{GlcA}2\text{S}$	E, E', E''
II) $\text{GalNAc}(1\rightarrow4)\text{GlcA}2\text{S}(1\rightarrow3)\text{GalNAc}6\text{S}$	E
III) $\Delta\text{HexA}(1\rightarrow3)\text{GalNAc}(1\rightarrow4)\text{GlcA}2\text{S}(1\rightarrow3)\text{GalNAc}6\text{S}$	E, F, F', F''
IV) $\Delta\text{HexA}(1\rightarrow3)\text{GalNAc}(1\rightarrow4)\text{GlcA}(1\rightarrow3)\text{GalNAc}6\text{S}$	D''
V) $\Delta\text{HexA}(1\rightarrow3)\text{GalNAc}(1\rightarrow4)\text{GlcA}(1\rightarrow3)\text{GalNAc}4\text{S}$	D''
VI) $\Delta\text{HexA}(1\rightarrow3)\text{GalNAc}6\text{S}(1\rightarrow4)\text{GlcA}(1\rightarrow3)\text{GalNAc}6\text{S}$	I

Table 2.9: Relative amount of each component in the different fractions A

Components	Fr-A	Fr-A	Fr-A
	<i>Raja bochyura</i> cartilage	<i>Torpedo nobiliana</i> cartilage	<i>Raja bochyura</i> skin
A, A'	5.5%	5.7%	--
B, B'	6.6%	5.1%	--
C, C'	20.8%	4.8%	--
D, D', D''	10.4%	6.1%	19%
E, E', E''	30.8%	38%	11.1%
F, F', F''	25.9%	40.3%	17.3%
G	--	--	9.4%
I	--	--	22.5%
L	--	--	6.1%
M	--	--	10.1%

Structural analyses revealed the occurrence of several artefact; the oligosaccharide I (Tab. 2.8) was explained on the basis of peeling reaction which degraded the terminal reducing GalNAc6S. Peeling reaction were caused by the chromatographic support, as proved by the injection, in the same conditions of the pure tetrasaccharide III, since the corresponding chromatogram profile contained also a peak at the same retention time of the trisaccharide I.

The origin of product II was understood later and it was related to the presence of an exoglycosidase contaminating the commercial AC lyase.

The structures of molecules (I, II, III), however, suggested the resistance to the AC I lyase action of oligosaccharide made up by CS-OS units linked to CS-D motifs.

In the sequences of all the oligosaccharides isolated (Tab. 2.8), moreover, CS-A motif was found in a little amount; these datum suggested again that the commercial AC lyase was contaminated by a sulfoesterase, probably selective for motifs containing GalNAc4S unit, therefore a test to prove the presence of this contaminants become necessary.

Test of AC lyase enzymatic activity on reference compounds

Starting from a mixture of unsaturated disaccharides (CS-A, CS-C and CS-D) a treatment with the AC lyase was performed and the resulting products were analyzed by SAX-HPLC.

The condition applied for the enzymatic digestions were similar to those used for the depolymerisation of the skate and torpedo CS materials.

SAX-HPLC analytical runs were performed at regular time intervals on the digestion products, and a net increment of CS-0S motif together with a decrement of the monosulfated motifs was detected during the time.

In particular after 1,5 h a decrement of both the CS-A and CS-C motifs was encountered while after 5 h a net reduction of only the CS-A motif was detected (Fig. 2.13). In this way the presence of sulfoesterase in the commercial enzymatic preparation was established and moreover a selectivity for CS-A motifs was proved.

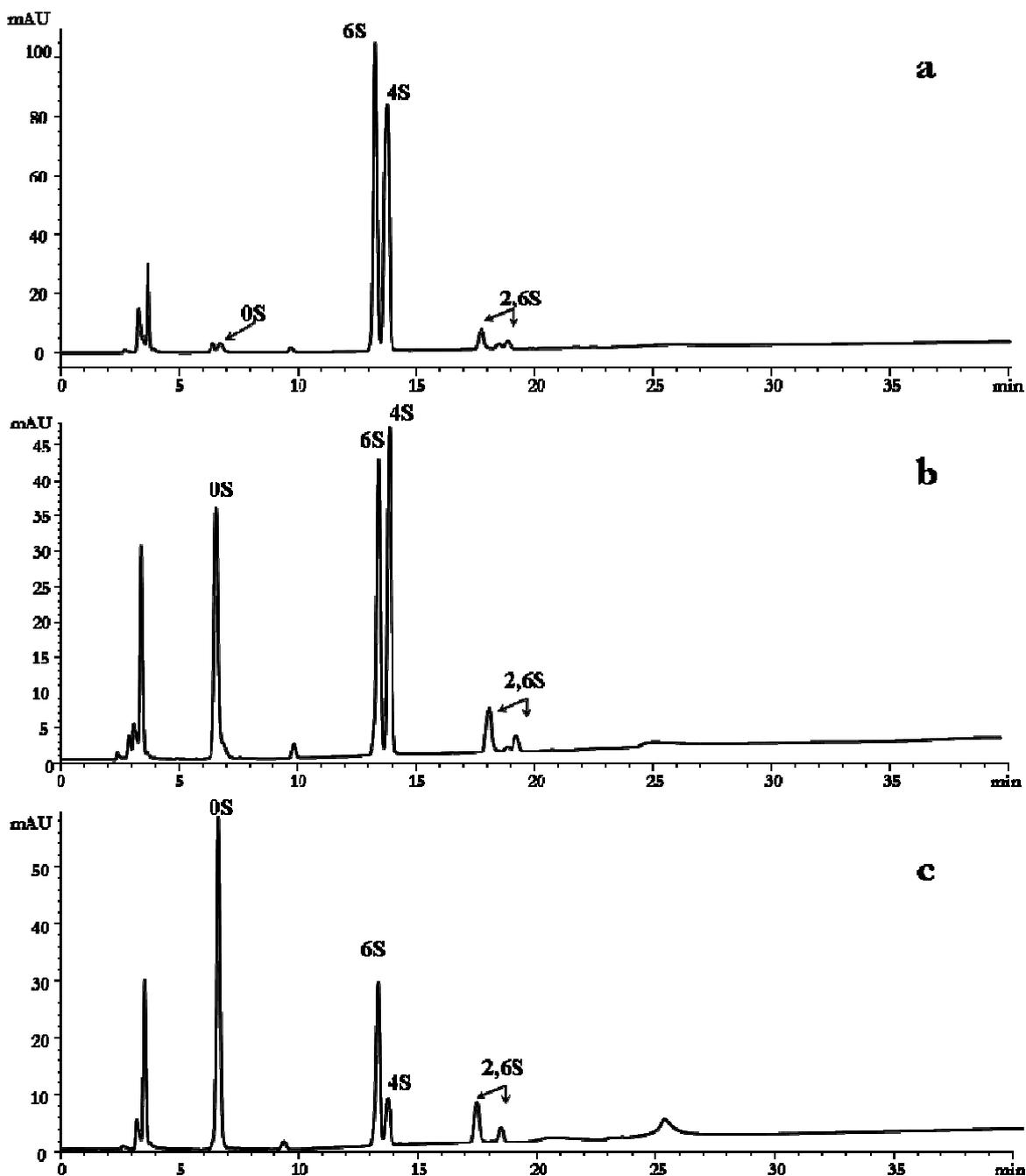


Figure 2.13: Chromatogram profiles ($\lambda=232$ nm) of the SAX-HPLC analysis of the AC digestion products on reference compounds to prove the presence of contaminants like sulfoesterase and exoglycosidases. Chromatogram of the reference compounds (a), of the digestion products after 1,5 h (b) and of digestion products after 5h (c).

2.2.4.1 NMR analyses

The NMR analysis through 1D and 2D experiments of the different components isolated through SAX-HPLC from the fractions A has led to the identification of the different oligosaccharides listed in Table 2.8. Some components were a mix of different oligosaccharides, so the structural investigation was more complex, but the high resolution of the spectra allowed the complete chemical shift assignment of each product and the residues sequencing.

The components B, B', D, D' and C, C' were analyzed just through 1D-NMR (Figure 2.14) due their low abundance while the components D'', E, E', F, F'' and I were analyzed by both 1D and 2D experiments.

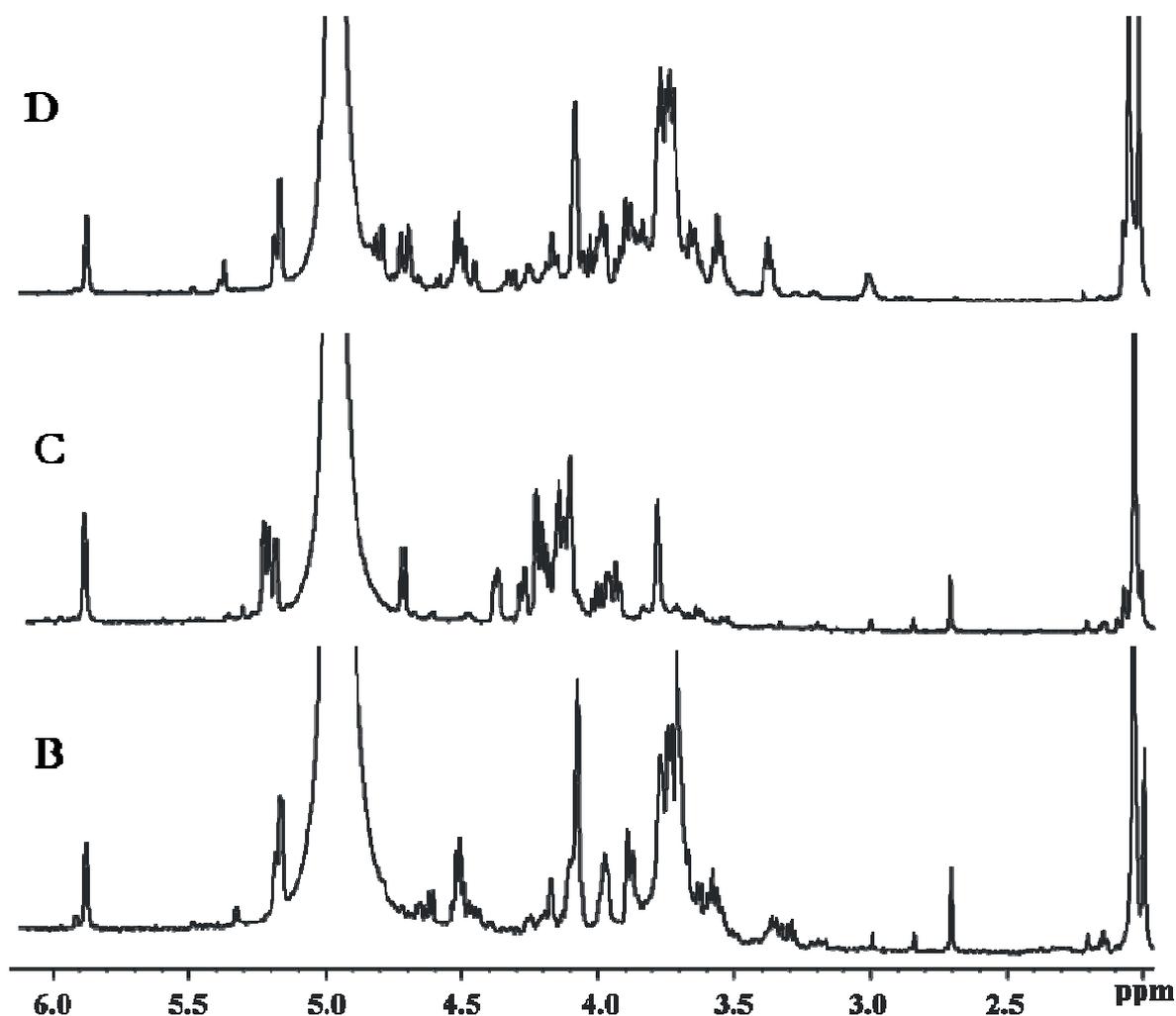


Figure 2.14: (600 MHz, D₂O, 283 K) ¹H-NMR spectra of components D, C, and B.

The ¹H-NMR spectra of all the oligosaccharides were characterized by three blocks of signals: one between 5.9 and 4.9 ppm containing the H1 and H4 signals of the unsaturated ΔHexA unit and the anomeric signals of the α reducing termini, a second region between 4.7 and 3.3 ppm

containing the anomeric signals of the β units and the carbinolic protons, and a third region around 2 ppm containing the signals of the methyl of the *N*-Acetyl groups.

The chemical shift assignments of the different oligosaccharides was performed through the combined study of homonuclear (DQF-COSY, T-ROESY and TOCSY) and heteronuclear spectra (HSQC and HMBC), according to the general strategy illustrated below.

In all cases the chemical shift assignment of the Δ HexA units was easily performed through the clear and strong COSY and TOCSY correlations. Differently, the assignment of the GalNAc units was performed from H1 to H4 following the COSY and TOCSY correlations, while H5 and H6 were identified through NOE contacts with H4. The position of H5, in β configured sugar, moreover, was confirmed also by the typical intra-residue NOE contact between H1-H5.

The characterization of the GlcA2S units was performed on the basis of the TOCSY correlations and the intra-residue NOE contacts, because COSY correlations of such residues, despite the *gluco* configuration, were not very clear. The chemical shift assignment of the ring protons of the unsulfated GlcA units was easily performed due to the presence of strong COSY cross-peaks, helped by TOCSY correlations linking the anomeric proton to all the others.

All the carbon chemical shifts were assigned through the analysis of the HSQC and HMBC spectra.

The sequence of each oligosaccharide was defined on the basis of the inter-residue NOE contacts involving the anomeric protons. The linkage positions were confirmed by the analysis of β and α glycosilation shifts in the HSQC spectrum and by the long-range C-H correlations in the HMBC spectrum.

Figure 2.15 shows an expansion of the anomeric region of TOCSY and T-ROESY spectra superimposed of the component F was reported, in order to give an idea of the TOCSY correlations found for each spin system and the NOE contacts used for the definition of the sequences.

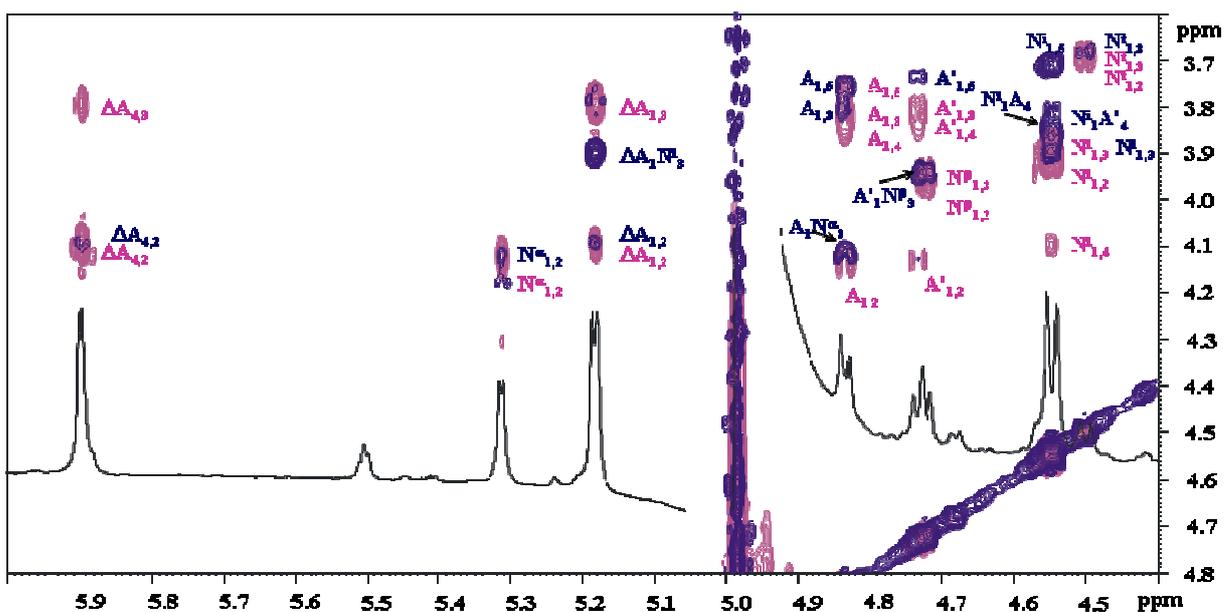


Figure 2.15: (600 MHz, D₂O, 283 K) Expansion of the anomeric region of the TOCSY (pink) and T-TOESY (blue) spectra superimposed of the component F. TOCSY correlations (pink) and NOE contacts (blue) were displayed as well. ΔA is the ΔHexA unit, N^α and N^β are the respectively the α and β forms of the reducing GalNAc6S, Nⁱ is the internal GalNAc unit, N^t is the terminal GalNAc and A and A' are the two forms of the internal GlcA2S (see text).

The 2D-NMR analysis established that both E' and E'' components contained only the trisaccharide I, because the anomeric signal at 5.18 ppm belonged to the ΔHexA unit, the one at 4.46 ppm to the internal β-GalNAc unit and the one at 5.51 to the reducing α-GlcA2S unit (Fig. 2.16).

The surprising datum, highlighted by such NMR analysis, was the absence of the β anomeric form of the reducing end, indicating that the sulfate group in position 2 stabilized the α anomer and did not favour the presence of the classical equilibrium between both the α and β forms.

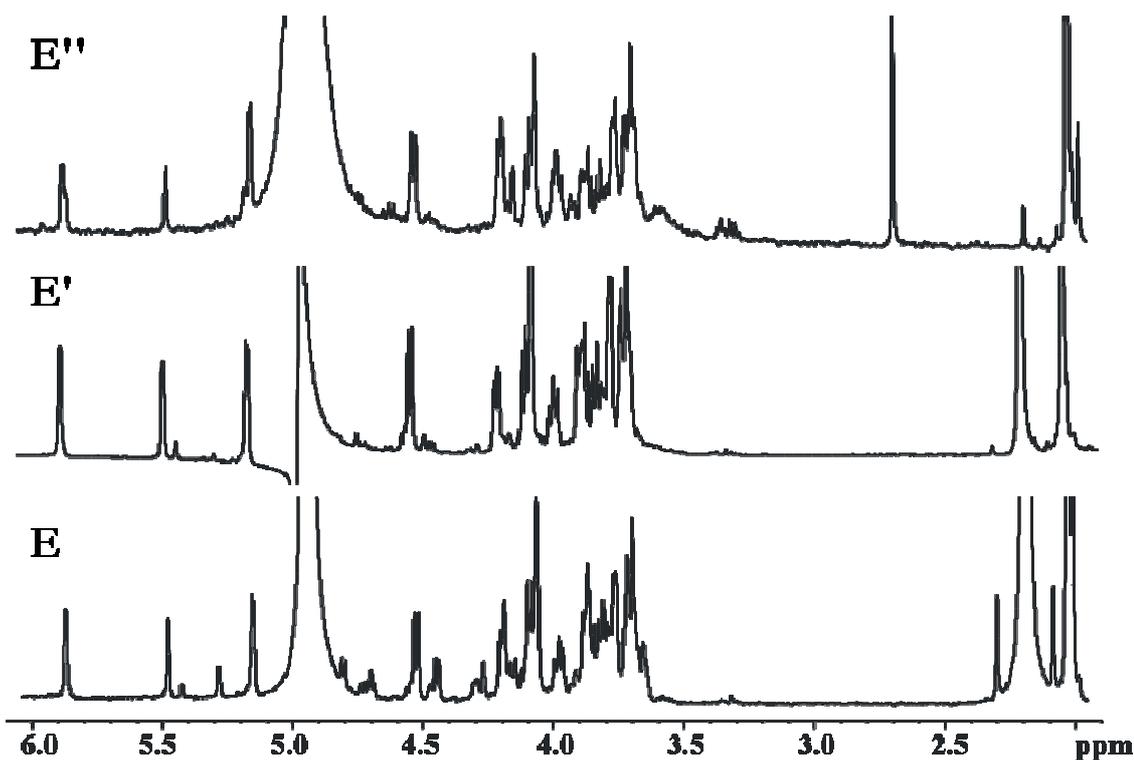


Figure 2.16: (600 MHz, D₂O, 283 K) ¹H-NMR spectra of components E, E' and E''.

Differently, the 2D-NMR analysis on component E indicated that it contained a mixture of tri and tetrasaccharides I, II and III (Tab 2.8), because additionally to the same spin systems found in E' and E'', an internal β -GlcA2S unit (4.83 ppm), a terminal β -GalNAc unit (4.48 ppm) and a reducing GalNAc6S unit in both α and β forms (5.31 ppm and 4.72 ppm, respectively) were identified as well (Fig. 2.16).

The complete chemical shift assignments of the three oligosaccharide are reported in Tables 2.10 and 2.11.

The study of the 2D spectra indicated that the components F, F' and F'' contained only the tetrasaccharides III, so the following spin systems were identified (Fig. 2.17): a Δ HexA unit at 5.18 ppm, a reducing GalNAc6S in both the α (5.31 ppm) and β (4.72 ppm) forms, an internal β -GalNAc unit at 4.55 and an internal β -GlcA2S unit that on the basis of the anomeric configuration of the reducing end showed a different chemical shift value (Tab 2.11).

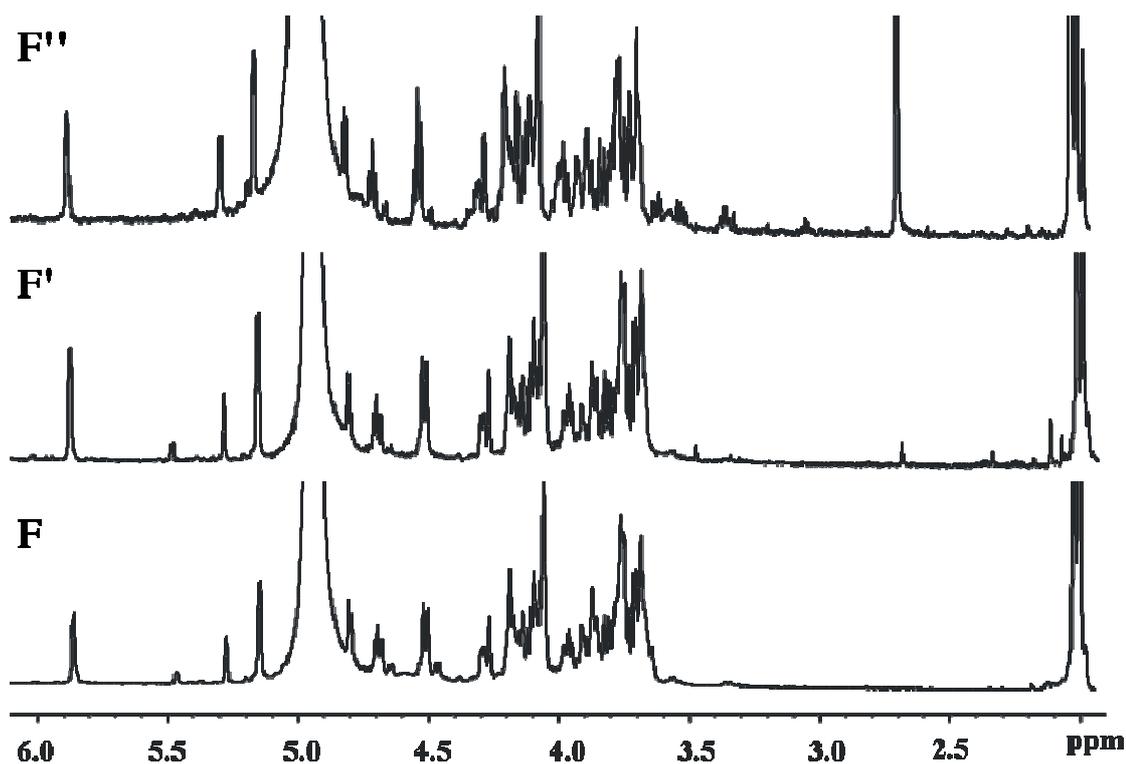


Figure 2.17: (600 MHz, D₂O, 283 K) ¹H-NMR spectra of components F, F', and F''.

The component D'', after the analysis of the 2D spectra, resulted characterized by a mixture of the tetrasaccharides IV and V (Tab 2.8). In its spectra (Fig. 2.18), so, an anomeric signal at 5.18 ppm of the ΔHexA unit and two overlapped signals at 5.20 ppm belonging to two different reducing ends, a GalNAc6S and GalNAc4S were found in the down field region, while in the crowded region between 4.8 and 4.4 ppm there were the two H4 signals of the α and β forms of reducing GalNAc4S, the anomeric signals of the two β reducing ends (GalNAc6S at 4.68 ppm and GalNAc4S at 4.71 ppm), the anomeric signal of the internal β-GalNAc unit at 4.53 and anomeric signals of two different β-GlcA units, that showed different chemical shift values due to the anomeric form of the reducing end. The NMR data are listed in Table 2.12 and 2.13.

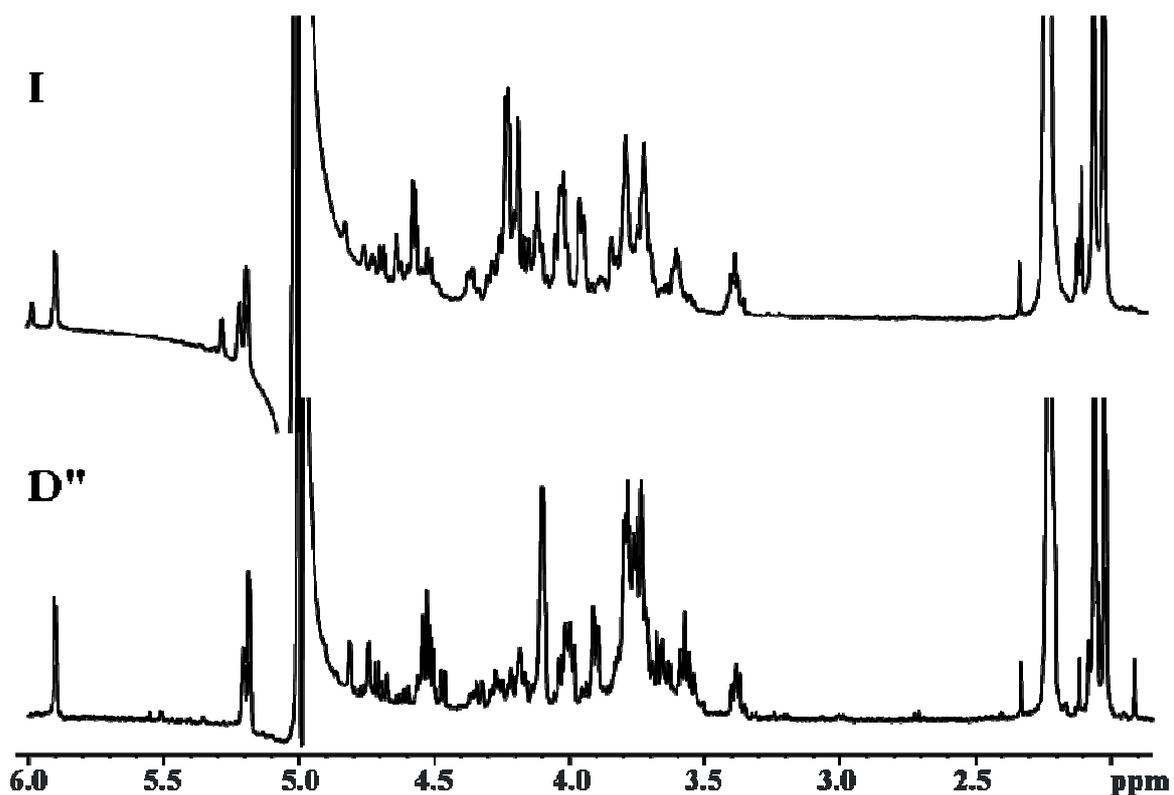


Figure 2.18: (600 MHz, D₂O, 283 K) ¹H-NMR spectra of components D'' and I.

The component I contained a mixture of species not completely characterized in which the oligosaccharide VI is the more abundant. Through the analysis of the high resolved 2D spectra the identification of all the protons belonging to the following rings was allowed: the Δ HexA unit (5.18 ppm), the internal GlcNAc6S unit (4.56 ppm), the reducing GalNAc6S in both anomeric forms (α 5.20 ppm and β 4.68 ppm) and the internal β -GlcA, characterized by two different set of chemical shifts on the basis of the anomeric configuration of reducing end (Fig. 2.18). The complete chemical shift assignment of the oligosaccharide VI is reported in Table 2.14.

Table 2.10: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the two unsaturated trisaccharides (components E, E' and E'') isolated via SAX-HPLC chromatography from the Fr.A belonging to the AC lyase digestions of the CS material isolated from the cartilage of skate and torpedo. The 2D-NMR experiments were performed at 283 K. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton and carbon chemical shifts of methyl group of the *N*-Acetyl are around 2.02 and 23.1 ppm. Respectively.

ΔHexA-(1\rightarrow3)-GalNAc-(1\rightarrow4)-GlcA2S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.18	3.79	4.10	5.90	--	--
	<i>102.6</i>	<i>70.7</i>	<i>66.9</i>	<i>108.6</i>	<i>145.4</i>	<i>170.7</i>
3)-β-GalNAc-(1\rightarrow	4.46	4.01	3.91	4.09	3.73	3.80-3.74
	<i>101.8</i>	<i>52.3</i>	<i>81.2</i>	<i>68.7</i>	<i>76.4</i>	<i>62.5</i>
4)-α-GlcA2S	5.51	4.23	3.89	3.84	4.12	
	<i>91.5</i>	<i>78.6</i>	<i>70.2</i>	<i>80.6</i>	<i>72.8</i>	<i>176.5</i>
GalNAc-(1\rightarrow4)-GlcA2S-(1\rightarrow3)-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
GalNAc-(1\rightarrow	4.48	3.88	3.69	3.90	3.68	3.74-3.81
	<i>101.8</i>	<i>53.4</i>	<i>72.1</i>	<i>69.0</i>	<i>76.3</i>	<i>62.3</i>
4)-β-GlcA2S-(1\rightarrow	4.83	4.13	3.81	3.86	3.75	--
	<i>102.5</i>	<i>80.8</i>	<i>73.8</i>	<i>79.8</i>	<i>77.5</i>	<i>175.7</i>
3)-β-GalNAc6S	4.72	3.90	3.95	4.22	3.95	4.20-4.14
	<i>97.0</i>	<i>53.7</i>	<i>81.0</i>	<i>68.7</i>	<i>74.0</i>	<i>69.2</i>
3)-α-GalNAc6S	5.31	4.18	4.11	4.30	4.33	4.20-4.14
	<i>91.9</i>	<i>50.3</i>	<i>77.9</i>	<i>68.7</i>	<i>69.4</i>	<i>69.2</i>

Table 2.11: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the unsaturated tetrasaccharide (component F, F' and F'') isolated via SAX-HPLC chromatography from the Fr.A belonging to the AC lyase digestions of the CS material isolated from the cartilage of skate and torpedo. The 2D-NMR experiments were performed at 283 K. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton chemical shifts of methyl group of the *N*-Acetyl are located in the region between 2.03 and 2.07 ppm, while the carbon chemical shifts are around 23.8 ppm.

ΔHexA-(1\rightarrow3)-β-GalNAc-(1\rightarrow4)-β-GlcA2S-(1\rightarrow3)-α-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.18	3.79	4.10	5.90	--	--
	<i>102.7</i>	<i>70.9</i>	<i>67.0</i>	<i>108.7</i>	<i>145.3</i>	<i>170.7</i>
3)-β-GalNAc-(1\rightarrow	4.55	4.00	3.90	4.10	3.72	~3.8
	<i>101.9</i>	<i>52.3</i>	<i>81.4</i>	<i>68.9</i>	<i>76.6</i>	<i>62.6</i>
4)-β-GlcA2S	4.83	4.13	3.81	3.86	3.76	--
	<i>102.6</i>	<i>81.0</i>	<i>74.2</i>	<i>80.2</i>	<i>77.5</i>	<i>175.4</i>
3)-α-GalNAc6S	5.31	4.19	4.12	4.30	4.33	4.21-4.14
	<i>92.1</i>	<i>50.5</i>	<i>78.1</i>	<i>68.9</i>	<i>69.5</i>	<i>69.5</i>
ΔHexA-(1\rightarrow3)-β-GalNAc-(1\rightarrow4)-β-GlcA2S-(1\rightarrow3)-β-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.18	3.79	4.10	5.90	--	--
	<i>102.7</i>	<i>70.9</i>	<i>67.0</i>	<i>108.7</i>	<i>145.3</i>	<i>170.7</i>
3)-β-GalNAc-(1\rightarrow	4.55	4.00	3.90	4.10	3.72	~3.8
	<i>101.9</i>	<i>52.3</i>	<i>81.4</i>	<i>68.9</i>	<i>76.6</i>	<i>62.6</i>
4)-β-GlcA2S	4.74	4.13	3.81	3.86	3.74	--
	<i>103.0</i>	<i>81.0</i>	<i>74.2</i>	<i>80.2</i>	<i>77.5</i>	<i>175.4</i>
3)-β-GalNAc6S	4.72	3.90	3.94	4.23	3.94	4.23-4.17
	<i>97.0</i>	<i>53.7</i>	<i>81.0</i>	<i>68.7</i>	<i>74.0</i>	<i>69.5</i>

Table 2.12: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the unsaturated tetrasaccharide (component D") isolated via SAX-HPLC chromatography from the Fr.A belonging to the AC lyase digestions of the CS material isolated from the skin of skate. The 2D-NMR experiments were performed at 283 K. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton chemical shifts of methyl group of the *N*-Acetyl are located in the region between 2.01 and 2.07 ppm, while the carbon chemical shifts are around 23.6 ppm. The HMBC experiment was not performed due the low abundance of the sample.

ΔHexA-(1\rightarrow3)-β-GalNAc-(1\rightarrow4)-β-GlcA-(1\rightarrow3)-α-GalNAc4S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.18	3.79	4.10	5.90	--	--
	<i>102.5</i>	<i>70.6</i>	<i>66.8</i>	<i>108.4</i>	--	--
3)-β-GalNAc-(1\rightarrow	4.53	4.00	3.90	4.10	3.73	~3.78
	<i>102.0</i>	<i>52.2</i>	<i>81.2</i>	<i>68.6</i>	<i>76.3</i>	<i>62.2</i>
4)-β-GlcA	4.51	3.39	3.57	3.78	3.67	--
	<i>104.8</i>	<i>73.4</i>	<i>74.9</i>	<i>81.0</i>	<i>77.6</i>	--
3)-α-GalNAc4S	5.21	4.33	4.18	4.81	4.26	3.73-3.79
	<i>92.5</i>	<i>50.8</i>	<i>73.9</i>	<i>78.9</i>	<i>71.5</i>	<i>62.2</i>
ΔHexA-(1\rightarrow3)-β-GalNAc-(1\rightarrow4)-β-GlcA-(1\rightarrow3)-β-GalNAc4S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1\rightarrow	5.18	3.79	4.10	5.90	--	--
	<i>102.5</i>	<i>70.6</i>	<i>66.8</i>	<i>108.4</i>	--	--
3)-β-GalNAc-(1\rightarrow	4.53	4.00	3.90	4.10	3.73	~3.78
	<i>102.0</i>	<i>52.2</i>	<i>81.2</i>	<i>68.6</i>	<i>76.3</i>	<i>62.2</i>
4)-β-GlcA	4.47	3.39	3.58	3.78	3.66	--
	<i>104.8</i>	<i>73.4</i>	<i>74.9</i>	<i>81.0</i>	<i>77.6</i>	--
3)-β-GalNAc4S	4.71	4.03	4.01	4.74	3.82	~3.78
	<i>96.1</i>	<i>52.3</i>	<i>78.1</i>	<i>77.7</i>	<i>75.8</i>	<i>62.2</i>

Table 2.13: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the unsaturated tetrasaccharide (component D") isolated via SAX-HPLC chromatography from the Fr.A belonging to the AC lyase digestions of the CS material isolated from the skin of skate. The 2D-NMR experiments were performed at 283 K. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton chemical shifts of methyl groups of the *N*-Acetyl are located in the region between 2.01 and 2.07 ppm, while the carbon chemical shifts are around 23.6 ppm. The HMBC experiment was not performed due the low abundance of the sample.

ΔHexA-(1→3)-β-GalNAc-(1→4)-β-GlcA-(1→3)-α-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1→	5.18	3.79	4.10	5.90	--	--
	<i>102.5</i>	<i>70.6</i>	<i>66.8</i>	<i>108.4</i>	--	--
3)-β-GalNAc-(1→	4.53	4.00	3.90	4.10	3.73	~3.78
	<i>102.0</i>	<i>52.2</i>	<i>81.2</i>	<i>68.6</i>	<i>76.3</i>	<i>62.2</i>
4)-β-GlcA	4.55	3.37	3.58	3.74	3.70	--
	<i>104.8</i>	<i>73.4</i>	<i>74.9</i>	<i>81.0</i>	<i>77.6</i>	--
3)-α-GalNAc6S	5.21	4.28	4.03	4.25	4.36	~4.15
	<i>92.5</i>	<i>51.3</i>	n.d.	<i>69.5</i>	<i>69.5</i>	n.d.
ΔHexA-(1→3)-β-GalNAc-(1→4)-β-GlcA-(1→3)-β-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1→	5.18	3.79	4.10	5.90	--	--
	<i>102.5</i>	<i>70.6</i>	<i>66.8</i>	<i>108.4</i>	--	--
3)-β-GalNAc-(1→	4.53	4.00	3.90	4.10	3.73	~3.78
	<i>102.0</i>	<i>52.2</i>	<i>81.2</i>	<i>68.6</i>	<i>76.3</i>	<i>62.2</i>
4)-β-GlcA	4.48	3.37	3.58	3.74	3.59	--
	<i>104.8</i>	<i>73.4</i>	<i>74.9</i>	<i>81.0</i>	<i>77.6</i>	--
3)-β-GalNAc6S	4.68	4.01	3.82	4.18	3.94	4.18
	n.d.	<i>52.2</i>	<i>75.4</i>	n.d.	n.d.	n.d.

* not determinate (n.d.)

Table 2.14: (600 MHz, D₂O) proton (plain) and carbon (italic) chemical shift assignment of the unsaturated tetrasaccharide (component I) isolated via SAX-HPLC chromatography from the Fr.A belonging to the AC lyase digestions of the CS material isolated from the skin of skate. The 2D-NMR experiments were performed at 283 K. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton chemical shifts of methyl group of the *N*-Acetyl are located in the region between 2.03 and 2.07 ppm, while the carbon chemical shifts are around 23.6 ppm. The HMBC experiment was not performed due the low abundance of the sample.

ΔHexA-(1→3)-β-GalNAc6S-(1→4)-β-GlcA-(1→3)-α-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1→	5.18	3.79	4.11	5.89	--	--
	<i>102.6</i>	<i>70.7</i>	<i>67.1</i>	<i>108.1</i>	--	--
3)-β-GalNAc6S-(1→	4.56	4.02	3.94	4.19	4.01	4.22-4.19
	<i>102.6</i>	<i>52.2</i>	<i>81.1</i>	<i>68.5</i>	<i>73.9</i>	<i>68.8</i>
4)-β-GlcA	4.48	3.39	3.59	3.68	3.68	--
	<i>105.5</i>	<i>73.6</i>	<i>75.1</i>	<i>82.7</i>	<i>77.5</i>	--
3)-α-GalNAc6S	5.20	4.34	4.18	4.82	4.36	~4.2
	<i>92.4</i>	<i>50.8</i>	<i>73.9</i>	<i>78.9</i>	<i>69.5</i>	<i>68.8</i>
ΔHexA-(1→3)-β-GalNAc-(1→4)-β-GlcA-(1→3)-β-GalNAc6S						
	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6/C-6
t-ΔHexA-(1→	5.18	3.79	4.11	5.89	--	--
	<i>102.6</i>	<i>70.7</i>	<i>67.1</i>	<i>108.1</i>	--	--
3)-β-GalNAc6S-(1→	4.56	4.02	3.94	4.19	4.01	4.22-4.19
	<i>102.6</i>	<i>52.2</i>	<i>81.1</i>	<i>68.5</i>	<i>73.9</i>	<i>68.8</i>
4)-β-GlcA	4.51	3.83	3.59	3.71	3.70	--
	<i>105.5</i>	<i>73.6</i>	<i>75.1</i>	<i>83.1</i>	<i>77.5</i>	--
3)-β-GalNAc6S	4.68	4.01	3.83	4.18	3.94	4.22-4.18
	<i>96.5</i>	<i>52.2</i>	<i>81.5</i>	<i>68.5</i>	<i>74.1</i>	<i>68.7</i>

2.3 The keratan sulfate of lesser spotted dogfish

The extraction process performed on the fresh cartilage of lesser spotted dogfish led to the isolation of keratan sulfate chains, too.

A mix of KS and CS was obtained, so many steps of purification were performed before obtaining a pure sample. The procedures used for the purification and analysis of the KS samples were similar to those used for chondroitin sulfate chains.

2.3.1 Isolation and purification of the polysaccharide

The keratan sulfate material, belonging to the cartilage of the lesser spotted dogfish, was recovered from the ethanolic supernatant obtained after the proteolytic digestion of the fresh tissues.

After the anionic-exchange chromatography (Q-sepharose), the KS material was recovered in the fractions belonging to the elutions with the NaCl 700 mM and 1 M buffers.

A first chemical analysis indicated the presence of a mix of CS and KS, due the presence of amounts of GlcA, Gal, GlcNAc, GalNAc and sialic acid. The two chromatographic fractions were put together and treated twice with ABC lyase and more times with protease, to remove all contaminants: the chondroitin sulfate and proteins.

After some chromatographic purifications on Sephadex S100 and Biogel P2, a sialylated keratan sulfate polysaccharide was isolated.

At first, a structural analysis through NMR on the polysaccharide was carried out, then the sulfated molecule was hydrolyzed with NaOH in a reductive buffer in order to remove the proteic portion linked to the polysaccharide reducing end. The positive result of the hydrolytic treatment indicated a *O*-linked nature of the sulfate glycan²⁶.

The molecule, obtained in this way, underwent a deep NMR survey and an estimation of its molecular weight.

2.3.2 Molecular weight measurement

A solution of the proteic-free KS sample (1 mg/ml) was analyzed by SEC-HPLC in order to establish the molecular weight of the polysaccharide. Using sulfated dextran standards (9-20 kDa, 500 kDa), a molecular weight approximately of 10 kDa was estimated for the molecule, as indicated by the chromatographic profiles reported in the Figure 2.19.

²⁶ Montreuil, J., Bouquelet, S., Debray, H., Lemoine, J., Michalski, J.C., Spik, G., Strecker, G. Cap 5, Carbohydrate analysis a practical approach II edition M.F., Chaplin, J.F., Kennedy, Eds, IRL press, 1994.

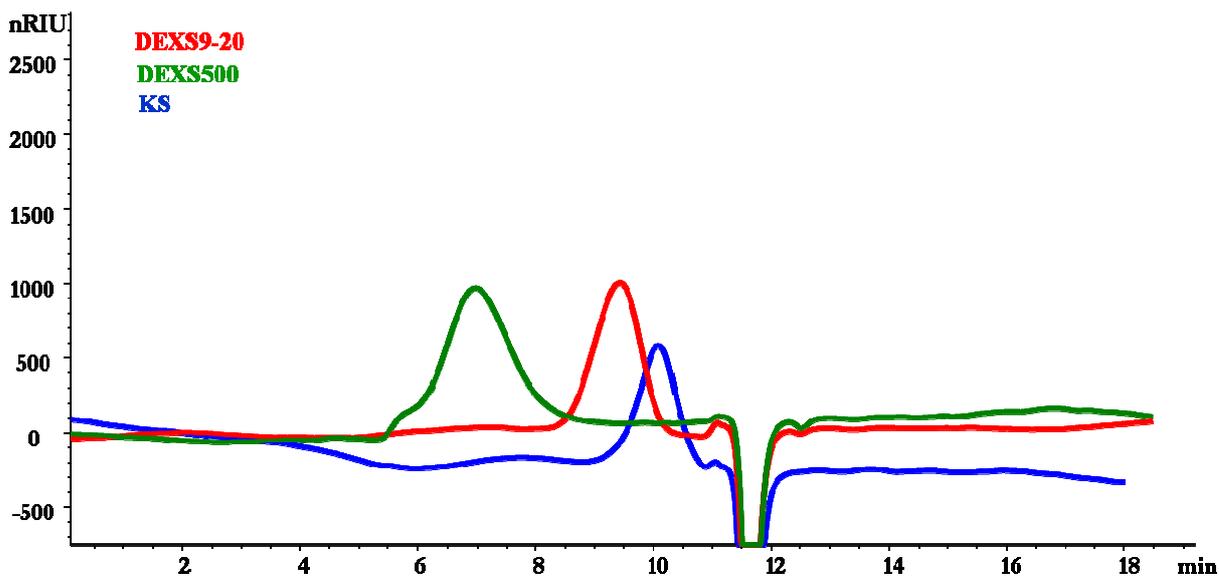


Figure 2.19: SEC-HPLC (TSK-G_{XL} 5000, RID) profiles of Keratan sulfate (blue line), sulfate dextran standard 9-20 KDa (red line) and sulfate dextran standards 500 KDa (green line).

2.3.3 Structural characterization by NMR

As typical of polysaccharide sample, the ¹H-NMR spectrum of proteic-free KS sample (Fig 2.20) was characterized by broad signals. Two main groups of signals were individuated: one between 4.8 and 3.5 ppm containing the signals belonging to the anomeric and carbinolic protons of the sugar backbone, one between 2.8 and 1.7 ppm containing the signals belonging to the *N*-acetyl group of the GlcNAc and sialic acid units, and those of the diastereotopic protons in position 3 and 3' of the sialic acid residues.

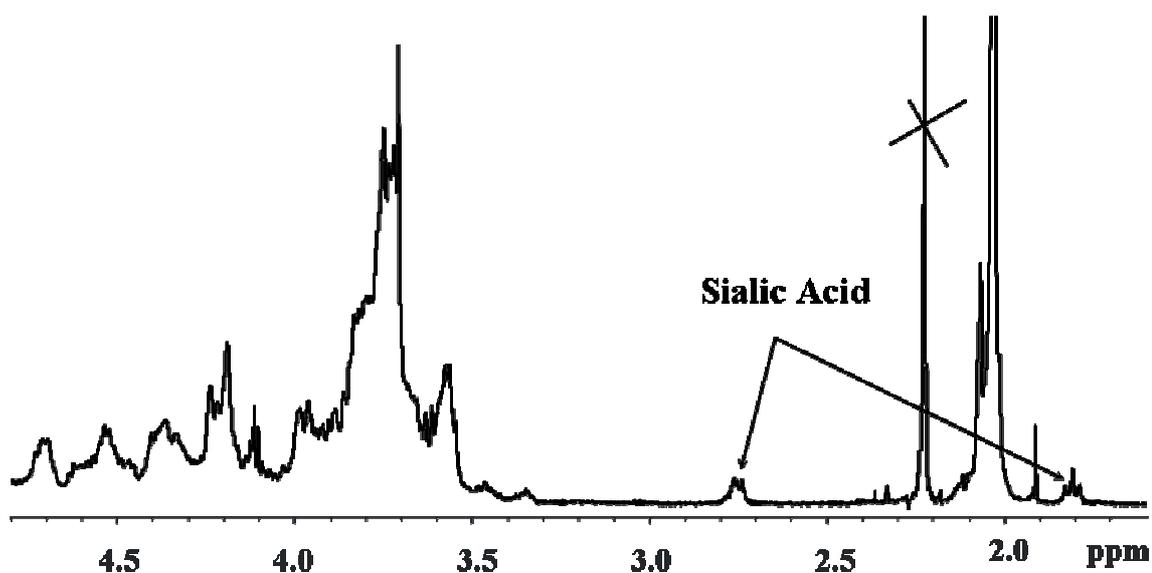


Figure 2.20: (600 MHz, 288 K, D₂O) ¹H-NMR spectrum of the keratan sulfate of the lesser spotted dogfish.

The proton and carbon chemical shift assignments of all the different spin systems (Tab 2.15) were performed using the classical approach described above for the CS of the lesser spotted

dogfish, using homonuclear (COSY, TOCSY, NOESY) and heteronuclear (HSQC, HMBC) spectra.

Table 2.15: (600 MHz, 288 K, D₂O) proton (plain) and carbon (italic) chemical shift attribution of the keratan sulfate isolate from the lesser spotted dogfish. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). The proton and carbon chemical shifts of methyl group of the *N*-Acetyl are 2.03 and 23.5 ppm

	1	2	3	4	5	6		
4-)-β-GlcNAc6S-(1\rightarrow	4.71	3.82	3.73	3.75	3.73	4.36		
N^I	<i>104.25</i>	<i>56.4</i>	<i>76.4/73.4</i>	<i>83.53</i>	<i>76.4/73.4</i>	<i>67.6</i>		
4-)-β-GlcNAc6S-(1\rightarrow	4.69	3.82	3.73	3.75	3.73	4.36		
N^{II}	<i>104.25</i>	<i>56.4</i>	<i>76.4/73.4</i>	<i>83.53</i>	<i>76.4/73.4</i>	<i>67.6</i>		
3-)-β-Gal-(1\rightarrow	4.63	3.55	4.12	3.96	3.71			
G^I	<i>103.301</i>	<i>71.05</i>	<i>76.67</i>	<i>68.77</i>				
3)-β-Gal6S-(1\rightarrow	4.54	3.57	3.74	4.24	3.97	4.19		
G^{II}	<i>104.06</i>	<i>71.09</i>	<i>83.57</i>	<i>69.62</i>	<i>73.86</i>	<i>68.88</i>		
3)-β-Gal-(1\rightarrow	4.52	3.56	3.72	4.19	3.98	3.76		
G^{III}	<i>104.06</i>	<i>71.09</i>	<i>83.57</i>	<i>69.86</i>	<i>73.86</i>	<i>62.19</i>		
3)-β-Gal-(1\rightarrow	4.46	3.57	3.66	3.89	3.72			
G^{IV}	<i>104.14</i>	<i>71.09</i>	<i>73.78</i>	<i>73.29</i>				
	1	2	3	4	5	6	7	8
α-NeuA-(2\rightarrow	--	--	1.81-2.74	3.63	3.86	*	*	*
NA	<i>175.8</i>	<i>101.2</i>	<i>41.01</i>	<i>69.71</i>	<i>53.06</i>	n.d.	n.d.	n.d.

n.d. not determinate

* The signals were found in the crowded region around 3.7 ppm

The anomeric protons of the GlcNAc units merged in one signal at approx. 4.7 ppm, while those of the Gal units in the one at c.a. 4.5 ppm.

The signals of H1 to H5, belonging to the different GlcNAc units, were individuated through the COSY and TOCSY correlations, a lot of difficulties were found in the identification of the H3, H4 and H5 signals, due to the high similarity in the chemical shift values. The diastereotopic H6 protons in position 6 were easily found through the NOE contact with the H4 and H5. The signals of these protons were found at approx. 4.36 ppm, indicating the presence of a sulfate

group. This datum was then confirmed by the carbon chemical shift at 67 ppm. The chemical shift assignment of the two GlcNAc units was different just for the anomeric protons.

The signals of H1 to H3 of the Gal units were identified through the COSY and TOCSY correlations. The proton H4 of the sulfated Gal units was individuated through a NOE contact with H3 at approx 4.24 ppm and confirmed by a weak TOCSY signal with the anomeric proton, while the H5 and H6,6' protons were identified through NOE contacts with H4. The H4 of unsulfated Gal units was identified at 4.19 ppm and the H5 and H6,6' of the same units were localized in the high crowded region around 3.6 ppm. In particular the H5 was collocated at 3.76 ppm on the basis of NOE contact with the anomeric proton, characteristic of a β configured sugar. The H4 of both Gal types and the H6,6' of the Gal6S units, so, were localized in the same region and their overlapping caused in the HSQC spectrum a partial cancellation of the corresponding signals, due the different phase signs. The same happened to the carbon signals belonging to the C4 and C6 of the GalNAc6S unit in the HSQC spectrum of the CS of lesser spotted dogfish (see 2.3.3). The H4 of Gal6S was found at a higher δ value with respect to the one of the unsulfated unit due the presence of the sulfate group in position 6, that influences the neighbour protons and carbons too.

The chemical shift assignment of the protons H3, H3', H4 and H5 of the sialic acid was performed through strong COSY and TOCSY correlations. The other protons of the residue were not easily identified so, on the basis of literature data²⁷, they were collocated in the region between 3.6 and 3.8 ppm. The identification of the carbon signals was performed using the HMBC spectrum too, in particular the analysis of this spectrum led to the estimation of the carbon chemical shifts of the C1 and C2, quaternary carbons not detectable in the HSQC spectrum. The signals of the protons H6, H7 and H8-H8' were collocated in the crowded region around 3.7 ppm, while the corresponding carbon chemical shifts were not determinate.

Through the comparison with literature data available for bovine tracheal cartilage KS²⁸ a classification of the different sugar units was performed. In particular the units N^I, N^{II}, G^{II} and G^{III} were classified as units of the two internal blocks: the one sulfated only on the hexosamine unit and of that characterized by sulfate group at both sugars of the repeating unit. Moreover, on the basis of the NOE contacts found in the NOESY spectrum, the residues NI and GII belonging to the disulfate block, and the N^{II} and G^{III} units to the other one.

The unit G^I was classified as a Gal unit at the non reducing end capped at the position 3 by the NeuA unit (NA). The presence of a little α -glycosylation shift effect on the carbon C3 with the

²⁷ Huckerby, T.N., 2002 *Progress in nuclear magnetic resonance spectroscopy* 40:35-110.

²⁸ Huckerby, T.N., Lauder, R.M.2000 *EUR J. BIOCHEM.* 267:3360-3369.

respect to the one detected for the other unit of Gal confirmed the ketosidic nature of the linkage.

The presence of a α -(2→3) linkage between the NA unit and the G^I was confirmed by the comparison of the proton chemical shift of the sialic acid with literature data²⁹.

The unit F, was classified as the terminal Gal unit of the core oligosaccharide (Cap 1, Fig. 1.7), the absence of a sialic acid as capping sugar was determined on the basis of the carbon chemical shift of 73.78 ppm of the C3, peculiar for an unbound sugar.

The analysis of the HSQC spectrum, supported by the integration of some densities, has indicated a higher content of unsulfated Gal units with respect to the sulfate ones.

2.5 Conclusions

Cartilaginous fishes are a typical source of glycosaminoglycans chains, with particular attention for the chondroitin sulfate due to the presence of a large content of CS-Proteoglycans in their cartilaginous skeleton.

These structural studies have revealed that the CS isolated from the different tissue of some specimens of Mediterranean *Elasmobranches* had a similar composition. In particular the same four structural units were detected: CS-0S [4]-GlcA-(1→3)-GalNAc-(1-), CS-A [-4]-GlcA-(1→3)-GalNAc4S-(1-), CS-C [-4]-GlcA-(1→3)-GalNAc6S-(1-), and CS-D [-4]-GlcA2S-(1→3)-GalNAc6S-(1), but occurring in different percentages as reported in the table below

Tissue	CS0S	CS-A	CS-C	CS-D
<i>Scyliorhinus canicula</i> Cartilage	8.2%	41%	32%	19.8%
<i>Raja bochyura</i> Cartilage	9.5%	36.5%	39.0%	15%
<i>Raja bochyura</i> Skin	5.7%	37.1%	38.4%	18.8%
<i>Torpedo nobiliana</i> Cartilage	0.9%	30.9%	38.2%	30%

All the CS samples showed a high density of negative charge and among the four sample investigated, the one belonging to the torpedo specimen was the one with the higher charge density.

In all the cases atypical components like fucose, mannose or glucose or rare disulfated disaccharide units were not found and an amount of iduronic acid around 2% was estimated, so all the chains were not classifiable as CS/DS hybrid chains.

Compositional analyses of the CS samples were carried out using both typical protocols and new methods. The classical approach was based on enzymatic depolymerisations and HPLC

²⁹ Huckerby, T.N., Nieduszynski, I.A., Bayliss, M.T., Brown, G.M. 1999 *EUR J. BIOCHEM.* 266:1174-1183.

separations of the corresponding oligosaccharide mixtures. One of the new compositional methods applied was a spectroscopic one, named Q-HSQC, based on the integration of a restricted group of HSQC densities. This method was applied to the HSQC spectrum of the intact CS isolated from the cartilage *Scyliorhinus canicula* without the modification of the acquisition parameters and of the sequence and led to the estimation of compositional percentages similar to those obtained through the SAX-HPLC analysis. The other method was the RPIP-HPLC techniques that led to For the separation of the mixture of unsaturated disaccharides and of the mixture of glycopeptides belonging to the linkage region of the CS sample isolated again from the cartilage *Scyliorhinus canicula* the RPIP-HPLC technique was applied. The carbohydrate species, quantitatively separated through this technique, were directly analyzed by 2D-NMR, demonstrating that the ion-pairing reagents did not hinder the spectroscopic assignment of such molecules and modify the chemical shift values, as well.

These studies, moreover, led to a big collection of NMR data about intact CS, glycopeptides of the linkage region, disaccharides, trisaccharides and tetrasaccharides species with different sulfation patterns. In this way, the influence of the sulfate group on the chemical shift values was estimated for both intact and depolymerised CS.

Structural investigations are usually focused on the isolation and characterization of new and rare motifs with peculiar properties, like those able to promote the neurite outgrowth. In this work, rare or atypical motifs were no identified, but a complete characterization of natural CS material was performed. In this way, among the four CS samples, those isolated from the cartilage of *Scyliorhinus canicula* and of *Raja bochyura*, showed a sulfation pattern compatible with human physiology that can be eventually considered as an alternative natural source of CS.

The chondroitin sulfate is not the unique galactosaminoglycans isolated from fresh cartilaginous tissues, keratan sulfate is another component of proteoglycans of the cartilage. From the lesser spotted dogfish cartilage a little amount of this GAG was isolated as well. This molecule was characterized by a structure analogue to the one reported for shark KS.

2.6 Material and Methods

2.6.1 Proteolytic digestion of fresh tissues

Galactosaminoglycans were isolated from the fresh tissue as described in litterature². The dices tissue was suspended in a buffer solution (0.1 M of sodium acetate, 10 mM of cysteine-HCl and 2.4 mM EDTA, pH 6.8, 400 ml) and digested with papain (1U for 100 mg of tissue, 18 h at 65°C, papain from papaya latex, crude powder 2.2 units/mg solid, P3375 Sigma-Aldrich). The digestion was terminated denaturing the enzyme at 100°C for 15', the unreacted solid was removed by

centrifugation and the supernatant treated with 4 volumes of ethanol to allow the precipitation of the polysaccharide components (GAGs). The solution was cooled and left at 4°C overnight. The precipitate was collected by centrifugation (5000 rpm, 30 min., 4°C) suspended in 50 mM sodium acetate (10 ml), precipitated upon the addition of two volumes of ethanol, centrifuged and the resulting solid was precipitated again in the same conditions and freeze-dried. All the supernatants were pooled and were purified successively.

The amount of cartilage belonging to the lesser spotted dogfish specimens was 180 g and at the end of the extraction process 9 g of precipitate were recovered, so the yield respect to the cartilage tissue was 5.0%.

The amount of cartilage belonging to the skate specimen was 120 g and the precipitate recovered was 1.8 g, so the resulting yield was 15%. The amount of skin was 90 g and the yield of precipitate was 7%.

The amount of cartilage of torpedo specimens was 35 g and the resulting precipitate was 0,4 g with a yield of 12%.

2.6.2 Anionic exchange chromatography: Q-sepharose

An aliquot of the solid was purified via anion exchange chromatography using Q-Sepharose fast flow as adsorbent (Amersham Biosciences). This support was made up by cross-linked agarose (6%) and had these characteristics: total ionic capacity of 0.18–0.25 mmol/ml media, exclusion limit of 4×10^6 Da for globular proteins, flow rate supported 400–700 cm/h and a chemical stability for all commonly used aqueous buffers.

The column size was 2.5 x 18 cm and the flow rate applied was 3 ml/min. Gel was equilibrated with 10 mM NaCl and sample was eluted increasing stepwise NaCl concentration (100, 200, 400, 700 and 1000 mM), each solution covered twice the column volume and the eluate was monitored reading the absorbance at 220 nm. Peaks containing carbohydrate material were visualized according to the phenol assay³⁰, and were found at high ionic strength; the most abundant fractions were eluted with 700 mM NaCl or 1 M buffers. Fractions from the 400, 700 and 1000 mM elutions were pooled accordingly, dialyzed and freeze-dried.

450 mg of solid belonging to the cartilage of lesser spotted dogfish were purified in this way and the main amount of polysaccharide was achieved by the elution with 700 mM buffer. The amount of purified chondroitin sulfate polysaccharide was 150 mg, so the resulting yield scaled with respect to cartilage tissue was 1.5%.

Starting from 600 mg of precipitate from the cartilage of skate, 120 mg of purified CS were recovered from the fractions eluted at the higher ionic strength (1000 mM). The yield of purified

³⁰ Taylor, K.A.C.C, 1995 *Applied Biochemistry and Biotechnology* 53:207-214.

material with respect to the fresh tissue was 3%. All the precipitate obtained from the skin of skate (640 mg) was purified and the amount of purified CS was 180 mg, with a yield of 2%. Also in this case, the polyanionic material was collected through the elution with NaCl 1M buffer. The precipitate obtained from the cartilage of torpedo specimen (0,4 g) was purified and the corresponding amount of pure polysaccharide was 140 mg, with a yield of 3.5%.

Keratan sulfate material was isolated from the cartilage of the lesser spotted dogfish too. In particular it was recovered from the fractions belonging to the elution with NaCl 700 nM and 1 M of the Q-sepharose purification of the supernatant isolated after the extraction process of cartilage tissue.

2.6.3 Chemical Analysis conditions

Monosaccharides were analyzed as acetylated O-methylglycoside derivatives as reported¹⁹. GC-MS analyses were performed with an Agilent 5973 instrument, using a SPB-5 capillary column (Supelco, 30 m x 0.25 i.d. flow rate 0.8 ml/min, He as carrier gas). Mass spectra were recorded at ionization energy of 70 eV and an ionizing current of 0.2 mA. The temperature program used for the analysis of the mixture of acetylated O-methylglycoside derivatives was: 150°C for 3 minutes, 150→300°C at 10°C/min, 300°C for 12 minutes.

2.6.4 SDS-PAGE conditions

The four different CS samples were screened by discontinuous SDS-PAGE (Sodium Dodecyl Sulphate polyacrylamide Electrophoresis), using a 8% gel. Samples were run at constant voltage (150V) on a miniprotean gel system of Bio-Rad.

The gel was at first stained with alcian blue dye³¹ and then stained according to the procedure of Kittelberger and Hilbink³².

For the electrophoretic screening 4 µl of CS solutions (1 mg/ml) were used. Solution of sulfate dextrans (9-20 KDa, 500 KDa) were used as molecular weight standards.

2.6.5 Enzymatic degradations

2.6.5.1 ABC/ AC and B lyase digestions

An amount of CS ranging between 40-100 mg (60 mg of lesser spotted dogfish CS, 11 mg of the cartilage of skate, 87 mg of the skin of skate, 28 mg of cartilage of torpedo) was solved in Tris-HCl buffer (0.1 M, pH 8.0, 20 mg/ml) and depolymerised with ABC lyase at 37°C for 18 h. The amount of enzyme used for the depolymerisation was established on the basis of the proportion

³¹ Cowman, M.K., Slahetka, M.F., Hittner, D.M., Kim, J., Forino, M., Gadelrab, G. 1984 *Biochem. J.* 221:707-716

³² Kittelberger, R., Hilbink, F., J. 1993 *Biochem. Biophys. Methods* 26:81-86.

1U/100 mg chondroitin (1 U is defined as the amount of enzyme that catalyze the formation of 1 μ mol of unsaturated disaccharides from CS per minute at 37°C). The ABC lyase (C2905 Sigma-Aldrich) was a bacterial lyase isolated *Proteus vulgaris* available as lyophilized powder (0.85 units/mg solid), stabilized with BSA.

After digestion, ABC lyase was heat inactivated (100°C, 15 min) and the solution treated with trypsin to degrade ABC lyase and other proteic components eventually present. Finally, trypsin was heat inactivated as above, and the solution cooled and freeze-dried. The corresponding mixture of oligosaccharides was purified by SEC chromatography.

15 mg of CS samples isolated from skate and torpedo were treated with 0.25 U of AC lyase (1U/100 mg chondroitin, AC lyase from *Flavobacterium heparinum* lyophilized, 0.85 units/mg solid, C2905 Sigma-Aldrich) in the same condition reported above. The resulting oligosaccharides were fractionated by SEC chromatography.

The digestions with B lyase (from *Flavobacterium heparinum* lyophilized, 0.85 units/mg solid, C2905 Sigma-Aldrich) was performed on 10 mg/ml solutions of the CS samples isolated by skate and torpedo using 5 U of enzyme. The digestion products were analyzed by NMR without any purification step.

2.6.5.2 Protease digestion

The sample was dissolved in Tris-HCl buffer (0.1 M, pH 8.0) and treated with protease type XIV (bacterial protease from *Streptomyces griseus* P5147 Sigma-Aldrich) for 16 h at 37°C. The enzyme (5 mg) was added three times, at regular intervals.

The enzyme was inactivated by heating as above, and the solution cooled, freeze-dried and then chromatographically purified.

2.6.6 Alkaline hydrolysis in reductive condition

The protocol applied is the one typical for the O-linked glycans²⁶.

The samples were dissolved in a volume of NaOH 0.1 M, NaBH₄ 2 M to obtain a final concentration of 10 mg/ml. The reaction was carried out at 45°C for 18 h. The excess of NaBH₄ was eliminated with the addition of acetic acid and the solution was purified by dialysis (cut off 3500 Da) or by SEC chromatography.

2.6.7 Size Exclusion Chromatography (SEC) conditions

2.6.7.1 BioGel supports

Bio-Gel P gels are porous polyacrylamide beads prepared by copolymerization of acrylamide and N-N'-methylene-bis-acrylamide. These supports are extremely hydrophobic and essentially free

of charge and provide efficient, gentle gel filtration of sensitive compounds. Bio-Gel P gels are compatible with diluted organic acids and bases, with chaotropic agents, reducing reagents, detergents and miscible organic solvents (<20%). For the separation, however, buffer of >50 mM ionic strength are recommended.

The support used in all the SEC-separation was a Bio-Gel P2, due the high efficiency in rapid carbohydrate and small peptide separations and desalting. The fractionation range is 100-1800 Da.

The separations were performed using a column of 180 ml (1.5 x 116 cm) using 50 mM ammonium bicarbonate as buffer at a flow rate of 0.2 ml/min. The column eluate, each time, was monitored in continuous with a refractive index refractometer (K-2310 Knauer) and carbohydrate containing fractions were identified with the phenol test or UV measurements.

The chromatographic separation on BioGel P2 of CS oligosaccharides obtained through ABC lyase digestions in all the cases has led to the isolation of two main carbohydrate fractions, one in the void volume column (Fr-A) and another one (Fr-B). The yield of two fractions A and B were every time around the 15% and 75% of the starting polysaccharide material.

The separation of AC products in all the cases has led to the isolation of five different fractions, the biggest one in the void volume of the column (Fr-A) and other four (Fr-B-E). The yields of those fractions with respect to the starting polysaccharide material were around : 35% for Fr.A, 15% for Fr.B, 10% for Fr.C, 10% for Fr.D and 5% for Fr.E.

2.6.7.2 Sephacryl supports

Sephacryl supports are porous beads of allyl-dextran cross-linked with N-N'-methylene bisacrilamide. They are stable to all commonly used buffers.

The support used in the separation of KS sample was a S-100 HR, characterized by a fractionation range, estimated for globular proteins, of 1-100 kDa.

The separation was performed on a column of 120 ml (1.5 x 65 cm), using 50 mM NH_4HCO_3 buffer and a flow rate of 0.2 ml/min. The chromatographic separation was monitored in continuous with a refractive index refractometer (K-2310 Knauer) and carbohydrate containing fractions were identified with the phenol test.

2.6.8 HPLC analyses

All the HPLC analyses were performed on a Agilent 1100 HPLC instrument, equipped with a binary pump and a UV/VIS detector and a RID detector.

2.6.8.1 Strong Anionic Exchange conditions

Each disaccharide mixture was analyzed by strong anion exchange chromatography⁶, using SUPELCO LC-SAX (4.6 x 250 mm) column equilibrated with 10 mM NaCl, oligosaccharides were eluted with a linear NaCl gradient (5 min at 10 mM NaCl then to 200 mM NaCl over 30min, additional 15min in isocratic at 200 mM) and the elution profile was monitored reading at 232 nm.

Each fraction A, belonging to the AC lyase digestions of cartilage material was analyzed by strong anion exchange chromatography as well.

Fr-A belonging to skate and torpedo cartilage: after the equilibration of the column with 10 mM NaCl, oligosaccharides were eluted with a linear NaCl gradient (from 10 mM NaCl to 400 mM NaCl over 40min, additional 10min in isocratic at 1000 mM NaCl) and the elution profile was monitored reading at 232 nm.

Fr-A generated from the skate skin : column equilibration at 100 mM NaCl, then from 100 mM NaCl to 400 mM over 90 min and additional 10 min in isocratic at 1000 mM NaCl.

2.6.8.2 Reverse Phase Ion-Pairing conditions

RPIP separations⁹ were performed on a 5 μ m Supelco RP-C8 column (4.6 x 250 mm) and eluate was monitored at 232 nm, and solvent gradient was optimized for each sample.

Fr-1: eluent A was water/acetonitrile (9:1) and the eluent B was water/acetonitrile (65:35), tributylamine (15 mM) and ammonium acetate (50 mM) were added to both eluents, and the pH was adjusted to 7.0 with acid acetic. Sample (10 mg/ml, 100 μ l each injection) was separated varying buffer B (flow rate 0.8 ml/min.) as follows: from 0 to 50% in 15 min., then 100% for 5 min.

Fr-2: eluents were the same as above except that acetonitrile in buffer A was 2 %. Sample (10 mg/ml, 100 μ l each injection) was separated varying buffer B as follows: from 0 to 30% in 15 min, from 30% to 60% in 5min., holding at 60% for other 5 min. at a flow rate of 0.8 mL/min was used for elution.

Peaks were collected from RPIP chromatography, acetonitrile was removed with a rotary evaporator and the solution was freeze-dried.

2.6.8.3 Size Exclusion conditions

The molecular weight estimations was performed using a TSK-GEL G500 PW_{XL} (Tosoh Bioscience, 7.8 x 300 mm, 10 μ m) characterized by a fractionation range of 50-250 kDa, estimated for dextrans. A guard column (Tosoh Bioscience, 6 x 40 mm) was applied too.

The analytical runs were performed using 50 mM NH_4HCO_3 as eluent at a flow rate of 1 ml/min. Isocratic conditions were applied and the elutions were monitored with both UV (220 nm) and RID detectors. 10 μl of sample and molecular weight standards solutions (1 mg/ml) were used for each injection.

of the intact chondroitin sulfate and of the two glycopeptides (Fr-1A and Fr-1B),

2.6.9 NMR analysis

^1H and ^1H - ^{13}C NMR experiments were carried out on a Bruker DRX-600 equipped with a cryogenic probe, spectra were calibrated with respect to internal acetone ($\delta_{\text{H}} = 2.225$ ppm; $\delta_{\text{C}} = 31.45$ ppm). For all the homonuclear spectra, experiments were measured with data sets of 2048 x 512 points, a mixing time of 200 and 120 ms was employed for NOESY and TOCSY, respectively. The mixing time applied for the TROESY experiments was 300 ms.

Each data matrix was zero-filled in both dimensions to give a matrix of 4096 x 2048 points and was resolution-enhanced in both dimensions by a shifted sine-bell function before Fourier transformation.

The HSQC and HMBC experiments were measured using a data set of 2048 x 512 points, 64 scans were acquired for each t_1 value, and pulse sequence was optimized for a 150 Hz coupling constant. During processing, matrix was extended to 4096 x 1024 points by forward linear prediction extrapolation.

The temperature of each 1D and 2D experiment was selected in order to move the signal of the residual HOD in a region free of signal of interest.

All NMR spectra were acquired, transformed and analyzed with Topspin 2.1 program.

^1H and ^1H - ^{13}C NMR experiments of the unsaturated disaccharides, were carried out on a Bruker DRX-600 equipped with a z-gradient reverse probe in the same conditions listed above.

CHAPTER 3

Molecular Modelling on glycosaminoglycans oligosaccharides

3.1 Conformational studies on a decasaccharide of hyaluronic acid with the support of Residual Dipolar Coupling¹

Hyaluronic acid is a polysaccharide constantly under study, and many efforts are dedicated to the description of its three-dimensional structure, a key parameter for the elucidation of its rheological properties² or its interaction mechanism with hyaloadherins³.

The detailed structure of this macromolecule has been defined at atomic level by crystallographic data on polysaccharide fibers⁴, or by the combined use of molecular modelling and NMR spectroscopy.

The detailed structure study, described below, was carried out through the combined analysis of data collected by molecular modelling and experimental data collected by a particular NMR spectroscopy technique based on the estimation of the Residual Dipolar Coupling (RDC) constants encoding for long-range structural information.

3.1.1 The Residual Dipolar Coupling⁵

The structural characterization of oligosaccharides in aqueous solutions, through traditional NMR methods, is based on the measurements of NOEs and scalar couplings. The r^{-6} dependence of NOEs allows detection of NOEs only between protons at a distance up to 5 Å. Scalar couplings give dihedral angle restraints for only those groups of atoms separated by three bonds or less. These classical structural data, consequently, are short range and local. The dipolar couplings, however, are through-space interactions that arise between two magnetically active nuclei and depend on their distance, orientation and dynamics. In aqueous solution the possibility

¹ Manuscript in preparation

² Hardingham, T. Solution properties of Hyaluronan, in Chemistry and Biology of Hyaluronan Eds: Gang, H.G., Hales, C.A. Elsevier Ltd. chap. 1:1-19.

³ Day, R.M., Mascarenhas, M.M. Signal Transduction Associated with Hyaluronan, in Chemistry and Biology of Hyaluronan, Eds: Gang, H.G., Hales, C.A Elsevier Ltd., chap. 7:153-188.

⁴ Structures of Polyoligosaccharides, in Conformation of Carbohydrates Eds: Rao, V.S.R., Qasba, P.K, Balaji, P.V., Chandrasekaran, R., Harwood Academic Publishers, Amsterdam, Netherland, chap. 8:255-275.

⁵ Lipsitz, R.S., Tjandra, N. 2004 *Annu. Rev. Biophys. Biomol. Struct.* 33 :387-413.

of a direct measurement of angular information disappears due to the Brownian motions of the molecules and the distance dependence can be only estimated indirectly through relaxation phenomena. To extract complete dipolar coupling data, the molecule must behave anisotropically, situation met using a cosolute able to cause a partial alignment with respect to the external magnetic field.

In an oriented medium, the tumbling of the molecule is fast enough to preserve the high resolution of the spectra, and the dipolar couplings do not completely average to zero, so a residual value of them can be estimated.

The anisotropic conditions are achieved dissolving the molecules in a special liquid crystal, able to induce a weak degree of alignment of the solute with respect to the static magnetic field. The alignment effect is the result of steric and electrostatic interactions of the solute with the liquid crystal. When neutral media are used the electrostatic contribution can be neglected, so the alignment reflects the asymmetries in the shape of the molecule and its tensor of inertia encodes the structural information⁶. On the basis of this correspondence, a method named TRAMITE (Tracking Alignment from the Moment of Inertia Tensor) has been developed for the prediction of dipolar couplings.

In TRAMITE method, the three principal axes of the alignment tensor are chosen to correspond to the three axes of the inertia tensor (i.e., their eigenvectors are parallel)⁶. The three eigenvalues of the inertia tensor, (I_x, I_y, I_z) , were then normalized to $I_x + I_y + I_z = 1$ to represent the probability ellipsoid for the alignment. The three alignment magnitudes A_x, A_y , and A_z were obtained by subtraction of $1/3$ to the normalized I_x, I_y , and I_z value, to satisfy the null-trace condition $A_x + A_y + A_z = 0$ ⁷.

The calculation of the average RDC from an MD trajectory proceeds as follows: for each conformation in an MD trajectory, the inertia tensor is calculated and diagonalized. The three inertia tensor eigenvectors were ordered by the size of the corresponding eigenvalue $I_z \geq I_x \geq I_y$. The molecule is then rotated so as to make the principal axis system of the alignment parallel to the laboratory frame. For a conformer in this orientation, the calculation of the instantaneous RDC for any spin vector, $r (r_x, r_y, r_z)$, is given by the equation below⁷:

$$D = \frac{k' \cdot k_{\max}}{R^3} \left(\cos^2 \theta \frac{3}{2} \right) = \frac{k' \cdot k_{\max}}{R^3} A_x r_x^2 + A_y r_y^2 + A_z r_z^2$$

where D is the expected RDC, $\cos^2 \theta$ is the angle of the vector with the alignment frame, k_{\max} is a constant that depends on the type of RDC, R is the C-H distance, and k' is a scaling factor,

⁶ Azurmendi, H. F., Bush, C. A., 2002, *J. Am. Chem. Soc. (Communications)* 124 :2426-2427.

⁷ Kramer, F., Deshmukh, M.V., Kessler, H., Glaser, S.J. 2004 *Concepts NMR* 21A:10-21.

which depends on the particular system and the experimental liquid crystal concentration. This scaling factor was optimized for each conformation of the Molecular Dynamics simulation to reach the maximum agreement with the experimental RDCs.

3.1.2 Isolation of the oligosaccharide and collection of spectroscopic data

The deca-saccharide was obtained applying the conditions of depolymerisation and chromatographic purification reported⁸. The concentration of both HA and HAse, and the ion strength used during hydrolysis catalyzed by hyaluronidases were optimal to obtain HA fragments with a molar mass ranging between 2000, and 5000 g/mol.

After the chromatographic purification, the oligosaccharide showed a purity of 98%, as indicated by the MS spectroscopic analysis (data not shown).

First, the deca-saccharide underwent a classical spectroscopic analysis by one-dimensional, and bi-dimensional NMR in isotropic conditions. The study of the homonuclear, and heteronuclear 2D-NMR spectra led to the complete assignment of all the proton, and carbon chemical shifts (Table 1.1).

Table 3.1: (600 MHz, 288 K) Proton, and carbon (italic) chemical shift attribution of the HA₁₀ in isotropic condition. Chemical shifts are expressed in δ relative to internal acetone (¹H at 2.225 ppm, ¹³C at 31.5 ppm). Proton, and carbon chemical shifts of methyl group of the *N*-Acetyl are 2.03, and 23.5 ppm.

	1	2	3	4	5	6
3)-α-GlcNAc	5.15	4.05	3.90	3.56	3.88	3.77-3.91
L'	<i>93.8</i>	<i>55.8</i>	<i>82.8</i>	<i>71.3</i>	<i>74.1</i>	<i>63.2</i>
3)-β-GlcNAc	4.71	3.82	3.72	3.52	3.48	3.77-3.91
L	<i>97.6</i>	<i>58.24/57.2</i>	<i>85.5</i>	<i>71.2</i>	<i>78.1</i>	<i>63.2</i>
3)-β-GlcNAc-(1\rightarrow	4.54	3.85	3.71	3.54	3.48	3.77-3.91
B,D,F,H	<i>103.4</i>	<i>57.1</i>	<i>85.4</i>	<i>71.2</i>	<i>78.2</i>	<i>63.2</i>
4)-β-GlcA-(1\rightarrow	4.51	3.36	3.59	3.72	3.72	--
C,E,G	<i>105.8</i>	<i>75.3</i>	<i>76.4</i>	<i>82.8</i>	<i>79.1/78.7</i>	--
4)-β-GlcA-(1\rightarrow	4.47	3.34	3.58	3.74	3.73	--
I	<i>105.9</i>	<i>75.3</i>	<i>76.4</i>	<i>82.8</i>	<i>79.1</i>	--
<i>t</i>-β-GlcA-(1\rightarrow	4.47	3.32	3.50	3.73	Ca3.73	--
A	<i>105.9</i>	<i>75.3</i>	<i>74.5</i>	<i>82.8</i>	<i>79.1/78.7</i>	--

Chemical shift assignment was easily performed for all the residues due to the presence of strong COSY and TOCSY correlations, as a consequence of the *gluco* configuration of all the units. The

⁸ Tranchepain, F., Deschrevel, B., Courel, M.N., Levasseur, N., Le Cerf, D., Loutelier-Bourhis, C., Vincent, 2006 *J. Anal. Chem.* 348:232-242.

units labelled as L' and L were the two forms (α and β) of the reducing GlcNAc unit, the unit I and A were classified respectively as the GlcA unit linked to the reducing GlcNAc and the GlcA unit at the non reducing end. The other spin systems B, D, F, H and C, E, G were classified as internal units. All these spectroscopic data and the classification of the different units agreed with those reported^{9,10}. The analysis of the NOESY spectrum (Fig. 3.1) confirmed the linkages between the units I and L, A and B, and between the internal units in accordance with the HA disaccharide repeating unit. Moreover, the presence of the typical intra-residual NOE contacts 1-3 and 1-5, typical of a β -configured sugar, indicated a 4C_1 conformation for all the residues.

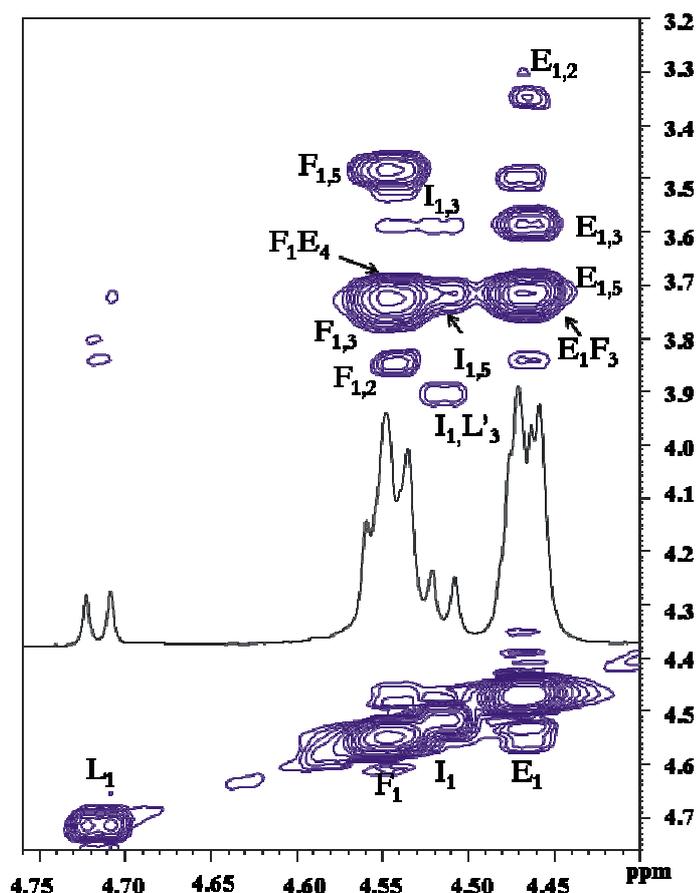


Figure 3.1: (D_2O , 600 MHz, 288K) expansion of the anomeric region of HA deca-saccharide NOESY spectrum. The principal NOE contacts are reported. Only E and F units are reported as internal units.

In order to perform a spectroscopic analysis on the oligosaccharide in anisotropic conditions, namely to measure Residual Dipolar Coupling constants, the deca-saccharide was dissolved in a ternary mixture of PBS, pentaethylene glycol mono-octyl ether (C_8E_5), and *n*-octanol with a ratio C_8E_5 / water of 3%, and a molar ratio of C_mE_n to *n*-alkyl alcohol of 0.87¹¹.

⁹ Blundell, A.D., Reed, M.A.C., Almond, A. 2006 *Carbohydr. Res.*, 341:2803-2815.

¹⁰ Blundell, C.D., Almond, A. 2007 *Magn. Reson. Chem.* 45:430-433.

¹¹ Rückert, M., Otting, G. 2000 *J. Am. Chem. Soc.* 122:7793-7797.

This kind of liquid crystalline medium was selected because it was uncharged, insensitive to pH, poorly sensitive to salt, with little binding capacity to macromolecules, and stable under a good range of temperature (0-40°C).

The mixture of C₈E₅/*n*-octanol was selected to perform 2D experiments at 288 K, due to the presence of a H1 signal of L unit, that otherwise was covered by the residual HOD signal.

The percentage of 3% for the ratio C₈E₅/ water was selected on the basis of the resulting quadrupolar splitting of 14.8 Hz in the ²H-NMR spectrum (data not shown).

The estimation of the residual dipolar carbon-proton couplings (¹D_{C,H}) was performed through the analysis of the coupled-HSQC spectra, recorded in both isotropic and anisotropic conditions. In particular the ¹D_{C,H} values were calculated by subtracting to the coupling constants measured in anisotropic conditions those estimated in the isotropic medium.

Analysis of the 2D-NMR spectra of the deca-saccharide led to the determination of only a big part of all possible ¹D_{CH} values (Table 3.3). In particular the overlap of the C6-densities of all the GlcNAc units with those of the alkyl chains of the medium prevented such determination. Moreover, the similarity of the carbon chemical shifts of the C2 of GlcA units and of C5 of the GlcNAc units, did not lead a correct measurement of the corresponding ¹D_{CH}.

For the GlcA units were calculated the ¹D_{C,H} for the carbons C1, C3, C4 and C5, for the GlcNAc units those of the C1, C2, C3 and C4 (Table 3.3).

These experimental data were then compared with those generated by an ensemble of conformers obtained through molecular dynamics calculation. This comparison led to select a group of structures that better reproduced the experimental values.

3.1.3 Molecular mechanics analysis

Both the disaccharides β-GlcA-(1→3)-β-GlcNAc and β-GlcNAc-(1→4)-β-GlcA were analyzed through molecular mechanics calculations to estimate the preferred Φ and Ψ of the anomeric torsions.

The Φ and Ψ dihedral angles were defined as H₁-C₁-O-C_n' and H_n'-C_n'-O-C₁.

The two structures were built up taking the *N*-Acetyl group in the typical *trans* disposition, the carboxylic group in the anionic form and the CH₂OH group in the *gt* conformation¹². After a minimization step they underwent a dihedral driving using the force field AMBER, implemented in Macromodel¹³ 8.0.

¹² Rockwell, G.D., Grindley, T.B. 1998 *J. Am. Chem. Soc.* 120:10953-10963.

¹³ Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., Still, W.C. 1990 *J. Comput. Chem.*, 11:440-446.

Both the glycosidic linkages displayed similar flexible maps (Fig. 3.2) and the low energy conformation agreed with the *exo*-anomeric effect.

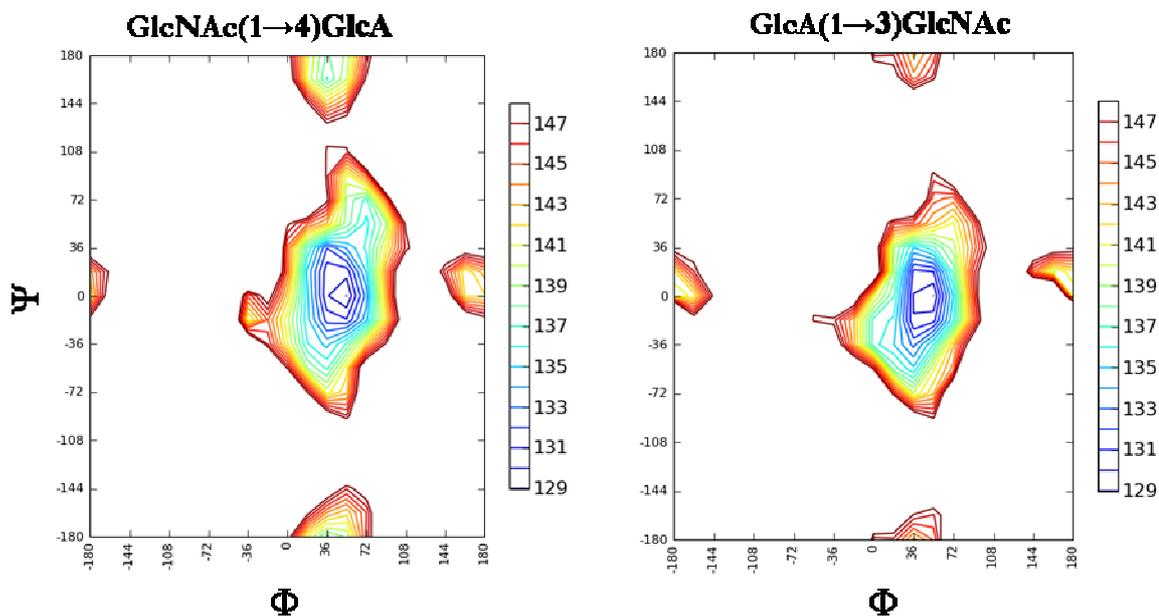


Figure 3.2: Flexible maps of the two glycosidic linkages of the HA backbone.

For the linkage β -(1 \rightarrow 4) the energy value associated with the minimum was 128.1 kJ/mol and the Φ and Ψ values were respectively 47.6 and -1.5 degrees. For the other disaccharide the energy value was 128.3 kJ/mol and the Φ and Ψ values were respectively 48.8 and 1.0 degrees. All these data agreed with those reported in the literature, for disaccharide of hyaluronan¹⁴.

3.1.4 Molecular Dynamics simulation in explicit water

Molecular dynamics (MD) simulation on the dodecasaccharide in explicit water was performed using the software AMBER¹⁵ 9 and applying the Glycam06e¹⁶ as force field.

The optimal values of Φ and Ψ estimated through molecular mechanics calculations were used to built up the decasaccharide. Adding 6859 molecules of water (TIP3P type of AMBER 9 software package) the water box (11 Å) was constructed around the molecule. The truncated octahedral type was selected for the box, because it gives an even distribution of the solvent around the solute.

¹⁴ Almond, A., Sheehan, J.K., Brass, A., 1997 *Glycobiology* 7:597-604.

¹⁵ Case, D.A., Darden, T.A Cheatham, T.E., Simmerling, C.L. Wang, J. Duke, R.E. Luo, R. Merz, K.M. Pearlman, D.A. Crowley, M. Walker, R.C. Zhang, W. Wang, B. Hayik, S. Roitberg, A. Seabra, G. Wong, K.F. Paesani, F. Wu, X. Brozell, S. Tsui, V. Gohlke, H. Yang, L. Tan, C. Mongan, J. Hornak, V. Cui, G. Beroza, P. Mathews, D.H. Schafmeister, C. Ross, W.S. and Kollman P.A. 2006 AMBER 9, University of California, San Francisco.

¹⁶ Kirshner, K.N., Yongye, A.B., Tschampel, S.M., Gonzales-Outeirino, J., Daniels, C.R., Foley, B.L., Woods, R.J. 2008 *J. Comput. Chem.* 29:622-655.

The system, after equilibration, was heated to 288 K and maintained to that temperature during the whole MD simulation. Periodic boundary conditions were applied to the system over 3.5 ns of simulation, the pressure was maintained constant at 1 atm and the box was free to relax around the molecule.

At the end of the 3.5 ns simulation, thermodynamic parameters were checked and showed the expected behaviour (data not shown) and then analysis of the simulation data was carried out.

In order to establish the goodness of the simulation data, a check of torsional angles and inter-protonic distances was performed, together with a comparison with experimental data collected in isotropic conditions.

All the residues during the 3,5 ns simulation assumed the typical 4C_1 conformation as demonstrated by the average values around 2.65 Å and 2.55 Å estimated respectively between the protons 1-3 and 1-5 of each ring. In this way the presence of the typical NOE contacts of β -configured sugars in the classical 4C_1 conformation detected in the NOESY spectrum was confirmed.

The *N*-Acetyl groups of all the GlcNAc residues were found in the typical *trans* disposition with respect to the proton in position 2 and the preferred rotamer C_5-C_6 of all the GlcNAc units was the *gt* one, as indicated by the value around -60° of the dihedral angle ω^6 , defined as $H_6-O_6-C_5-O_5$.

For all the glycosidic linkages a disposition in accordance with the *exo*-anomeric effect was detected, as demonstrated by the scattered plots in the Figure 3.3.

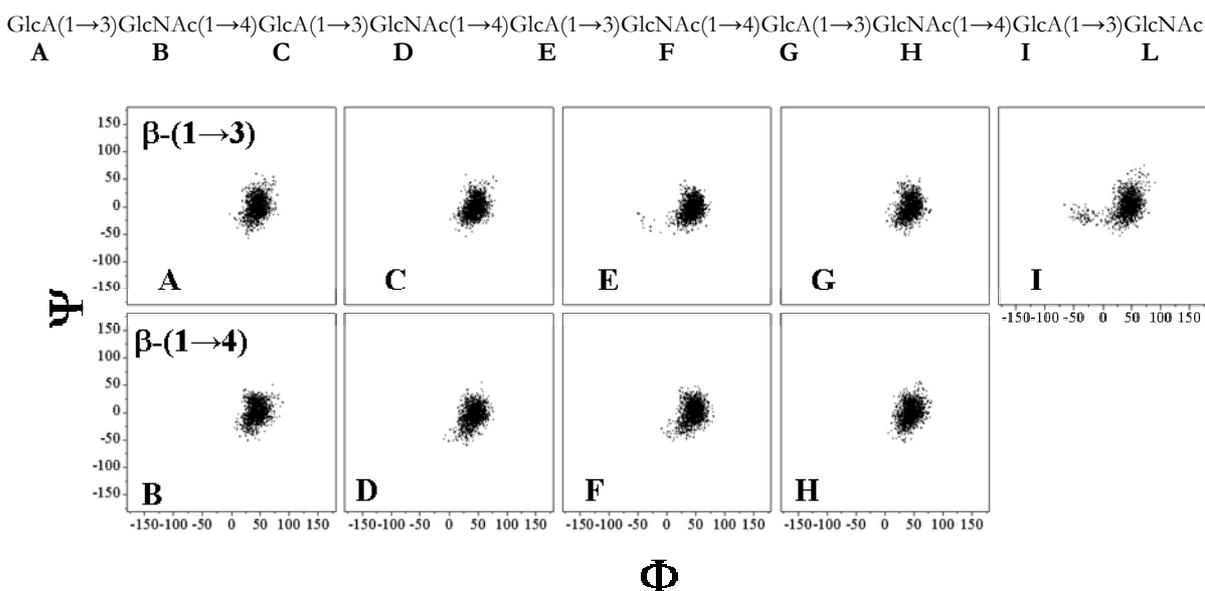


Figure 3.3: Scattered plots of the glycosidic torsions of the deca-saccharide. At the top the sequence of the deca-saccharide and the labels used are reported.

The averaged values for the dihedral angles Φ and Ψ of the β -(1 \rightarrow 3) linkage were respectively 44.99° and -0.77°, while those of the β -(1 \rightarrow 4) were 44.67° and -0.92°. These data confirmed those predicted by the molecular mechanics calculations.

The estimation of the inter-residual proton distances confirmed the data collected through the analysis of the NOESY spectrum. Distances lesser than 2.5 Å were found between protons across the glycosidic linkages confirming the presence of experimental strong NOE contacts. The absence of NOE contacts originated by no *exo*-anomeric disposition was established by the values higher than 4 Å of the corresponding inter-residual distances. All the values of the distances values and of the corresponding RMSD are listed in Table 3-2.

Table 3.2: Inter-protonic distances and corresponding RMSD.

NOE	Averaged distances (Å)
A1 \rightarrow B2	4.08 \pm 0.26
A1 \rightarrow B3	2.34 \pm 0.2
A1 \rightarrow B4	4.45 \pm 0.17
B1 \rightarrow C ₃	4.43 \pm 0.15
B1 \rightarrow C ₄	2.33 \pm 0.2
C ₁ \rightarrow D ₂	4.09 \pm 0.25
C ₁ \rightarrow D ₃	2.35 \pm 0.20
C ₁ \rightarrow D ₄	4.45 \pm 0.17
D ₁ \rightarrow E ₃	4.38 \pm 0.21
D ₁ \rightarrow E ₄	2.30 \pm 0.19
E ₁ \rightarrow F ₂	4.13 \pm 0.23
E ₁ \rightarrow F ₃	2.31 \pm 0.17
E ₁ \rightarrow F ₄	4.42 \pm 0.19
F1 \rightarrow G ₃	4.42 \pm 0.19
F1 \rightarrow G ₄	2.36 \pm 0.19
G ₁ \rightarrow H ₂	4.09 \pm 0.27
G ₁ \rightarrow H ₃	2.34 \pm 0.19
G ₁ \rightarrow H ₄	4.44 \pm 0.17
H ₁ \rightarrow I ₃	4.42 \pm 0.18
H ₁ \rightarrow I ₄	2.35 \pm 0.20
I ₁ \rightarrow L ₂	4.05 \pm 0.30
I ₁ \rightarrow L ₃	2.38 \pm 0.23
I ₁ \rightarrow L ₄	4.43 \pm 0.22

3.1.5 Analysis and refinement of simulation data

Starting from the collection of structures obtained through the molecular dynamic simulations in explicit solvent, the alignment tensor, the instantaneous RDCs of every snapshot and then the averaged RDCs over the all MD trajectory were determined (Tab 3.3). Mspin (Mestrelab research) software¹⁷, through the application of the method TRAMITE⁶, and the comparison

¹⁷ Mspin program information at: <http://mestrelab.com/Products/Mspin/Details.html>

with the experimental ${}^1D_{\text{CH}}$ allowed a refinement of the MD trajectory through the definition of a quality factor (Q): the lower the better.

Table 3.3: Experimental RDC values (D_{CH}) calculated as a difference between the carbon-proton coupling constants ${}^1J_{\text{CH}}$ measured in anisotropic conditions and those in isotropic condition, and the averaged theoretical values estimated through Mspin analysis among the 1750 MD structures.

Atom pair	Experimental D_{CH}	Theoretical D_{CH}
$\text{C}_1\text{-H}_1^{\text{A}}$	6.2 ± 1	1.14
$\text{C}_3\text{-H}_3^{\text{A}}$	13.54 ± 0.4	3.07
$\text{C}_4\text{-H}_4^{\text{A}}$	13.54 ± 0.8	-0.65
$\text{C}_5\text{-H}_5^{\text{A}}$	6.4 ± 0.2	1.86
$\text{C}_1\text{-H}_1^{\text{B}}$	9.5 ± 2	10.87
$\text{C}_2\text{-H}_2^{\text{B}}$	13.04 ± 1.3	10.51
$\text{C}_3\text{-H}_3^{\text{B}}$	15.8 ± 1.5	10.28
$\text{C}_4\text{-H}_4^{\text{B}}$	13.1 ± 1.3	10.28
$\text{C}_1\text{-H}_1^{\text{C}}$	6.2 ± 1	5.79
$\text{C}_3\text{-H}_3^{\text{C}}$	13.54 ± 0.4	8.20
$\text{C}_4\text{-H}_4^{\text{C}}$	13.54 ± 0.8	5.73
$\text{C}_5\text{-H}_5^{\text{C}}$	6.4 ± 0.2	7.07
$\text{C}_1\text{-H}_1^{\text{D}}$	9.5 ± 2	10.47
$\text{C}_2\text{-H}_2^{\text{D}}$	13.04 ± 1.3	10.02
$\text{C}_3\text{-H}_3^{\text{D}}$	15.8 ± 1.5	9.62
$\text{C}_4\text{-H}_4^{\text{D}}$	13.1 ± 1.3	10.46
$\text{C}_1\text{-H}_1^{\text{E}}$	6.2 ± 1	6.04
$\text{C}_3\text{-H}_3^{\text{E}}$	13.54 ± 0.4	8.32
$\text{C}_4\text{-H}_4^{\text{E}}$	13.54 ± 0.8	5.62
$\text{C}_5\text{-H}_5^{\text{E}}$	6.4 ± 0.2	7.25
$\text{C}_1\text{-H}_1^{\text{F}}$	9.5 ± 2	11.10
$\text{C}_2\text{-H}_2^{\text{F}}$	13.04 ± 1.3	10.66
$\text{C}_3\text{-H}_3^{\text{F}}$	15.8 ± 1.5	10.22
$\text{C}_4\text{-H}_4^{\text{F}}$	13.1 ± 1.3	10.92
$\text{C}_1\text{-H}_1^{\text{G}}$	6.2 ± 1	7.40
$\text{C}_3\text{-H}_3^{\text{G}}$	13.54 ± 0.4	9.13
$\text{C}_4\text{-H}_4^{\text{G}}$	13.54 ± 0.8	6.42
$\text{C}_5\text{-H}_5^{\text{G}}$	6.4 ± 0.2	8.22
$\text{C}_1\text{-H}_1^{\text{H}}$	9.5 ± 2	9.25
$\text{C}_2\text{-H}_2^{\text{H}}$	13.04 ± 1.3	9.01
$\text{C}_3\text{-H}_3^{\text{H}}$	15.8 ± 1.5	8.67
$\text{C}_4\text{-H}_4^{\text{H}}$	13.1 ± 1.3	9.74
$\text{C}_1\text{-H}_1^{\text{I}}$	6.2 ± 1	6.91
$\text{C}_3\text{-H}_3^{\text{I}}$	13.54 ± 0.4	8.23
$\text{C}_4\text{-H}_4^{\text{I}}$	13.54 ± 0.8	6.24
$\text{C}_5\text{-H}_5^{\text{I}}$	6.4 ± 0.2	7.64
$\text{C}_1\text{-H}_1^{\text{L}}$	8.4 ± 2	6.81
$\text{C}_3\text{-H}_3^{\text{L}}$	13.04 ± 1.3	6.32
$\text{C}_4\text{-H}_4^{\text{L}}$	15.8 ± 1.5	6.06
$\text{C}_5\text{-H}_5^{\text{L}}$	13.1 ± 1.3	7.29

On the basis of the experimental data, reported in Table 3.3, a quality factor of 0.6, over the all MD trajectory, was achieved. The software provided, also, a singular quality factor for each frame of the simulation and a selection of structures with a Q lesser than 0.5 was performed, collecting 30 % of the structures of the molecular dynamic simulation and decreasing the averaged Q factor to 0.4. In this way a subgroup of 571 structures (referred as MD_rdc) was created.

In order to understand the effect of the RDC filtering on the simulation data, the two conformer ensembles, referred MD_tot and MD_RDC, respectively, were compared, paying attention to the behaviour of the Φ , Ψ dihedrals and to molecular length.

Torsional angles (Φ , Ψ) of the RDC-filtrated data (MD_RDC) are reported in Table 3.4 together with those from the total MD ensemble (MD_tot).

Table 3.4: Averaged values of the dihedral angles (degrees) of the entire simulation (MD_tot) and of the RDC-filtrate subgroup (MD_RDC).

Angle	MD_tot	MD_RDC
Φ_A	45.9 ± 10.64	45.15 ± 11.13
Ψ_A	-0.29 ± 16.30	-5.61 ± 14.21
Φ_B	44.85 ± 11.34	45.05 ± 10.80
Ψ_B	1.07 ± 15.87	-0.87 ± 14.60
Φ_C	46.56 ± 11.14	43.62 ± 11.17
Ψ_C	-1.18 ± 15.47	-6.16 ± 13.29
Φ_D	43.47 ± 11.51	44.60 ± 10.88
Ψ_D	-4.52 ± 16.28	-2.33 ± 15.93
Φ_E	43.79 ± 12.31	42.28 ± 15.36
Ψ_E	-4.46 ± 14.88	-5.49 ± 15.09
Φ_F	45.89 ± 12.02	44.78 ± 12.72
Ψ_F	0.06 ± 16.66	-0.53 ± 17.50
Φ_G	44.72 ± 11.04	42.32 ± 11.90
Ψ_G	-0.59 ± 16.53	-7.38 ± 14.60
Φ_H	45.75 ± 11.37	46.65 ± 11.18
Ψ_H	0.29 ± 16.10	-0.88 ± 14.17
Φ_I	42.38 ± 20.45	41.77 ± 18.82
Ψ_I	1.95 ± 18.32	-0.98 ± 17.64

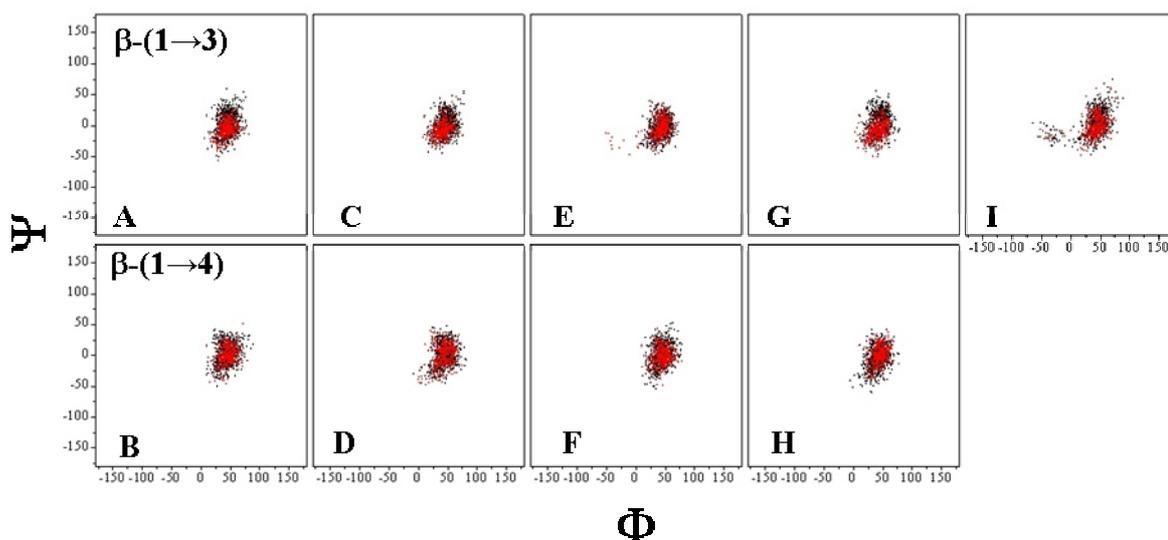


Figure 3.4 Superimposition as scattered plots of the (Φ, Ψ) values of each glycosidic linkage of the structures of the entire MD simulations (black points) and of the subgroup (red points).

Particular differences between the two sets of values were not detected, as confirmed also by the scattered plots in the figure 3.2, showing for both ensembles of structures one main large distribution with values always centred in the same area of the graphic. This datum indicated that apparently RDC did not lead to a selection of a specific conformer.

The effectiveness of the RDC filtration procedure appeared upon investigation of the length of the molecule. The molecular length was measured between the anomeric oxygen of the reducing glucosamine (unit L) and the oxygen at position four of the glucuronic acid at the non reducing end of the molecule (unit A). The graphic in Figure 3.5 showed that this parameter spanned of 38 to 52 Å.

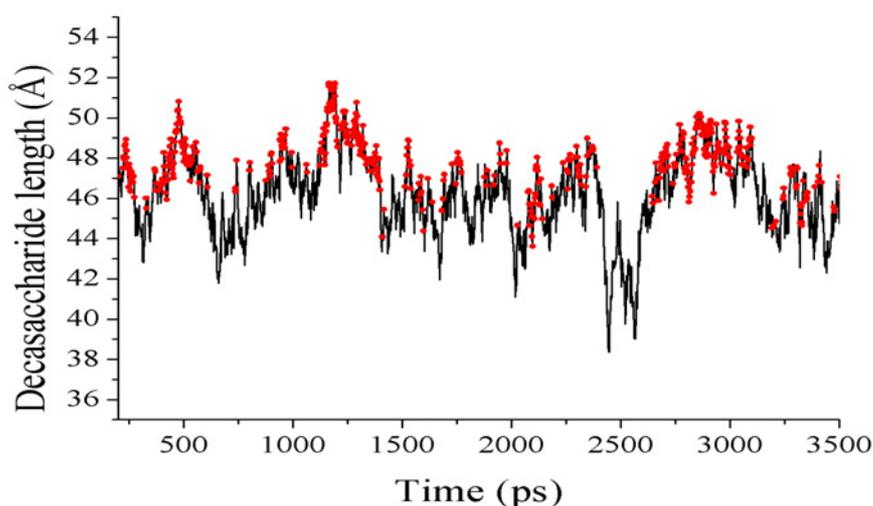


Figure 3.5: Comparison between the length of the structures of the entire MD simulation (black line) and the length of the structures of the subgroup (red points).

In particular, RDC-subgroup contained molecules with a rather elongated conformation in the range of 44-52 Å. This datum indicated that structures characterized by a length around 39-44 Å, namely characterized by a bent or compressed conformation, were not compatible with experimental restrains.

Dynamics simulation refinement through molecular length was a worthy choice, because it led to the selection of just some conformations among all the possible ones, namely not all the conformers were selected even though their length fell in the RDC selected range; this outcome reflected the fact that only specific combinations of the glycosidic junctions responded to the experimental constraints.

Additionally, it must be noted that, differently from NOE, RDC refinement took into account the whole molecule.

3.1.6 Helix parameters estimation

Analysis of the RDC filtrated ensemble become more quantitative describing with POLYS¹⁸ program the secondary structure of the conformers. This software, starting from dihedral angular values, evaluated the best fitting helix conformation which was described through the following parameters: h , the axial rise per repeating unit, and n , the number of unit per helical turn.

Starting from the average values of the dihedral angles of the internal junctions (between F and G for the β -(1 \rightarrow 3) and between E and F for the β -(1 \rightarrow 4)), without the optimization, the n values was not integer numbers, but 3.35, indicating that the corresponding conformation was of hybrid kind between a three- ($n=3$) and a four-folded ($n=4$) helix. The h value (not reported) was negative and indicated the occurrence of a left-handed coiling.

In this way it was established that two different left-handed helical structures were found to be compatible: a 3_2 geometry and a 4_2 one (Figure 3.6).

In the next step, the dihedral angles of each oligosaccharide were optimized fitting the molecule in a regular three- or four-folded helix; importantly, the optimized Φ , Ψ values (Tab. 3.5) fell in the dihedral range (averaged value and associated error) defined during the analysis (Table 3.4) revealing that both three- and four-folded helix geometries were possible.

¹⁸ Engelsens, S.B., Cros, S., Mackie, W., and Perez, S. 1996 *Biopolymers* 39:417-433.

Table 3.5: Averaged Φ and Ψ values of the inner junctions and optimized ones for the 3-fold and 4-fold geometries

Angle*	Averaged value (degree)	Optimized value (degree)	
		3_2	4_2
$\Phi_{1\rightarrow 3}$	-75.99 ± 11.20	-79.63	-70.28
$\Psi_{1\rightarrow 3}$	114.71 ± 16.78	111.46	119.79
$\Phi_{1\rightarrow 4}$	-73.78 ± 11.57	-78.28	-66.76
$\Psi_{1\rightarrow 4}$	-119.63 ± 14.79	-123.34	-113.83

* The dihedral angles Φ and Ψ were defined as $O_5-C_1-O-C_n$ and $C_1-O-C_n-H_{n+1}$, respectively

In addition, it was estimated that the extended conformation of the oligomer was possible with three folded helix geometry, whereas a more compact arrangement, namely a shorter molecular length value, could be accomplished only switching from a three- to a four-folded geometry.

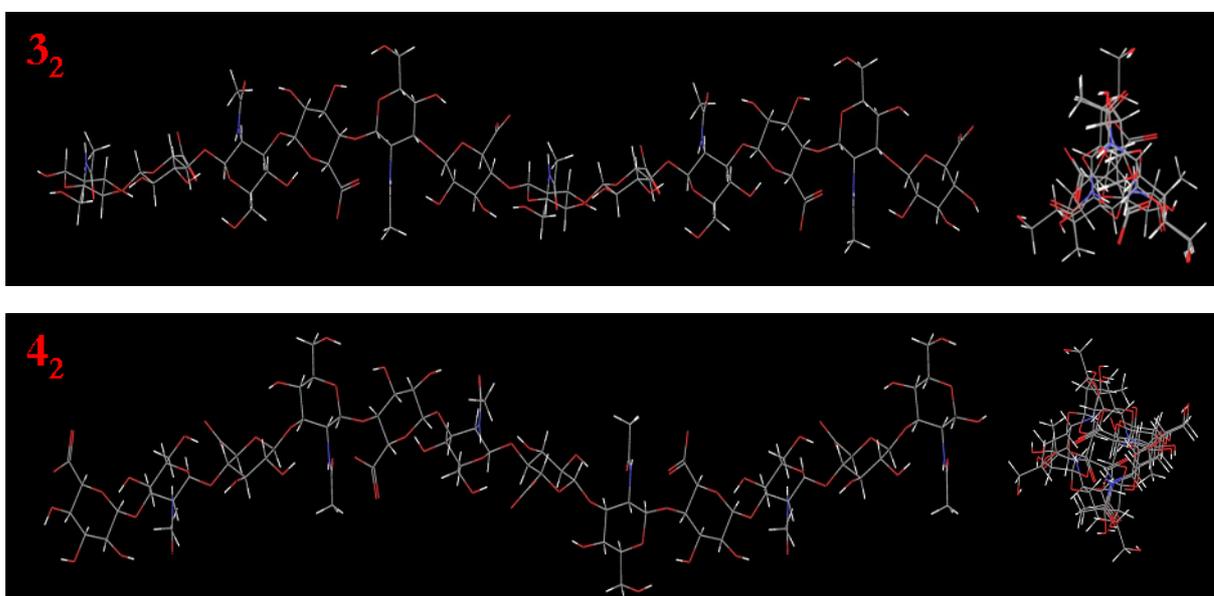


Figure 3.6: Helix structures generated with POLYS.

In order to ascertain the occurrence of both three and four-folded motifs in the RDC-filtrated ensemble, two helical structures in the pure 3_2 or 4_2 geometry (referred as HA_{n3} and HA_{n4} , respectively) were built up on the basis of the optimized dihedral angles and they were used as model in a comparison with the structures of the MD_rdc subgroup. The comparison was performed between the coordinates of the selected atoms of the reference model with those of the RDC ensemble. The divergence of the coordinated between one conformer and the model compound was expressed as RMSD and the smallest values were associated to a more stringent structural homology. The corresponding RMSD values of each frame are plotted in Figure 3.7.

Actually, since the conformers with the lowest RMSD values are those which show a higher similarity to the selected model, different values of RMSD were used as selection filter and the number of the oligomers within the limit reported in Table 3.6.

Table 3.6: Quantification of the conformers' amount with respect to their model and with respect to different values of the RMSD filter. Percentages in parenthesis are calculated with respect to the total RDC ensemble.

Model	RMSD 0.6	RMSD 0.75	RMSD 1.0
HA _{n3}	7	29 (5.1%)	143 (25%)
HA _{n4}	1	7 (1.2%)	56 (9.8%)

Clearly, small RMSD values (0.6) selected few conformers with a regular n-folded geometry, whereas higher values (1.0) defined a group more populated, but even more diverging from the model compound.

After the visual comparison of the RDC structures with the corresponding model compounds a value of 0.75 for RMSD was selected as reference: this value represented a good compromise since the number of conformers selected was reasonable while a good structural homology with the model compound was maintained.

As a result, the amount of structures in *pure helical conformation* within the RDC ensemble followed the proportion: HA_{n3} : HA_{n4} = 29 : 7, or nearly a ratio 4 : 1.

It must be noted that the pure three-folded helical conformations of the decamer accounted only for the 6.3% of the total ensemble, therefore it is worthy to retain that the majority of the conformers possessed a rather pronounced three-folded character, or that within each molecule, most of the glycosidic linkages were adopting the values relating to the three folded helix geometry whereas the other fell in the range of the four-folded motif.

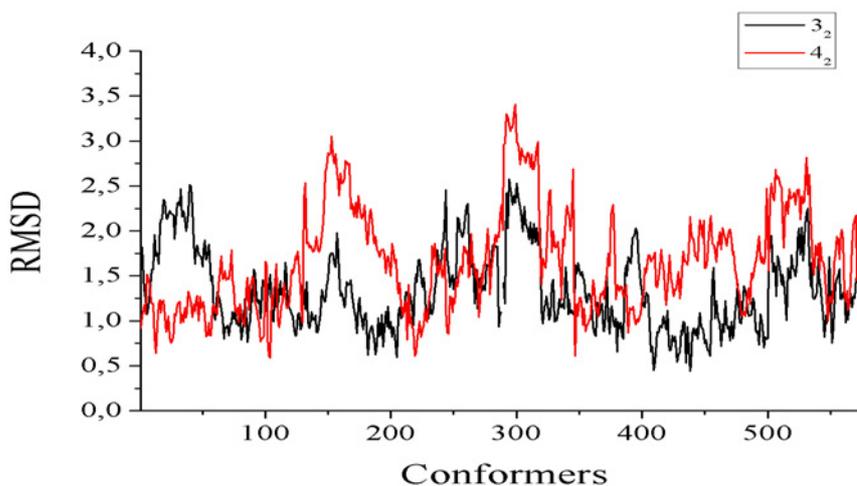


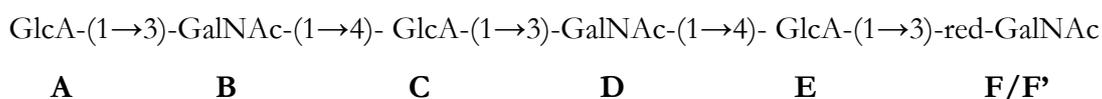
Figure 3.7: RMSD of the difference between the coordinates of each frame of the subgroup MD_RDC and of the two models HA_{n3} (black line) and HA_{n4} (red line).

3.2 MD simulation on an unsulfated hexasaccharide of chondroitin sulfate¹⁹

3.2.1 Isolation and NMR analysis of the hexasaccharide

The saturated hexasaccharide was isolated by SEC from an oligosaccharide mixture derived from the enzymatic depolymerisation of the de-fructosylated esopolysaccharide from *E. coli* K4²⁰.

The sequence of the isolated hexasaccharide and the labels used for each sugar ring are the following:



The oligosaccharide was first dissolved in D₂O to perform the typical sequence of 2D-NMR experiments necessary for the complete spectroscopic characterization of the molecule and then it was dissolved in a mixture of H₂O:D₂O (90:10) to study the behaviour of the NH protons too.

The combined study of the homonuclear (DQF-COSY, TOCSY, T-ROESY) and heteronuclear (HSQC and HMBC) spectra allowed the complete chemical shift assignment of the molecule (Table 3.7).

All the spectroscopic data of the internal GalNAc and GlcA units agreed with those reported for unsulfated unit of chondroitin sulfate polysaccharide²¹, while those of the GlcA unit of the non reducing end and of the reducing GalNAc units agreed with NMR data of unsulfated chondroitin sulfate disaccharide²².

In order to collect structural information, a deep analysis of the T-ROESY spectrum was performed. The classic ⁴C₁ chair conformation for all the monosaccharides was confirmed, on the basis of the intra-residue H1-H3 and H1-H5 NOE contacts, typical of β configured sugars. For the inner units, the NOE contacts typical for the β-(1→3) and β-(1→4) glycosidic linkages were found. For the GlcA unit at 4.62 ppm, a NOE contact between its anomeric proton and the one in the position 3 of the reducing α-GalNAc unit confirmed that it was the residue linked to the reducing unit (Figure 3.8).

¹⁹ Manuscript in preparation.

²⁰ Volpi, N. 2003 *Electrophoresis* 24:1063-1068.

²¹ Mucci, A., Schenetti, L., Volpi, N., 2000 *Carbohydr. Polym.* 41:37–45.

²² D'Arcy, S.M.T., Carney, S.L., Howe, T.J. 1994 *Carbohydr. Res.* 255:41-59.

The presence of overlapped signals in the anomeric region did not allow the correct integration of the NOESY densities needed to calculate inter protonic distances, so a qualitative classification of the NOE contacts was performed. Weak NOE (*w*) related to an inter-protonic distance ranging between 3.50 and 3.00 Å, medium densities (*m*) were referred to a distance between 2.5 and 3.0 Å and strong NOE (*s*) were related to a distance lesser than 2.5 Å. Table 3.8 shows the classification of all the NOE contacts of the anomeric region (Figure 3.8a).

Table 3.8: NOE contacts and corresponding visual classification of the their intensities in the T-ROESY spectrum. The letter *w* indicates a weak NOE contact corresponding to an inter-protonic distance between 3.50 and 3.00 Å, *m* a medium density referred to a distance between 2.5 and 3.0 Å and *s* a strong NOE contact related to a distance lesser than 2.5 Å. Only the X_nY_m NOE contacts are listed, the Y_mX_n ones are omitted.

Residue	NOE contacts
β-red-GalNAc F'	F _{1,2} ^s <i>w</i>
α-red-GalNAc F	F _{1,2} ^w
Internal GalNAc units B, D	F _{3,4} ^m , F ₃ E ₁ ^s
	F _{4,5} ^m , F _{4,6} ^w
	B _{1,3} ^s , B _{1,5} ^s , B ₁ C ₄ /A ₄ ^{s*}
	B _{3,4} ^m , B ₃ A ₁ /C ₁ ^{s*}
	D _{1,3} ^s , D _{1,5} ^s , D ₁ E ₄ /A ₄ /C ₄ ^{s*}
GlcA unit linked to reducing end E	D _{3,4} ^m , D ₃ C ₁ /A ₁ ^s
	D _{4,5} ^m , D _{4,6} ^w
	E _{1,5} ^m , E ₁ F ₃ ^s
Internal GlcA unit C	C _{1,3} ^s , C _{1,5} ^s , C ₁ D ₃ /B ₃ ^{s*}
	C _{4,2} ^m , C ₄ B ₁ /D ₁ ^{s*}
Terminal GlcA unit A	A _{1,3} ^s , A _{1,5} ^s , A ₁ B ₃ /D ₃ ^{s*}
	A _{2,3} ^m

* The overlapping of the corresponding signals did not allow a correct assignment.

These inter-protonic distances were then compared with those obtained through the molecular dynamic simulation, and a good agreement was estimated.

Analysis of the NMR spectra recorded in a mixture of H₂O:D₂O 9:1 allowed the characterization of the NH protons, namely the behaviour of the *N*-Acetyl groups.

For all the NH protons, two strong NOE contacts were detected (figure 3.3 b): one with the anomeric proton and another with the proton in position 3. For both the α and β reducing ends, a weak NOE contact with the proton in position 2 was identified too. These NOE contacts agreed with a trans disposition of the NH proton with respect to the carbinolic proton in position 2, they were also supported from the ³J_{NH,H} coupling constant value around 9.6 Hz.

In order to evaluate the involvement of NH protons in H-bond, their temperature coefficients were determined.

The evaluation of the change in the chemical shift was performed for the NH protons of the internal GalNAc units that merge in one signal and for the NH proton of the β-GalNAc at the

reducing end, the values of the α -red-GalNAc one was not calculated due to the overlap with the signals of the NH protons of the internal units.

All the data of the NH protons are reported in Table 3.9, and the temperature dependences are shown graphically in Figure 3.9.

Table 3.9: Chemical shift values of the NH protons of the internal and reducing GalNAc units at different temperatures

Temperature (K)	δ for internal GalNAc units (ppm)	δ for reducing β -GalNAc (ppm)
310	7.9812	8.1206
308	7.9956	8.1345
306	8.0104	8.1488
304	8.0258	8.1641
302	8.0409	8.1792
300	8.0562	8.1941
298	8.0731	8.2101
296	8.0904	8.2275
294	8.1058	8.2424
292	8.1214	8.2577
290	8.138	8.2738
288	8.1546	8.2899

The resulting temperature coefficient is -7.6 pbm/deg for the internal units and -7.4 pbm/deg for the reducing one. These values are larger with respect to those reported for NH protons in free exchange with water (-11 pbm/deg), but smaller with respect those for protons implicated in strong H-bond (between -4.6 and -1 pbm/deg)²³, therefore the involvement of the NH protons of the hexasaccharide in H-bond is around 40 %.

All these spectroscopic data collected for the *N*-Acetyl groups agreed with those reported for hyaluronic acid oligosaccharide, taken as reference^{24,25}.

²³ Cierpicki, T., Otlewski, J. 2001 *Journal of Biomolecular NMR* 21:249–261.

²⁴ Almond, A., DeAngelis, P.L., Blundell, C.D. 2006 *J. Mol. Biol.* 358:1256-1269.

²⁵ Blundell, C.D., Almond A. 2007 *Magn. Reson. Chem.* 45:430-433.

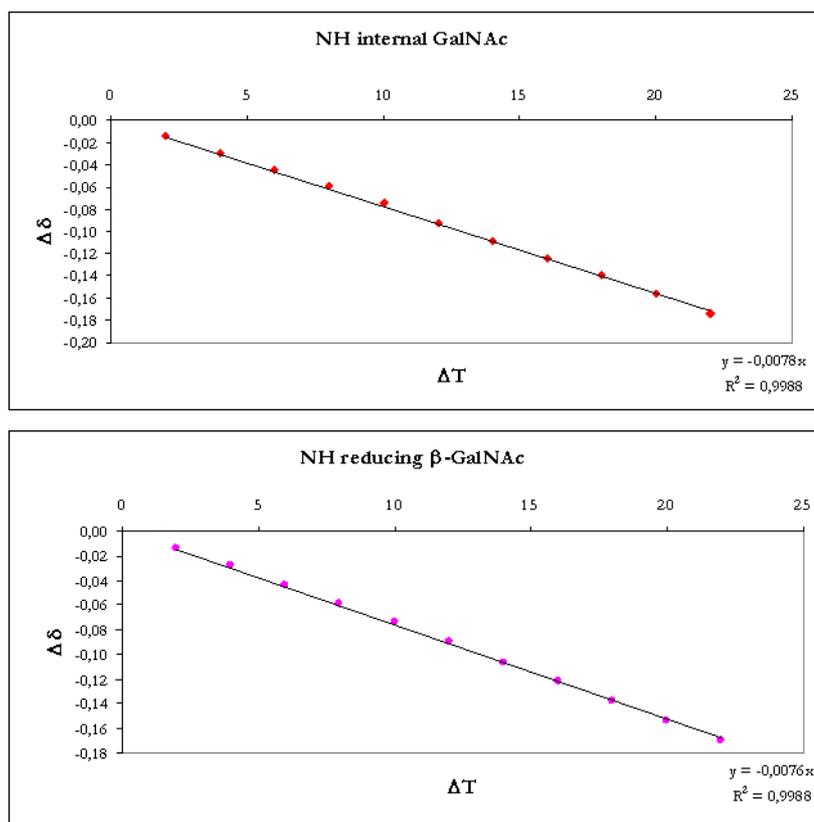


Figure 3.9: Graphical representation of the temperature dependence of the δ of NH protons

3.2.2 Molecular Mechanics calculations

The characterisation of the different glycosidic linkages through molecular mechanic calculation was performed to determinate the starting Φ and Ψ torsions. The two disaccharides β -GlcA-(1 \rightarrow 3)- β -GalNAc and β -GalNAc-(1 \rightarrow 4)- β -GlcA were built using Macromodel⁶ 8.0 and then minimized and analyzed using AMBER as force field. The *N*-Acetyl groups were set in the typical *trans* disposition, the carboxylic groups was considered in the anionic form and the CH₂OH group was set in the *gt* conformation, the stablest conformation for a *galacto* configuration¹².

Figure 3.10 shows the flexible maps of the two disaccharides. The conformation found for both the glycosidic linkages was in accordance with the *exo*-anomeric effect.

For the linkage β -(1 \rightarrow 4) one minimum was identified, the corresponding energy value was 111.64 KJ/mol and the Φ and Ψ values were respectively 46.0 and 2.5 degrees. For the disaccharide β -GlcA-(1 \rightarrow 3)- β -GalNAc more than one minimum was identified, the energy value of the lowest one was 127.74 KJ/mol and corresponding Φ and Ψ values at 48.0 and 7.8 degrees respectively. All these data agreed with those reported in the literature, for chondroitin sulfate disaccharide²⁶.

²⁶ Rodriguez-Carvajal, M.A., Imberty, A., Perez, S. 2003 *Biopolymers* 69:15-28.

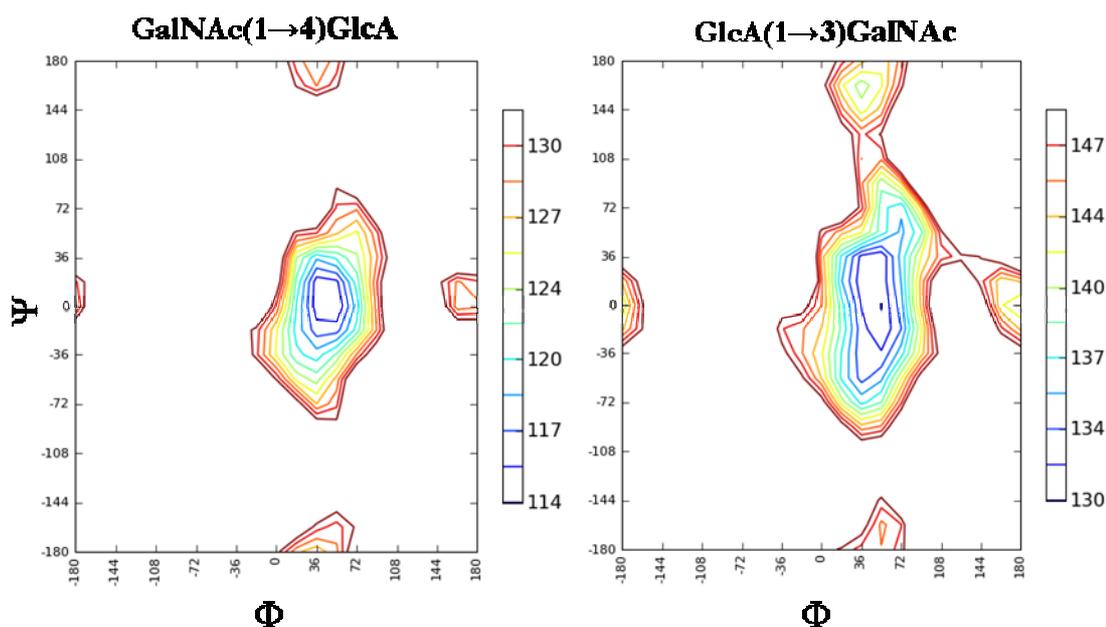


Figure 3.10: Flexible maps of the two glycosidic linkages of the unsulfated CS backbone

On the basis of these data, an evaluation of the possible NOE contacts between the protons across the glycosidic linkage was performed too. In Figure 3.11 the regions of conformational space consistent with different distances for both the glycosidic linkages are plotted as contours 2D-plots.

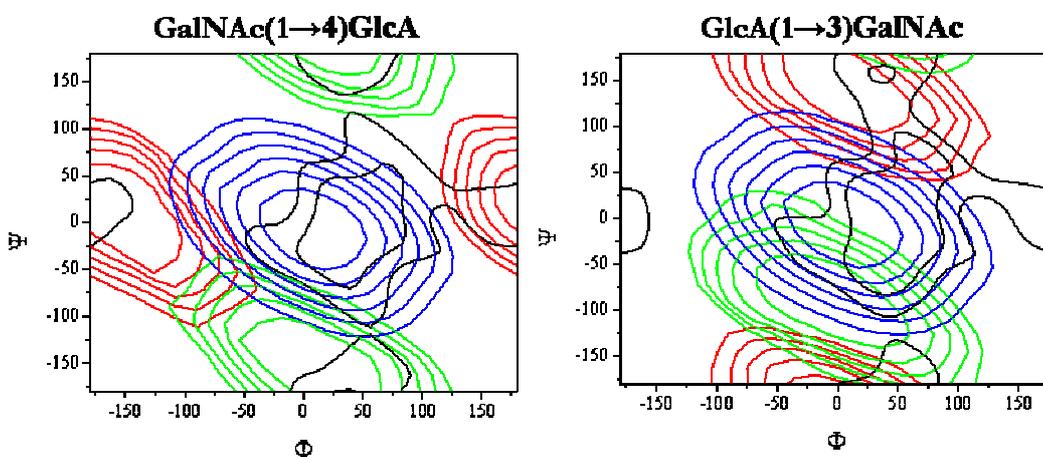


Figure 3.11: Contours plots representing the combination of the conformational space (black line) accessible by the linkages β -(1→4) and β -(1→3) and the one consistent with NOE contacts across the glycosidic linkages (colored lines). For the linkage β -(1→4) the blue line corresponds to H1-H4' distance, in green H1-H3' one and in red H2-H4' one. For the linkage β -(1→3) the blue corresponds to H1-H3' distance, in green the H1-H4' one and in red that H1-H2'. The distances are plotted from 2.1 Å (inner lever) in 0.3 Å intervals, up to the maximum that could be consistent with the experimental NOE measurements, 3.5 Å.

For the junction β -(1→4) the distance between the H1 of the GalNAc and the H4 of GlcA is lesser than 3.2 Å. The other distances (the one between H1 of GalNAc and H3 of GlcA or that between H2 of the GalNAc and H3 of GlcA), are outside the range observable via NOE measurement.

For the linkage β -(1 \rightarrow 3) the most probable NOE contact is the one between the two protons across the glycosidic linkage, the anomeric one and the one in position 3 of the GalNAc. In this case also the NOE contact between the anomeric proton and the one in position 4 must be considered, but just for Φ/Ψ combinations in accordance with an no *exo*-anomeric effect.

3.2.3 MD in explicit water

Applying the same approach used for the HA deca-saccharide (see 3.1.4), a molecular dynamics simulation in explicit water was performed on the unsulfated CS hexasaccharide too.

The hexasaccharide was built up starting from the optimal values of Φ and Ψ estimated for both the β -(1 \rightarrow 3) and β -(1 \rightarrow 4) glycosidic junctions. 4527 molecules of water TIP3P were used to construct an octahedral water box around the molecule.

The system, after the steps of equilibrations, was heated to 298 K and maintained to that temperature during the whole MD simulation. Periodic boundary conditions were applied to the system over 8 ns of simulation, the pressure was maintained constant at 1 atm and the box was free to relax around the molecule.

At the end of the 8 ns simulation, after the check of the consistency of the thermodynamic parameters (data not shown) the analysis of the calculation data was carried out.

3.2.4 Analysis of the MD trajectory and comparison with experimental data

During the MD simulation all the sugars of the hexasaccharide adopted a 4C_1 conformation, as indicated by the values around 2.7 and 2.5 Å calculated for the distances between the protons H1-H3 and H1-H5 of each ring, respectively. The *N*-Acetyl groups remained in the typical *anti* disposition and the preferred rotamer for the CH₂OH group of the three GalNAc units was mainly in the *gt*, as expected for a *galacto* configuration¹².

At the end of simulation an *exo*-anomeric orientation was detected for all the glycosidic linkages, as indicated by the scattered plots reported in Figure 3.12, demonstrating the goodness of the employed force field. Just a little amount of no-*exo*-anomeric effect (< 4%) was estimated for the linkage β -(1 \rightarrow 3). For both linkages, however, the conformational space covered during the MD simulation was restricted. In particular, the linkage β -(1 \rightarrow 4) populated a conformation clustered around 40° for the dihedral angle Φ and around -30° for the Ψ angle. For the junction β -(1 \rightarrow 3) a preferred value around 45° was estimated for Φ , whereas Ψ assumed a large array of values (Fig. 3.13).

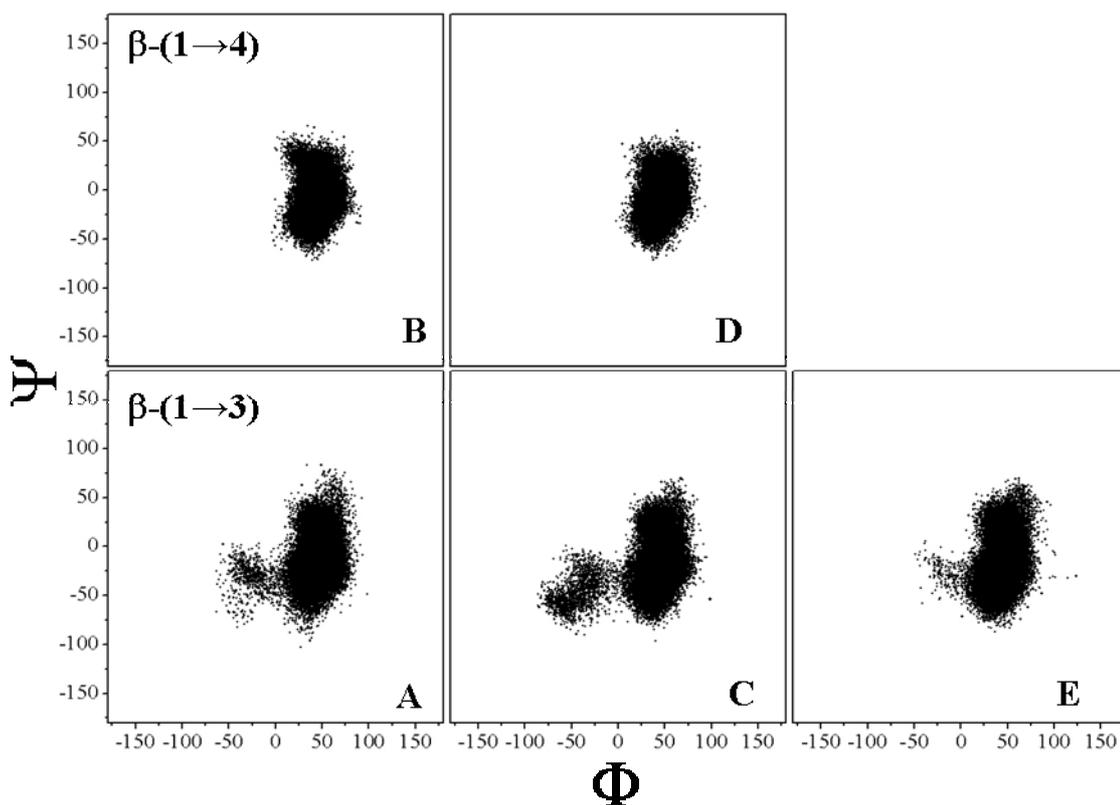


Figure 3.12: Scattered plots of the five glycosidic linkages of the unsulfated CS hexasaccharide

Counting the frames with a $\Psi > 0$, an estimation of the percentage of structures characterized by a positive Ψ value was performed. A number ranging between 24 and 30%, with respect the total number of structures, was estimated for the three different β -(1 \rightarrow 3) linkages.

The interconversion between the conformers corresponding to the two possible Ψ values is continuous over the whole simulation as visible from the large fluctuation between positive and negative values, as demonstrated by the graphics of Ψ *vs* time of the three β -(1 \rightarrow 3) linkages, reported in Figure 3.13.

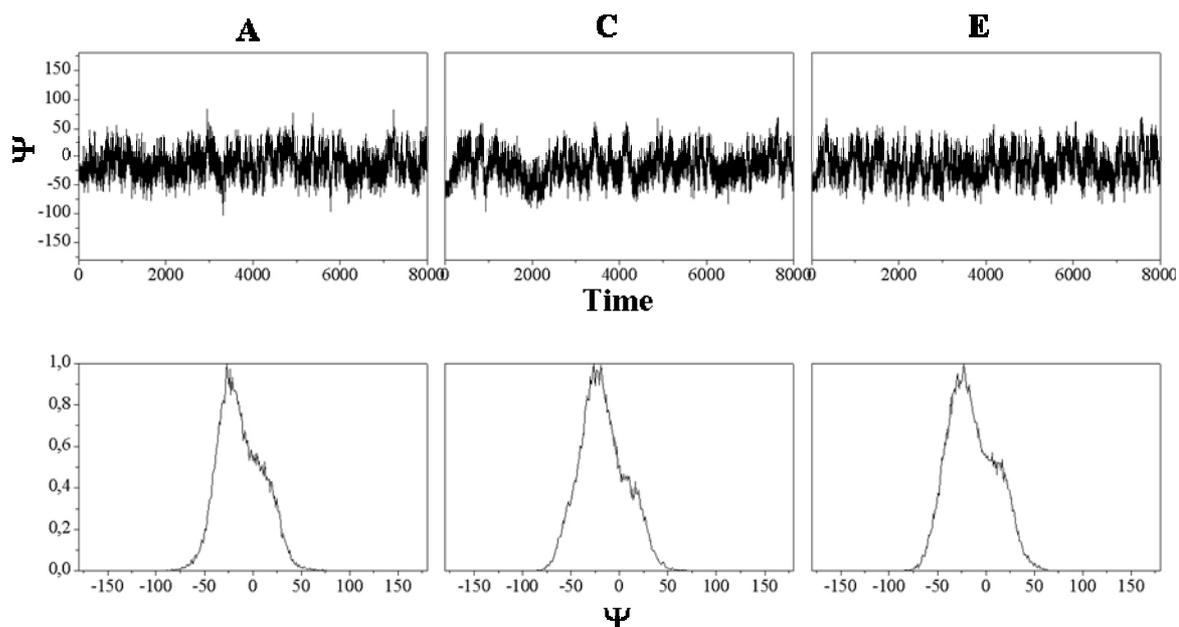


Figure 3.13: Graphics of Ψ vs time (ps) of the three β -(1 \rightarrow 3) linkages

The flexibility of this dihedral angle, is demonstrated by its higher RMSD as well with respect to the those calculated for the other dihedral angles.

The averaged values, the corresponding RMSD and the more frequent angular values of all the glycosidic junctions of the molecule are listed in Table 3.10.

Table 3.10: Averaged dihedral angular values (degrees) and corresponding RMSD. The more frequent angular value (degree) assumed during the MD simulation.

Angle	Averaged value	Frequent value
Φ_A	43.38 ± 15	43
Ψ_A	-12.49 ± 22.97	-27
Φ_B	44.72 ± 11.23	45
Ψ_B	-7.18 ± 18.84	-9
Φ_C	41.67 ± 21.01	45
Ψ_C	-17.00 ± 23.92	-23
Φ_D	46.01 ± 11.05	47
Ψ_D	-8.49 ± 18.25	-7
Φ_E	45.16 ± 12.31	47
Ψ_E	-14.30 ± 24.70	-23

In order to compare the experimental data to the simulated ones, the averaged inter-protonic distances over the whole MD trajectory were calculated. The averaged and the corresponding RMSD are listed in Table 3.11.

Table 3.11: Inter-protonic averaged distances and corresponding RMDS calculated over the whole MD simulation

NOE contact	Averaged distance (Å)
A ₁ →B ₂	4,17 ± 0.32
A ₁ →B ₃	2,33 ± 0.21
A ₁ →B ₄	3.68 ± 0.49
B ₁ →C ₃	4,34 ± 0.21
B ₁ →C ₄	2,34 ± 0.19
C ₁ →D ₂	4,20 ± 0.31
C ₁ →D ₃	2,34 ± 0.23
C ₁ →D ₄	3.56 ± 0.55
D ₁ →E ₃	4,34 ± 0.21
D ₁ →E ₄	2,35 ± 0.2
E ₁ →F ₂	4,17 ± 0.33
E ₁ →F ₃	2.33 ± 0.21
E ₁ →F ₄	3.63 ± 0.52

On the basis of the NOE classification (Table 3.8), reported above, a good correspondence between the experimental data and the theoretical one was estimated. All the calculated distances between the protons involved in strong NOE contacts were around or lesser than 2.6 Å. The distances between protons involved in medium or weak NOE contacts were consistent with the ranges established at the beginning in the NMR section. The inter-protonic distances diagnostic of *no-exo*-anomeric disposition were around 4-4.5 Å, and the corresponding dipolar couplings were not observed in the NOESY spectrum, as expected.

The MD data were then considered consistent with experimental data and further used to analyze the conformational behaviour of the oligosaccharide. For this purpose, the length of the oligomer was inspected. For each conformer, this parameter resulted from the combination of all the Φ , Ψ values, therefore, it was a simple way to observe their time-course behaviour.

During the 8 ns simulation, the hexasaccharide tended to assume a quite elongate disposition as indicated by the length around 27 Å; however, inspection of trajectory in Figure 3.14_a showed that the oligosaccharide could also assume rather compressed conformations with the a length in the range between 22 and 26 Å.

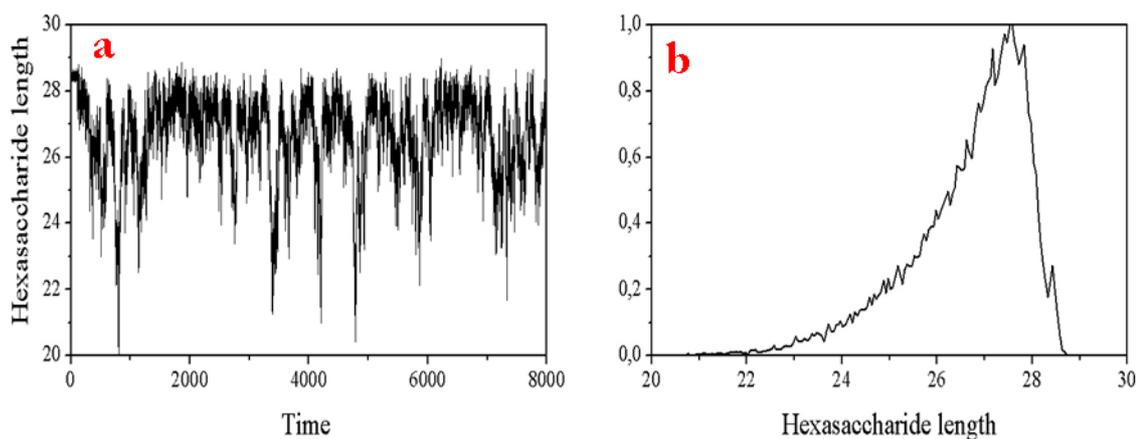


Figure 3.14: Graphic (a) and corresponding projection (b) of the hexasaccharide length during the 8 ns.

Studies performed on hyaluronan established that the interactions between adjacent monosaccharides are responsible of its secondary structure¹³, so the estimation of the presence of H-bonds in the hexasaccharide was carried out, too. The identification of H-bonds was performed setting the distance between the hydrogen donor (D) and the hydrogen acceptor (A) lesser than 3.5 Å and the angle D-H-A larger than 120 degrees. Intra e inter-molecular H-bonds were detected and in particular the more persistent interactions (100%) were those intra-residual, involving the *exo*-cyclic oxygens as acceptor. For the conformational characterization of the molecule, however, just the inter-residual ones were taken in account, in particular those across the glycosidic linkages involving the *endo*-cyclic oxygens or the amidic groups or the hydroxyl groups in position 2 of the GlcA and in position 4 of the GalNAc. The figure 3.15 reports the H-bonds considered during the investigation. The graphics in Figure 3.16 show the combination between the distance D-A and the angle D-H-A for some of the H-bonds considered during the analysis. For the two glycosidic linkages a different amount of structures agreed with the H-bond definition, indicating that not all these interactions were stable, consequently the possible presence of water bridges was also taken in account, as reported in the literature for glycosaminoglycans chains²⁷.

As a result, for the junction β -(1 \rightarrow 4) the most stable and persistent interaction was the H-bond between the *endo*-cyclic oxygen of the inner GalNAc units (B and D) and the OH in position 3 of the GlcA (Figure 3.15 type c, Figure 3.16 plots VII and VIII), it was identified in 70% of the conformers.

The other H-bond reported in literature, namely the one between the NH proton and the carboxylic group (type d Figure 3.15), occurred with a frequency around 20% (Figure 3.16), probably because it was destroyed by the interaction with water molecules. The investigation on

²⁷ Almond, A. 2005 *Carbohydr. Res.* 340:907-920.

this H-bond, was complicated by the presence on the carboxylate group of two equivalent oxygen atoms. For this reason, the estimation was performed for each carboxylic oxygen and the total persistence was calculated as sum of the two values, due the equivalence of the oxygens on a charged carboxylic group. The plots (Fig. 3.17), as well, were the sum of the data belonging from both the oxygens, two different colours were used to represent the points generated by an oxygen or by the other. The presence of water-bridges was not investigated. The persistence estimated for this H-bond agreed with the experimental datum belonging to the estimation of the temperature coefficient of the NH protons of the internal GalNAc units (see 3.2.1).

The lower presence of the H-bond between the NH proton and the carboxylic group was accompanied by the absence of the H-bond, across the β -(1 \rightarrow 3) junction, involving the axial OH of the GalNAc, namely the one between the OH in position 4 of the GalNAc unit and the *endo*-cyclic oxygen of the previous GlcA.

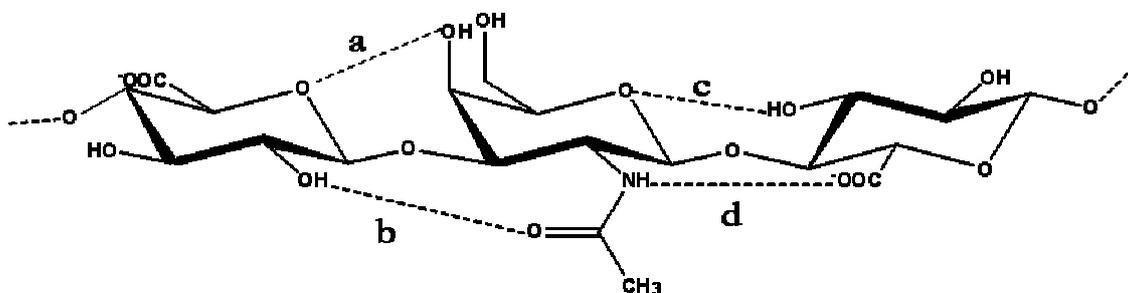
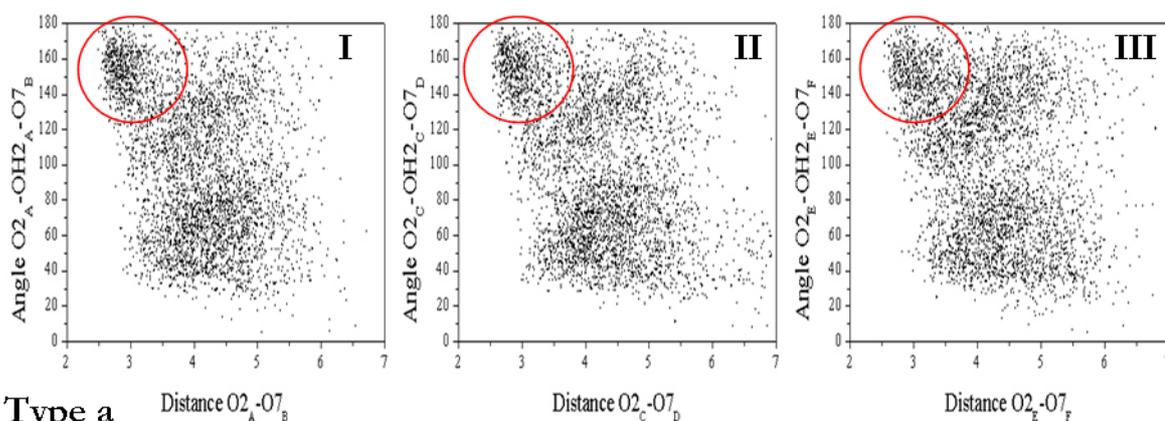
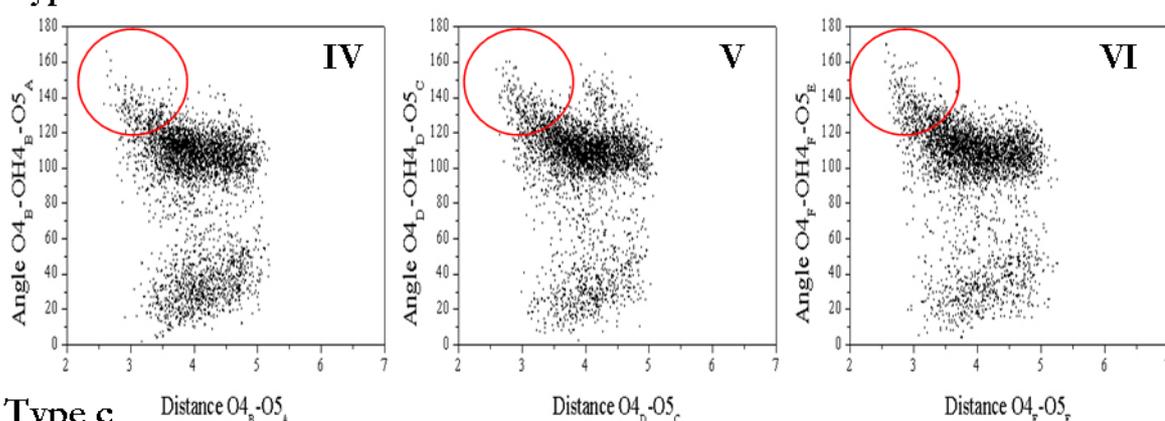


Figure 3.15: Principle H-bonds in the Chondroitin sulfate backbone

Type b



Type a



Type c

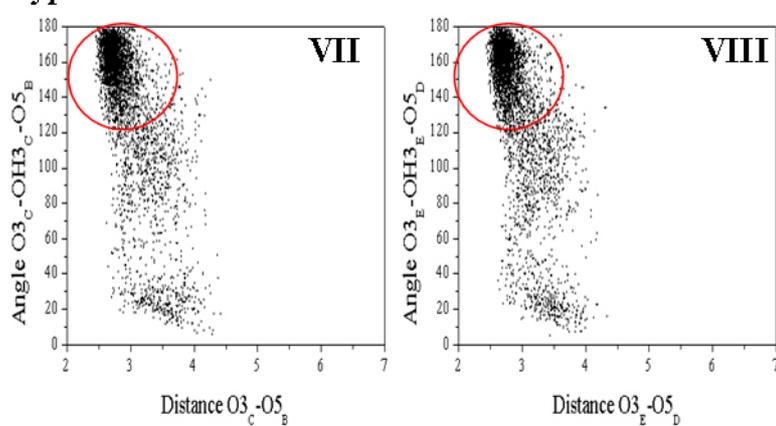


Figure 3.16: Scattered plots represented the distances and the angles of the H-bonds across the linkage β - $(1 \rightarrow 4)$ (c), and across the linkage β - $(1 \rightarrow 3)$ (a and b).

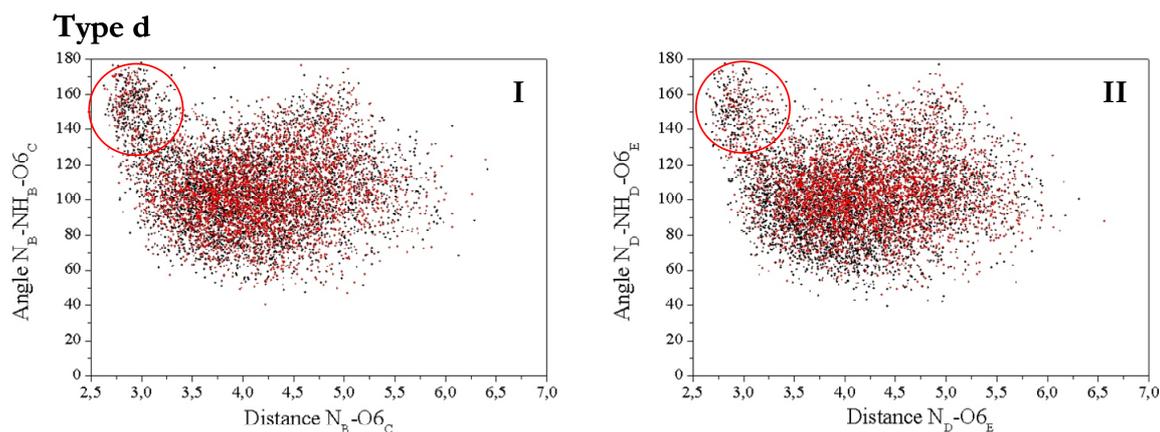


Figure 3.17: Scattered plots represented the distances and the angles of the H-bond type d

The axial orientation of the OH group in position 4 of the GalNAc might explain the poorness of H-bonds across the β -(1 \rightarrow 3) linkage.

In particular, the direct interaction one between the *exo*-cyclic oxygen of the GlcA unit and the OH in position 4 of the GalNAc (figure 3.11 type a, figure 3.12 plots IV, V and VI) was characterized by a persistence lesser than 10 % and that between the O7 of the GalNAc and the OH in position 2 of the GlcA (figure 3.11 type b, figure 3.12 plots I, II and III), showed a presence around 16 %.

However, a closer look at plots IV, V and VI in Figure 3.16, indicated that the *exo*-cyclic oxygen of the GlcA unit and the OH in position 4 of the GalNAc gave rise to dispositions characterized first of all by an D-H-A angle around 110° and a distance between 3.5 and 4.5 Å. In order to investigate, if those dispositions were caused by the presence of water-bridges, different frames of the MD simulation were isolated and studied. In particular, in this way the collocation of water molecules with the respect to H-bond donor and acceptor atoms was investigated. The frames isolated were characterized by an angle around 110° and different D-A distance values, trying to explore more possibilities: two frames showed a distance around 3.5 Å, two a distance around 4 Å and two a distance around 4.5 Å.

The analysis of these frames indicated that in no cases there was a water-bridge, suggesting that those dispositions corresponded to decoupled situations. This result agreed with literature data²⁷.

This analysis confirmed that the β -(1 \rightarrow 3) linkage in chondroitin sulfate is not stabilized by H-bond, differently from hyaluronan for which direct, strong and persistent H-bonds were found for both β linkages²⁷. Another aspect is that all the OH functions modifiable by sulfation are not involved in H-bond, so that all the H-bond connections stabilized the secondary conformation of the polymer should still occur in the sulfate analogue.

Helix parameters estimation

Using the software POLYS¹⁸, an evaluation of the secondary structure of the unsulfated chondroitin backbone was performed as well.

Applying the same approach used for the oligosaccharide of HA (see 3.1.6), the regular helix structures generable on the basis of the dihedral angles of the inner junctions were calculated. The number of repeating units per helical turn, n , was 2.92 and the axial rise was a negative value, so, also in this case, there was a left-handed coiling and the conformation was a hybrid between a two ($n=2$) and a three ($n=3$) helix.

Considering the structural similarity with the hyaluronic acid, however, a four-fold geometry was taken in account as well.

Not all the optimized Φ and Ψ values of the three regular helix structures (Tab 3.12) fell in the dihedral range (averaged value and associated error, RMSD) defined during the analysis, revealing that just the 3_2 geometry was possible (Fig. 3.18).

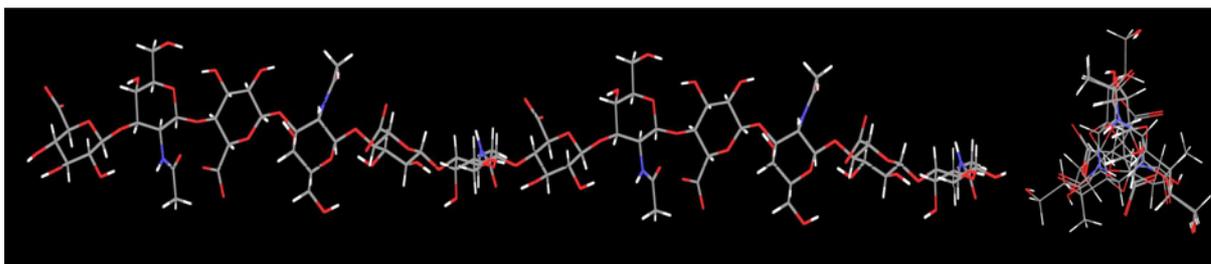


Figure 3.18: 3_2 left-handed helix structure generated with POLYS

Table 3.4 Averaged Φ and Ψ values of the inner junctions and optimized ones for the 3-fold and 4-fold geometries

Angle*	Average value (degree)	Optimized value (degree) 3_2	Optimized value (degree) 4_2	Optimized value (degree) 2_2
$\Phi_{1\rightarrow3}$	-76.38 ± 10.48	-75.312	-65.112	-95.093
$\Psi_{1\rightarrow3}$	-126.23 ± 17.15	-125.248	-115.880	-143.416
$\Phi_{1\rightarrow4}$	-77.77 ± 25.51	-76.804	-67.590	-94.672
$\Psi_{1\rightarrow4}$	-102.70 ± 22.59	103.615	112.343	83.690

* The dihedral angles Φ and Ψ were defined respectively as $O_5-C_1-O-C_n$ and $C_1-O-C_n-H_{n+1}$.

In order to quantify the number of conformers showing a good homology with the 3_2 helix geometry, a comparison between the frames of the entire MD simulation and an optimized 3_2 helix structure, was performed. The comparison was carried out in the same way described for the HA decasaccharide (see 3.1.6). The divergence between the structures was defined through the RMSD value (Fig. 3.19)

Selecting 0.5 as allowed RMSD values, it was estimated that the 10 % of structures had a conformation with a good structural homology with the 3_2 left-handed model.

This datum, indicated that, even in the case of the unsulfated chondroitin molecule, in solution, a conformation in accordance with a 3-folded left-handed helix geometry was preferred, as estimated for the hyaluronic acid.

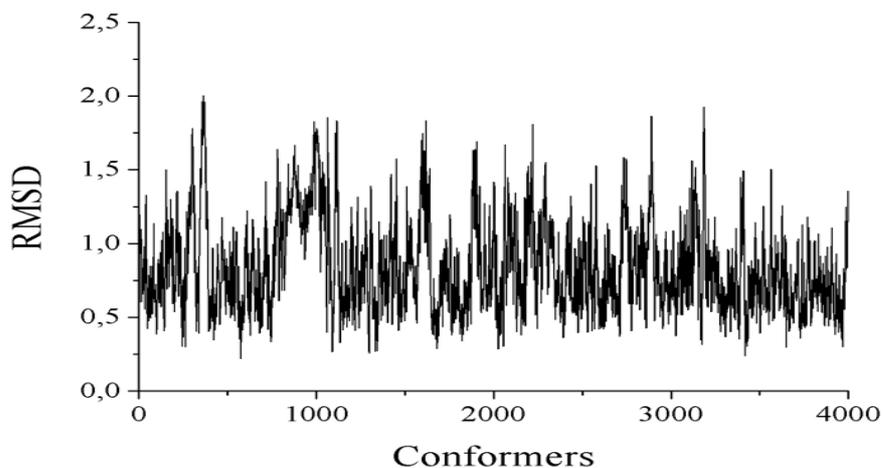


Figure 3.19: RMSD of the difference between the coordinates of each frame of the subgroup and of the two models.

3.4 MD simulation on a representative dodecasaccharide of shark chondroitin sulfate

Molecular modelling techniques were used also to establish the influence of sulfate groups on the behaviour of chondroitin backbone.

In particular, object of these studies was a dodecasaccharide (Fig. 3.20) representative of the CS isolated from the cartilage of lesser spotted dogfish (see 2.1.2); it contained the three repeating units CS-A, CS-C and CS-D.

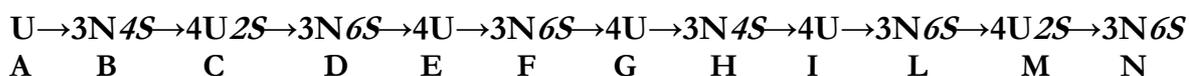


Figure 3.20: Dodecasaccharide sequence.

The repeating units were linked creating a model sequence, trying to generate different sulfation pattern across both the linkages $\beta\text{-(1}\rightarrow\text{3)}$ and $\beta\text{-(1}\rightarrow\text{4)}$.

The experimental data used to validate the theoretical ones, were those collected through the analysis of the NOESY spectrum of the shark CS (see. 2.1.2).

Molecular modelling studies were carried out as described previously: first, molecular mechanics analysis was performed on the monosulfated and disulfated disaccharide repeating units to estimate the preferred Φ and Ψ values for each linkage, which were successively used to build the molecule subjected to a molecular dynamics simulation in explicit water.

Finally, data collected were compared with those of the unsulfate structure.

3.4.1 Molecular mechanics studies on sulfated CS disaccharides

The presence of sulfate groups, obviously, influences the behaviour of the glycosidic linkages of the CS backbone. A first determination of how the sulfate groups, namely bulked and negative charged groups, modify the torsional angles, was achieved through molecular mechanics studies on disaccharides with different sulfation patterns. The data, collected in this way, were then used as starting Φ and Ψ values of the different junctions in the sulfated dodecasaccharide.

For both the linkages β -(1 \rightarrow 3) and β -(1 \rightarrow 4) different disaccharides were built up and analyzed. For all the structures, molecular mechanics calculations were performed applying the same conditions and using the same parameters described for the unsulfated disaccharides (see 3.2.2).

The disaccharide sequences, the energy values of the minima and the corresponding Φ and Ψ values, are listed in Table 3.13, together with those corresponding to the unsulfated structures. The flexible maps, plotted as 2D contours plots are shown the figure 3.21.

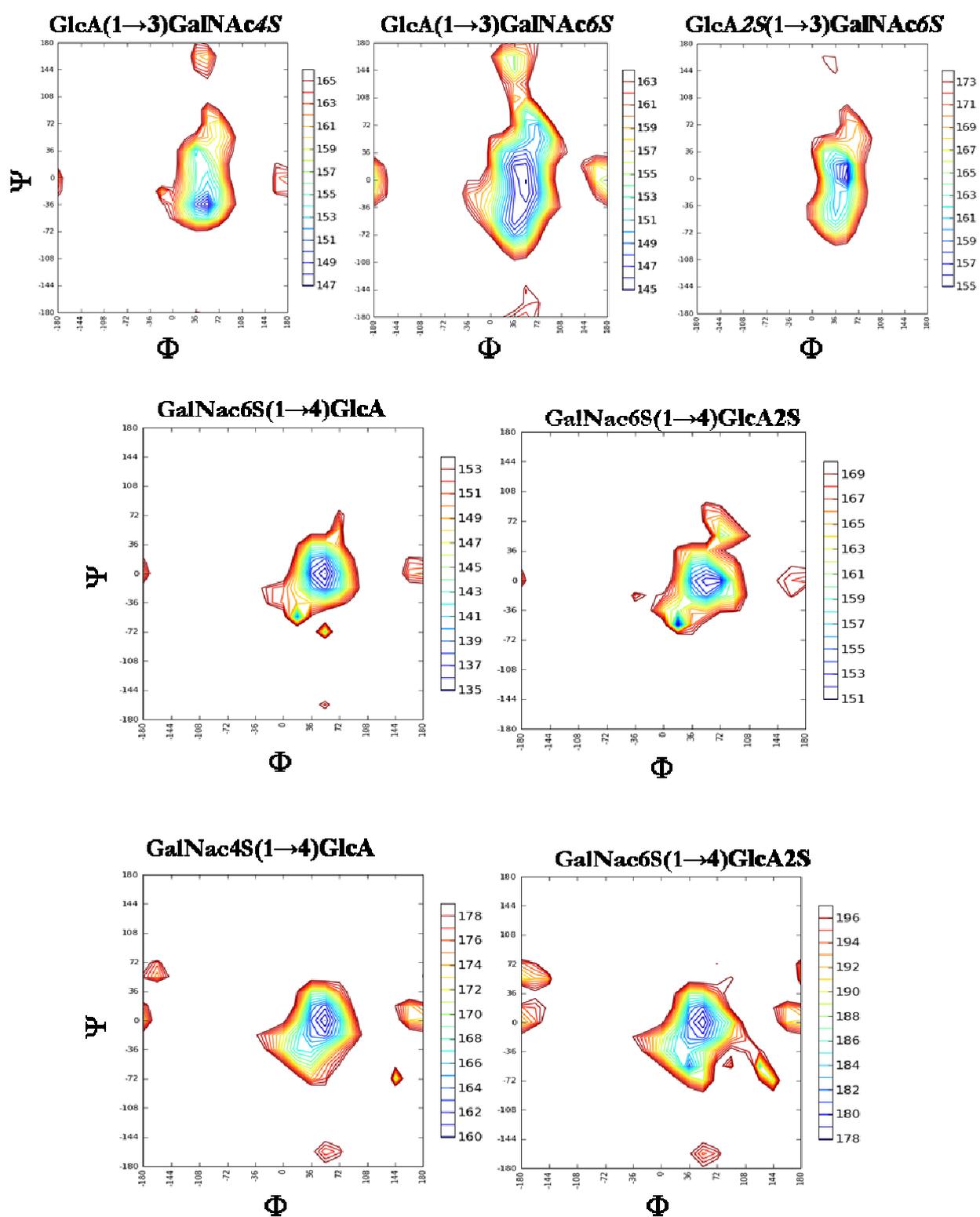


Figure 3.21: Flexible maps of the different CS monosulfated and disulfated disaccharides.

Table 3.5: Disaccharide sequences, dihedral angles and Energy values of different sulfated motifs. The values of the unsulfated disaccharides are reported too.

Linkage	Disaccharide sequence	Angle	Values (degree)	Energy (kJ/mol)
β -(1 \rightarrow 3)	GlcA-(1 \rightarrow 3)-GalNAc	Φ	48.0	127.74
		Ψ	7.8	
	GlcA-(1 \rightarrow 3)-GalNAc4S	Φ	54.1	144.58
		Ψ	-17.9	
	GlcA-(1 \rightarrow 3)-GalNAc6S	Φ	48.3	144.781
		Ψ	8.4	
GlcA2S-(1 \rightarrow 3)-GalNAc6S	Φ	56.3	154.943	
	Ψ	-2.0		
β -(1 \rightarrow 4)	GalNAc-(1 \rightarrow 4)-GlcA	Φ	46.0	111.64
		Ψ	2.5	
	GalNAc4S-(1 \rightarrow 4)-GlcA	Φ	51.1	159.448
		Ψ	-0.9	
	GalNAc6S-(1 \rightarrow 4)-GlcA	Φ	49.4	134.613
		Ψ	-2.4	
	GalNAc4S-(1 \rightarrow 4)-GlcA2S	Φ	50.8	177.401
		Ψ	-0.8	
	GalNAc6S-(1 \rightarrow 4)-GlcA2S	Φ	49.8	140.613
		Ψ	-2.8	

In all the cases, a principal minimum could be individuated and the conformations were in accordance with the *exo*-anomeric effect.

Making a comparison among the Φ and Ψ values and the maps of the three different structures containing β -(1 \rightarrow 3) linkages and also with respect to those of the unsulfated molecule, obviously, a real difference between the different sulfated structures was detected, due to the different distribution of sulfate groups across the linkages.

In particular, the presence of a sulfate group in position 4 of the GalNAc shifted the minimum position within the region covered from the map.

The presence of a sulfate group in position 6 of the GalNAc, instead, favoured a conformation with Φ and Ψ values, similar to those reported for the unsulfated motif. More over, the presence of a second sulfate group, on the acid residue in the disaccharide GlcA2S(1 \rightarrow 3)GlcNAc6S, did not produce a behaviour more different than the one estimated for the monosulfated structure.

Among the four different disaccharides containing a β -(1 \rightarrow 4) linkage, a similar behaviour was detected, with a conformation characterized by similar Φ and Ψ values.

3.4.2 Molecular Dynamics in explicit water

The dodecasaccharide having the sequence reported in Figure 3.20 was built up, the glycosidic torsions were set to the values estimated through molecular mechanics studies and the atomic partial charges were calculated through the software Gaussian03²⁸.

The molecule then was put in a truncated octahedral box, made up by 12429 TIP3P water molecules and the corresponding system was minimized, equilibrated and then heated to 288 K. The molecular dynamic simulation of 2,5 ns was performed applying periodic boundary conditions and a cut off of 14 Å.

The thermodynamic parameters (volume, pressure, water density and temperature) showed a regular behaviour over the entire MD so the analysis of the simulation data was carried out.

3.4.3 Analysis of the simulation data

The first parameters checked were the torsional angles, the estimation of the inter-protonic distances (intra and inter-residues) and the presence of H-bonds.

All the residues, during 2,5 ns maintained the classical ⁴C₁ conformation, as demonstrated by the inter-protonic distances between the protons H1-H3 and H1-H5 around 2.65 Å and 2.55 Å, respectively.

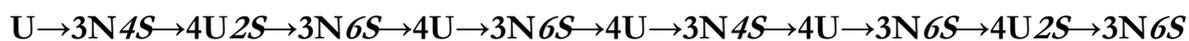
The averaged dihedral angles estimated for each glycosidic junction (Tab. 3.14) showed a good accordance with those estimated by the molecular mechanics calculations.

Table 3.14: Averaged values (degrees) of the dihedral angles Φ and Ψ of all the glycosidic linkages.

Linkage β -(1 \rightarrow 3)		Linkage β -(1 \rightarrow 4)	
Angle	Averaged values	Angle	Averaged values
Φ_A	48.6 \pm 12.85	Φ_B	46.2 \pm 11.2
Ψ_A	-16.7 \pm 15.5	Ψ_B	10.8 \pm 16.6
Φ_C	51.5 \pm 9.6	Φ_D	48.4 \pm 10.5
Ψ_C	-2.9 \pm 22.4	Ψ_D	7.5 \pm 11.4
Φ_E	47.8 \pm 12.1	Φ_F	46.9 \pm 12.24
Ψ_E	-17.4 \pm 22.1	Ψ_F	-1.5 \pm 17.8
Φ_G	50.2 \pm 11.3	Φ_H	48.4 \pm 10.5
Ψ_G	-20.1 \pm 13.5	Ψ_H	5.7 \pm 13.8
Φ_I	44.3 \pm 11.3	Φ_L	48.3 \pm 12.1
Ψ_I	-20.9 \pm 19.6	Ψ_L	5.6 \pm 14.6
Φ_M	50.4 \pm 10.9		
Ψ_M	1.79 \pm 27.9		

²⁸ Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. et al Gaussian, Inc., Pittsburgh PA, 2003.

As showed by the scattered plots (Fig. 3.22 and 3.23), for each glycosidic linkages, a disposition in accordance with the *exo*-anomeric effect was detected, confirming the goodness of the employed force field (Glycam06) and the behaviour predicted by the flexible maps in Figure 3.21. The *N*-Acetyl group of all GalNAc units was found in the *anti* disposition and the preferred rotamer for the CH₂OH group was partially affected by the sulfation pattern. For all the GalNAc units the *gt* rotamer was the one preferred, but for the GalNAc6S units also a little amount of *tg* one has to be taken in account, as indicated by the plots in Figure 3.24.



A
B C D E F G H I L M N

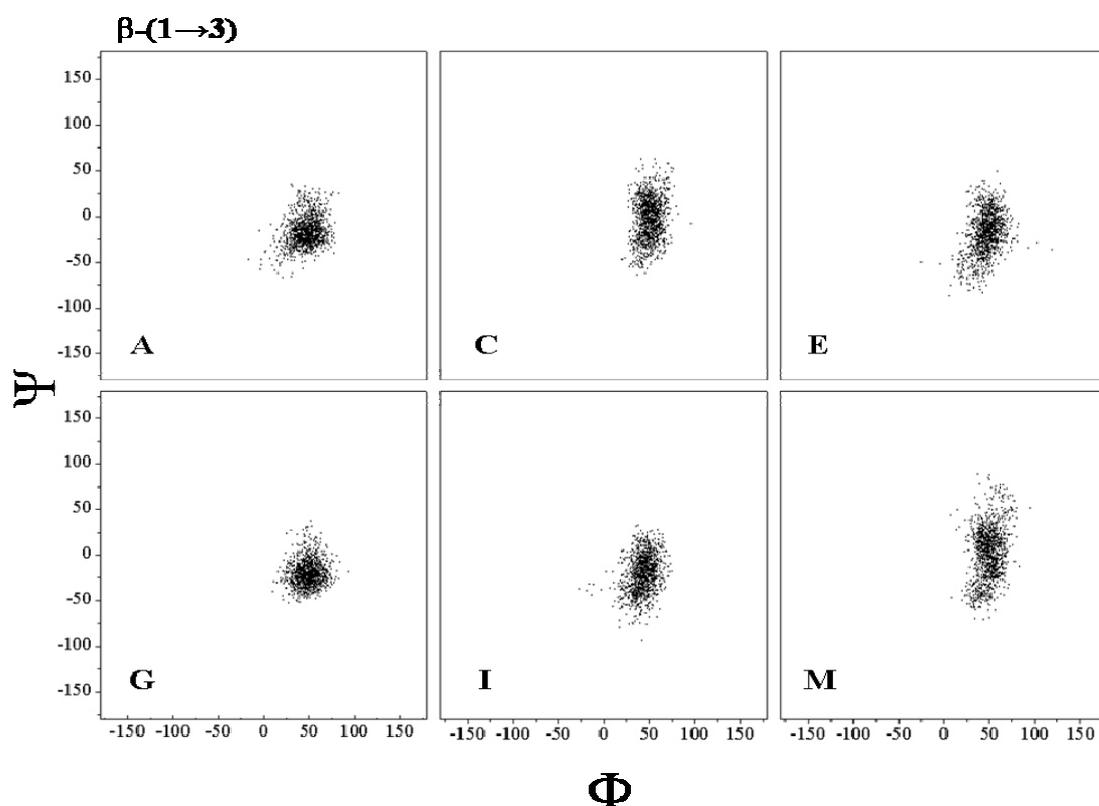


Figure 3.22: Scattered plots of all the β -(1→3) linkages of the dodecasaccharide.

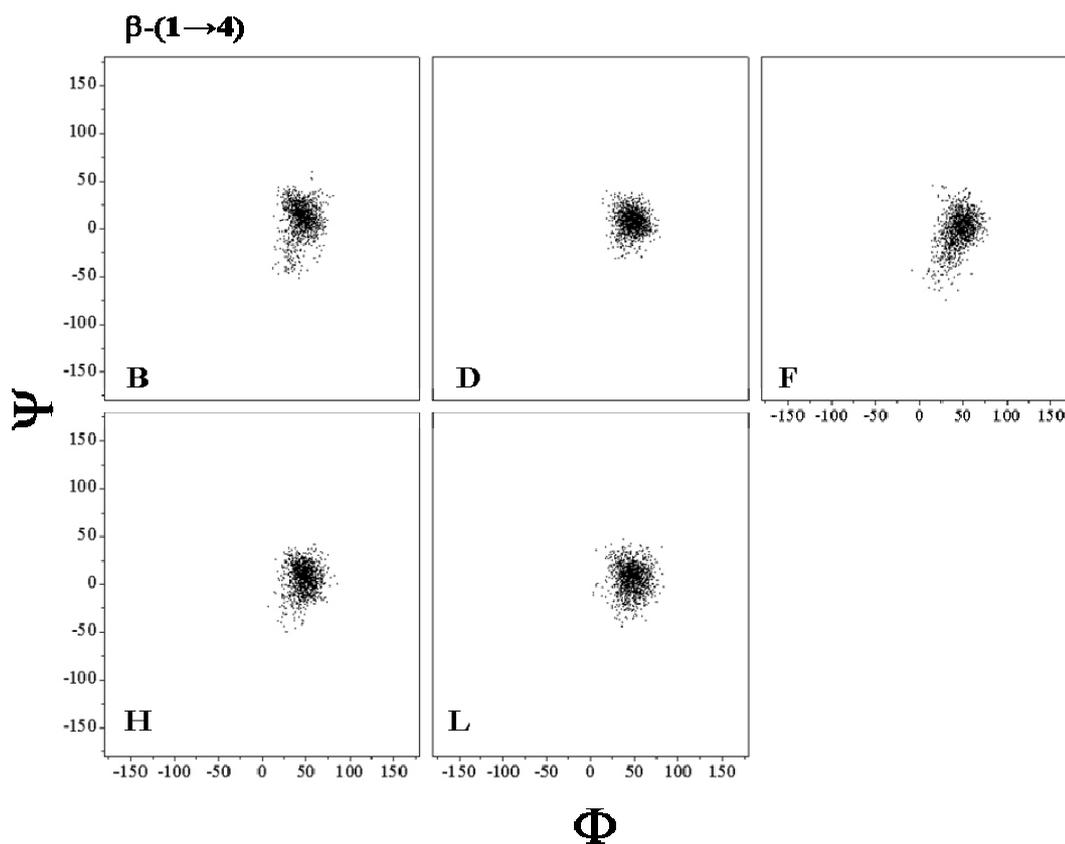


Figure 3.23: Scattered plots of all the β -(1 \rightarrow 4) linkages of the dodecasaccharide.

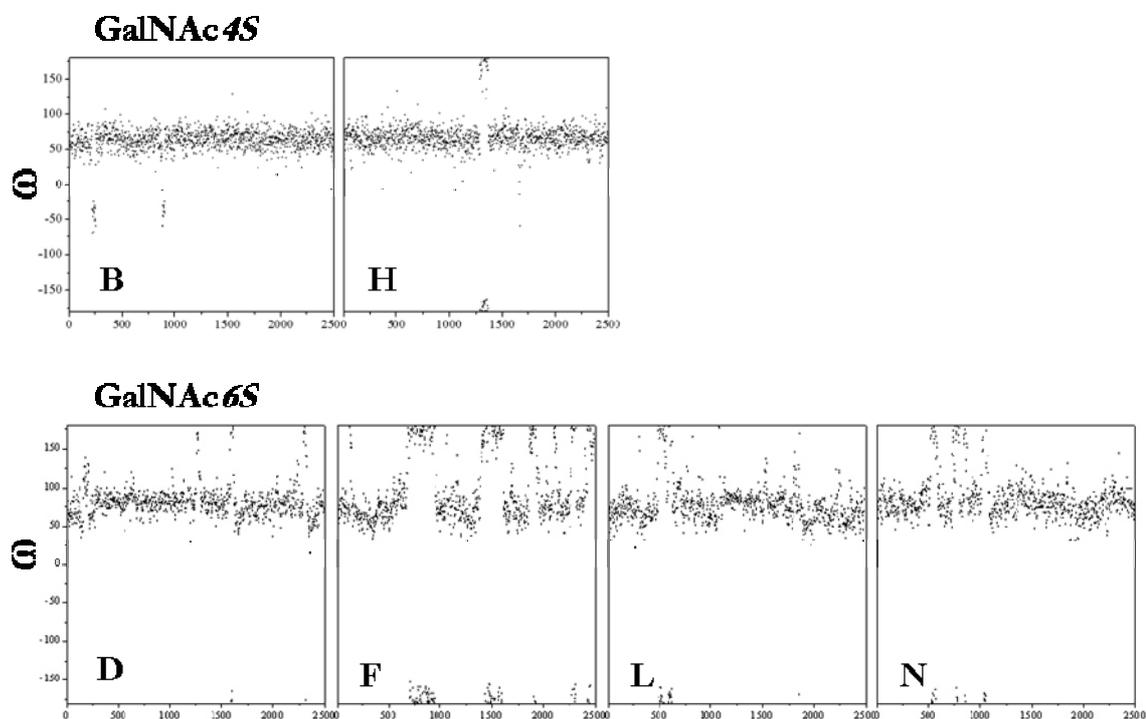


Figure 3.24: Plots of the dihedral angle ω (degrees) *vs* time (ps) of the CH₂OH group of all the GalNAc units

The disposition of the sulfate group was established through the analysis of the behaviour of the dihedral angle defined as $C_{x-1}-C_x-O-S$. The three dihedral angles analyzed were graphically represented by the Newman projection in Figure 3.25.

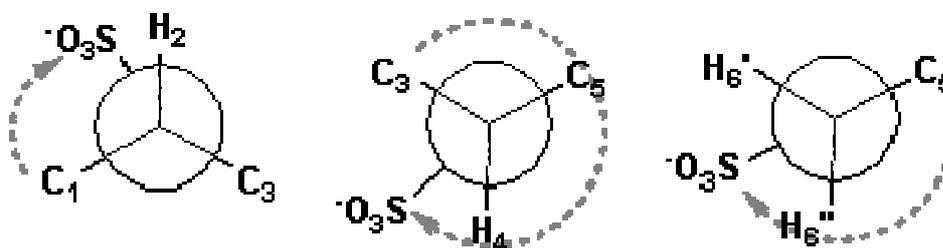


Figure 3.25: Newman projections of the 3 sulfate group dihedral angles

For the sulfate group in position 4 of GalNAc unit and for that in position 2 of the GlcA, a disposition in agreement with literature data was detected²⁹. In particular, for the group on the acidic residues an averaged angular value around 125° was estimated, indicating an eclipsed orientation with respect to H-2 (Fig. 3.26, a). For the group in position 4 of the GalNAc units a value around -115° was found, so an eclipsed orientation with respect to H-4 was deduced (Fig 3.26, b)

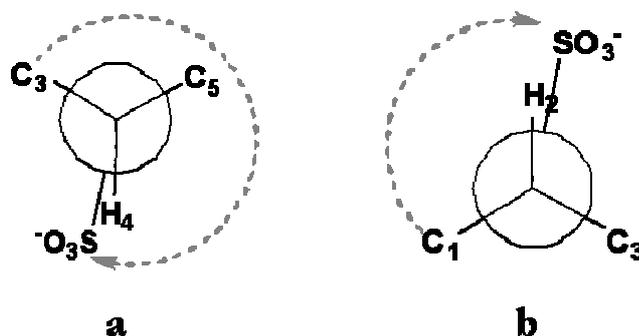


Figure 3.26: a) Newman projection of the sulfate group on the GlcA. b) Newman projection of the sulfate group in position 4 of the GalNAc

In both cases the preference for eclipsed orientations was explained by the presence of low steric interactions between atoms.

For the sulfate group in position 6 on the GalNAc a fluctuation between different angular values was observed, as highlighted by the scattered plots in Figure 3.27.

²⁹ Blanchard, V., Chevalier, F., Imbert, A., Leeftang, B.R., Basappa, Sugahara, K., Kamerling J.P., 2007 *Biochemistry* 46(5):1167-1175.

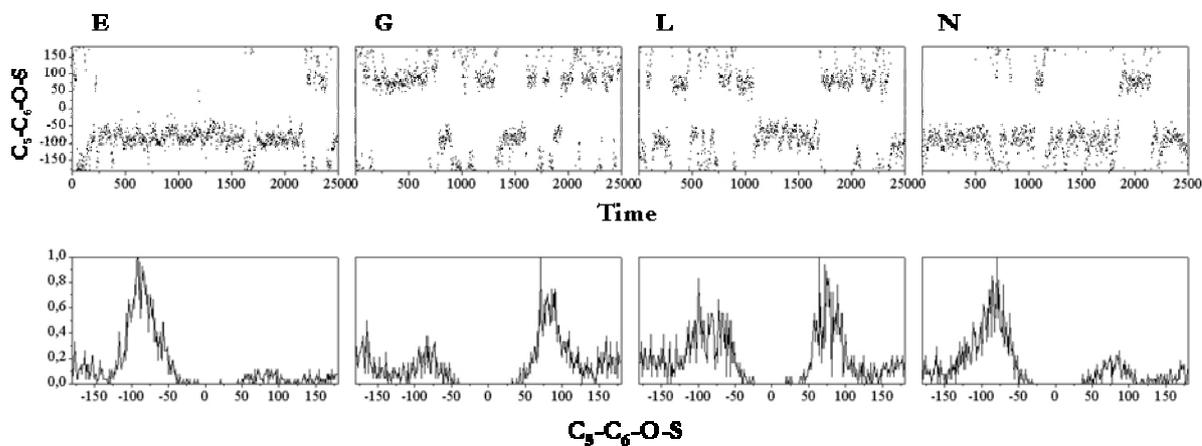
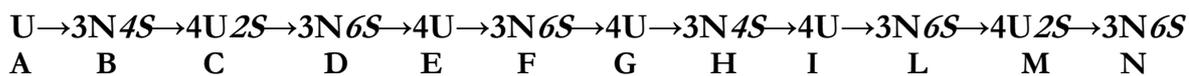


Figure 3.27: Scattered plot of the sulfate group dihedral angle in position 6 of the GalNAc vs time.

The behaviour of the torsional angles of the sulfate groups in position 6 of the GalNAc was affected by the sulfation pattern of the GlcA unit linked in position 3. When the acidic unit was unsulfated a value around -80° was calculated, indicating a quite eclipsed disposition with respect to the H-6' (E and N, Figure 3.28). When the GlcA unit brought a sulfate group in position two, the dihedral angle value estimated was around 80° indicating a quite eclipsed disposition with respect the H6'' (G and L, Figure 3.28).

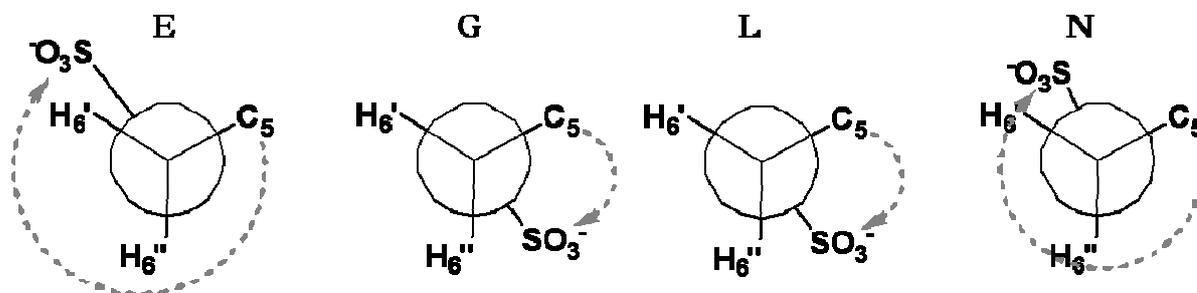


Figure 3.28: Newman projections for the four GalNAc6S units.

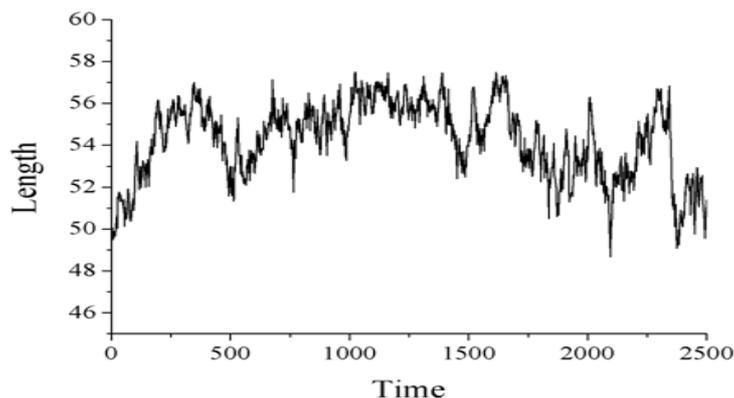


Figure 3.29: Fluctuation of the dodecasaccharide length during 2.5 ns of MD simulation

The length of the entire molecule was around 54.6 ± 1.8 Å, so as indicated by the plot in figure 3.29 and by the little RMSD value, the molecule assumed a quite elongated disposition, probably because more bent or compressed conformations led to unfavourable proximity between negative charged groups.

The estimation of the inter-protonic distances (Tab. 3.15) has indicated a NOE pattern in agreement with that measured for the lesser spotted dogfish; an expansion of the anomeric region of the NOESY spectrum reported (Fig. 3.30).

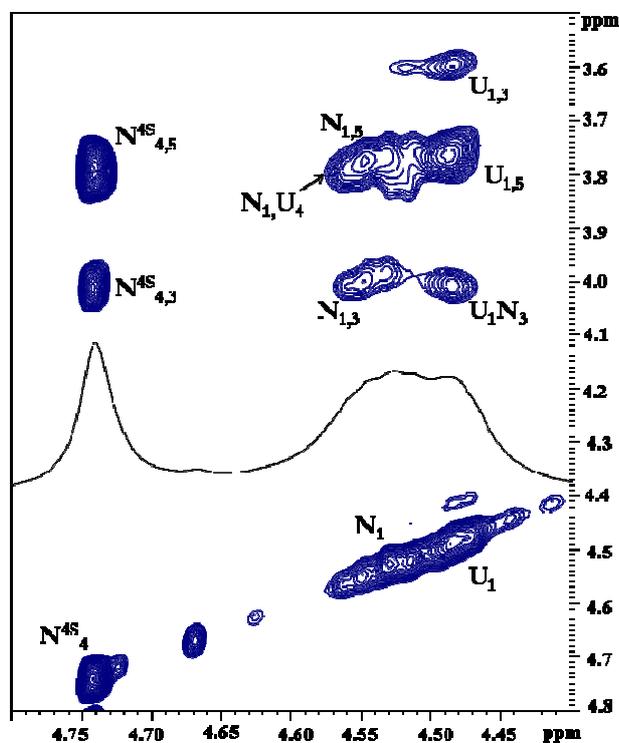


Figure 3.30: (D_2O , 600 MHz, 298 K) expansion of the anomeric region of CS NOESY spectrum. The principal NOE contacts are reported

Table 3.15: Averaged NOE distances

Linkage β -(1 \rightarrow 3)		Linkage β -(1 \rightarrow 4)	
NOE	Averaged distance	NOE	Averaged distance
A ₁ \rightarrow B ₂	4.30 \pm 0.20	B ₁ \rightarrow C ₃	4.45 \pm 0.2
A ₁ \rightarrow B ₃	2.3 \pm 0.2	B ₁ \rightarrow C ₄	2.45 \pm 0.2
A ₁ \rightarrow B ₄	3.7 \pm 0.4	D ₁ \rightarrow E ₃	4.5 \pm 0.1
C ₁ \rightarrow D ₂	4.05 \pm 0.23	D ₁ \rightarrow E ₄	2.4 \pm 0.2
C ₁ \rightarrow D ₃	2.42 \pm 0.20	F ₁ \rightarrow G ₃	4.4 \pm 0.2
C ₁ \rightarrow D ₄	3.9 \pm 0.4	F ₁ \rightarrow G ₄	2.4 \pm 0.2
E ₁ \rightarrow F ₂	4.2 \pm 0.3	H ₁ \rightarrow I ₃	4.5 \pm 0.2
E ₁ \rightarrow F ₃	2.4 \pm 0.2	H ₁ \rightarrow I ₄	2.4 \pm 0.20
E ₁ \rightarrow F ₄	3.6 \pm 0.5	L ₁ \rightarrow M ₃	4.4 \pm 0.2
G ₁ \rightarrow H ₂	4.3 \pm 0.2	L ₁ \rightarrow M ₄	2.4 \pm 0.2
G ₁ \rightarrow H ₃	2.3 \pm 0.2		
G ₁ \rightarrow H ₄	3.6 \pm 0.3		
I ₁ \rightarrow L ₂	4.3 \pm 0.20		
I ₁ \rightarrow L ₃	2.3 \pm 0.2		
I ₁ \rightarrow L ₄	3.5 \pm 0.5		
M ₁ \rightarrow N ₂	3.9 \pm 0.4		
M ₁ \rightarrow N ₃	2.5 \pm 0.3		
M ₁ \rightarrow N ₄	3.9 \pm 0.5		

In particular, the possibility of generating strong NOE contacts was estimated just between the protons across the glycosidic linkages, as indicated by the distances values up to 2.7 Å. The absence, in the NOESY spectrum, of contacts generated by non-*exo*-anomeric disposition was confirmed by the corresponding distances higher than 3.5 Å.

Applying the same approach used for the unsulfated molecule (see 3.2.4), the analysis of the existence and persistence of H-bond across the glycosidic linkages was carried out as well (Fig. 3.31), in to order to establish the effect on these parameters of the sulfate group presence.

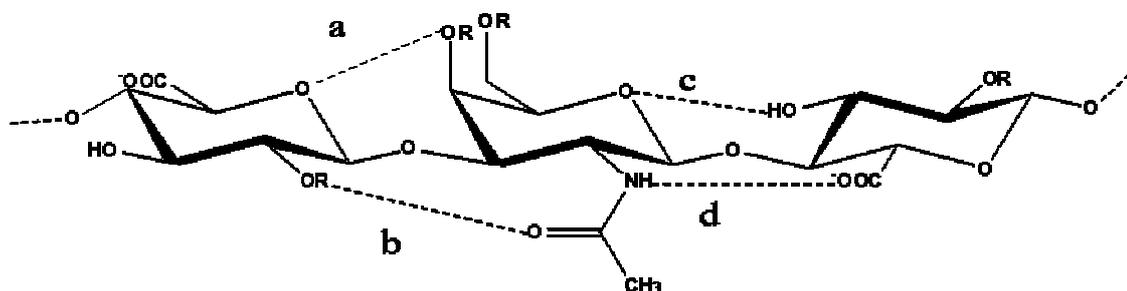


Figure 3.31: Graphical representation of H-Bond types investigated and labels used during the description. R=H or SO₃⁻ group.

Type a

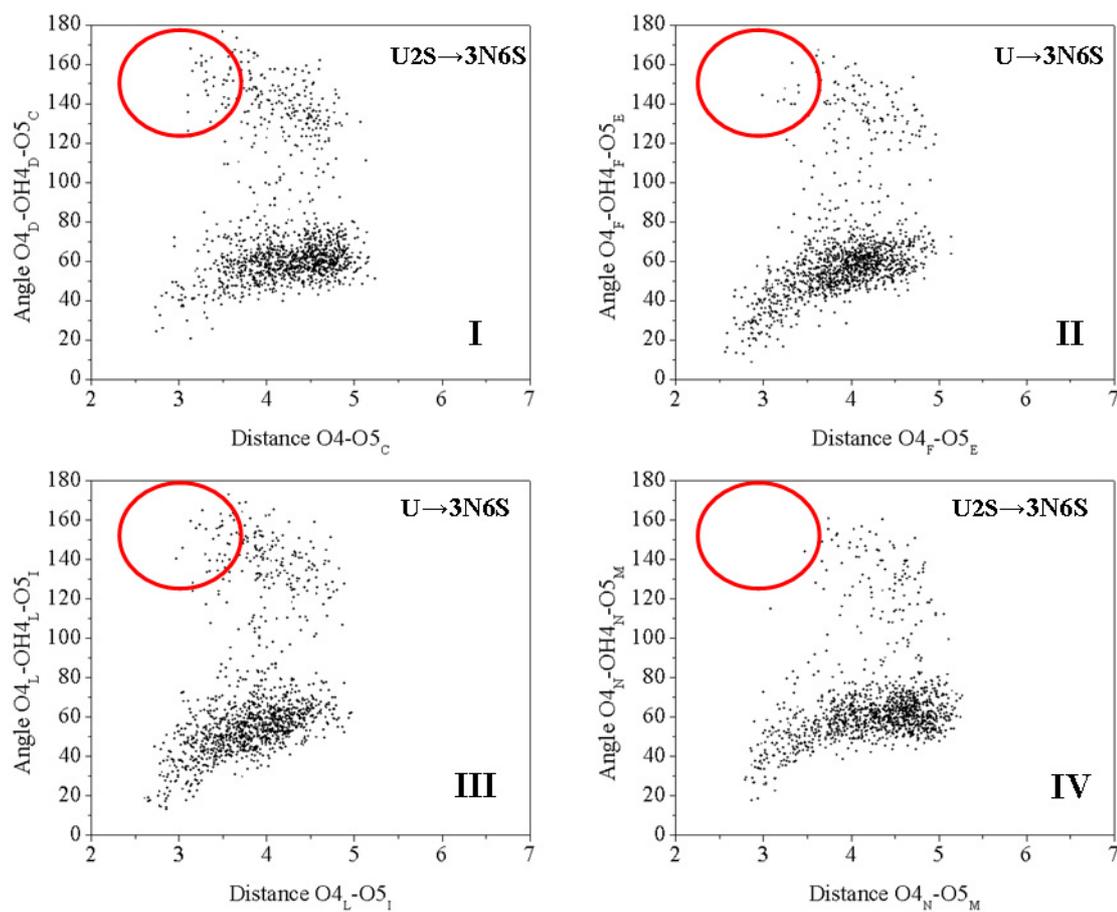


Figure 3.32a: Scattered plots representing the distances and the angles of the H-bonds type a across the linkage across the linkage β -(1→3).

Type b

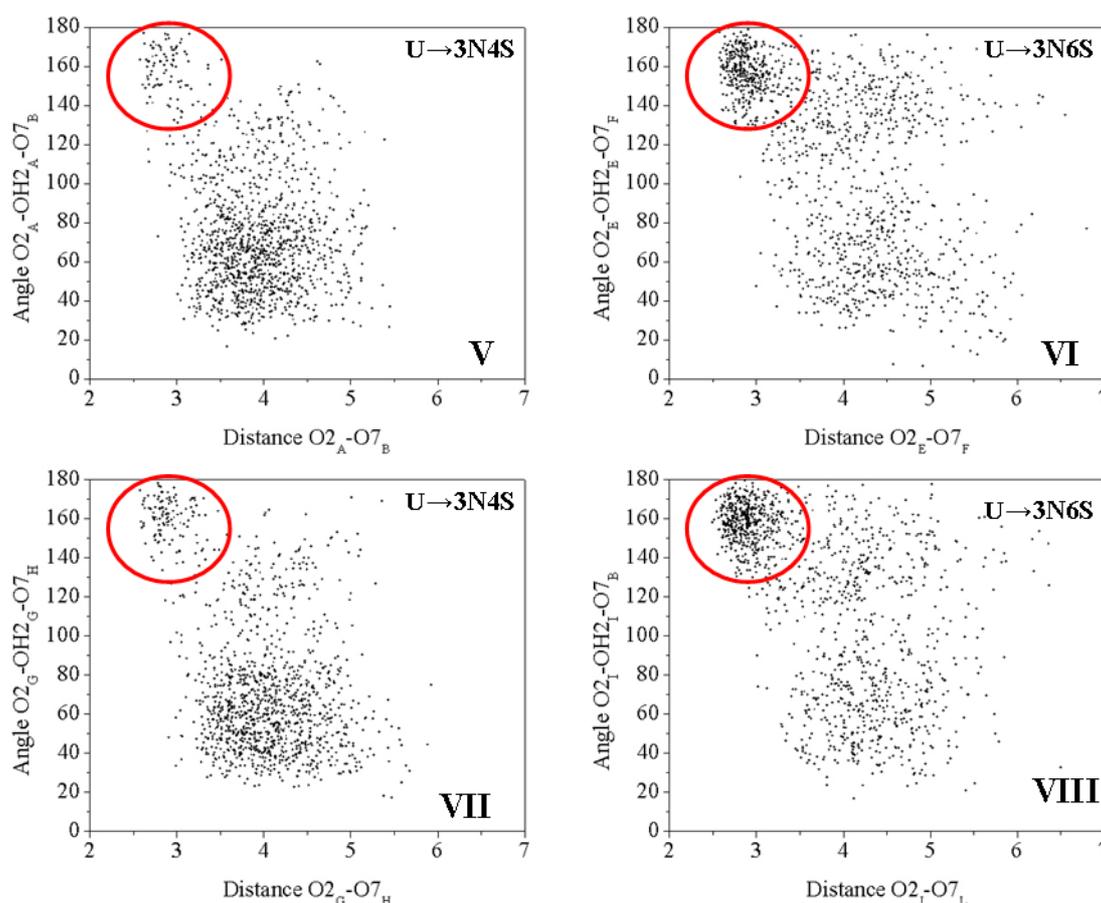


Figure 3.32b Scattered plots representing the distances and the angles of the H-bonds type b across the linkage across the linkage β -(1 \rightarrow 3).

Looking at the plots in Figures 3.32 and 3.33 for both the linkages, a medium strength H-bond was found in the 40% of the structures. In particular, linkage β -(1 \rightarrow 4) presented H-bond (\sim 35% of persistence) between the OH in position 3 of the GlcA units and the *endo*-cyclic oxygen of the previous GalNAc unit (Figure 3.31 type c), while for the linkage β -(1 \rightarrow 3) the H-bond present in the 40% of the structures was the one between the OH in position 2 of the GlcA unit and the carboxylic group of the *N*-Acetyl of the next GalNAc6S unit (Figure 3.31 type b, VI, VIII Figure 3.32b). The H-bond involving the OH in position 2 of the GlcA units and the carboxylic group of the *N*-Acetyl of the next GalNAc4S showed a lesser persistence (\sim 10 %)(Figure 3.31 type b, V, VII Figure 3.32b).

The other possible H-bond across the linkage β -(1 \rightarrow 3), namely the one between the OH in position 4 of the GalNAc unit and the *endo*-cyclic oxygen of the previous GlcA (type a Figure 3.31), in all was could be considered absent, counting lesser than 1%, as indicated in Figure 3.32a by the plots I-IV.

Type c

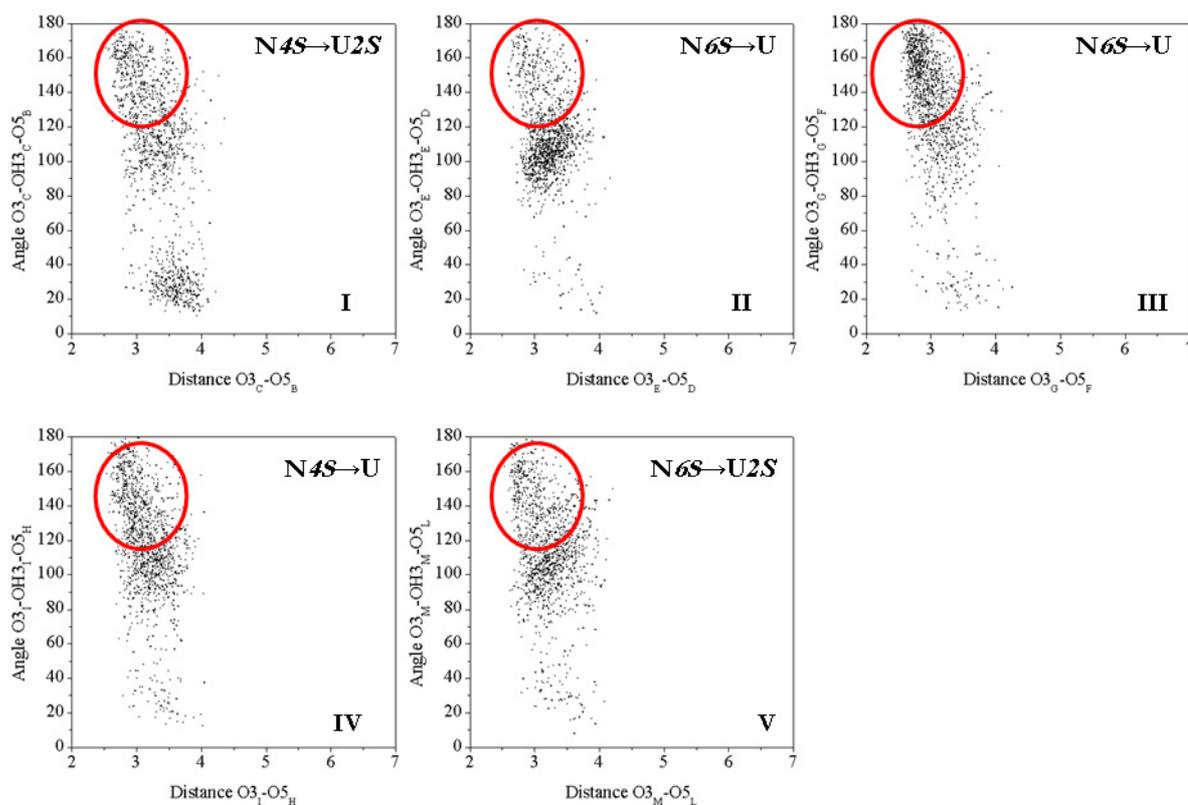


Figure 3.33a: Scattered plots representing the distances and the angles of the H-bonds type c across the linkage β -(1 \rightarrow 4).

The investigation on other H-bond characteristic of the β -(1 \rightarrow 4), namely the one between NH proton of the GalNAc unit and the carboxyl group of the next GlcA (type d Figure 3.31), was complicated by the presence on the carboxylate group of two equivalent oxygen atoms. For this reason, the estimation was performed for each carboxylic oxygen and the total persistence was calculated as a sum of the two values, due to the equivalence of the oxygens on a charged carboxylic group. The plots (Fig. 3.33b), as well, were the sum of the data belonging to both the oxygens, two different colours were used to represent the points generated by an oxygen or by the other. The H-Bond with the highest persistence value (\sim 70%) was the one between the units G and F (III, Figure 3.33b). The others showed a value in the range between 25-40%. The presence of water-bridge was not investigated.

Type d

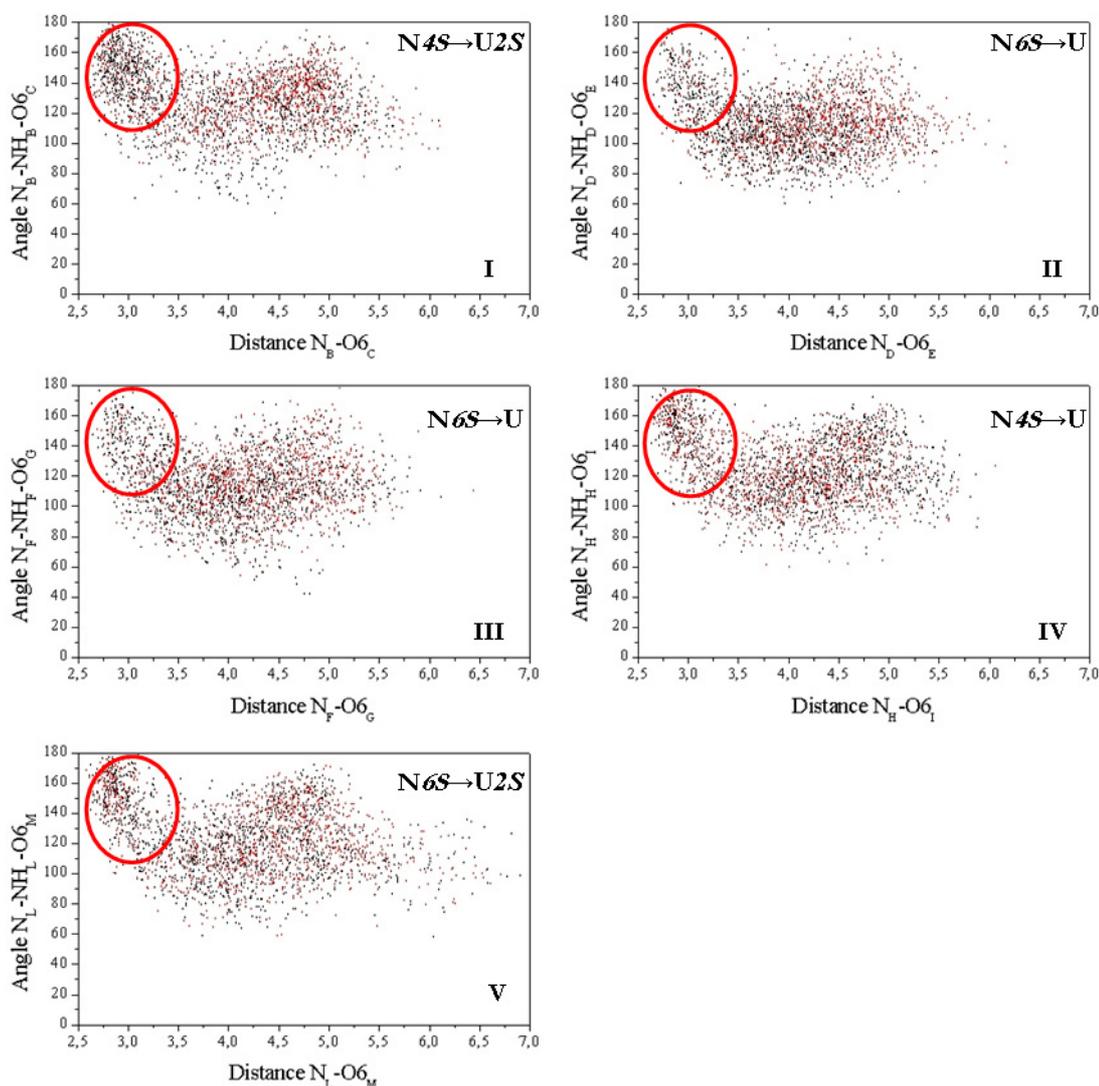


Figure 3.33b: Scattered plots representing the distances and the angles of H-bonds type d across the linkage across the linkage β -(1 \rightarrow 4).

3.4.4 Comparison with data of unsulfate molecule

Analysis carried out on the MD simulation indicated a good consistence with experimental data, so a comparison between these with those collected for the unsulfated one, was performed too.

In general, it was established that the presence of sulfate groups did not affect the sugar chair conformation, the *N*-Acetyl group disposition and the preferred rotamer of the CH₂OH group.

More over, the presence of a sulfate group did not cause a particular modification in the type of the H-bonds detected, in both molecules the linkages β -(1 \rightarrow 4) were more stabilized by direct H-bonds whit respect the β -(1 \rightarrow 3) one.

Relevant differences were detected for the torsional angles Φ and Ψ , as indicated by the plots in the figure 3.34 and 3.35, which display the scattered plots of the different sulfated motifs of the dodecasaccharide superimposed to those of the analogue unsulfate motifs.

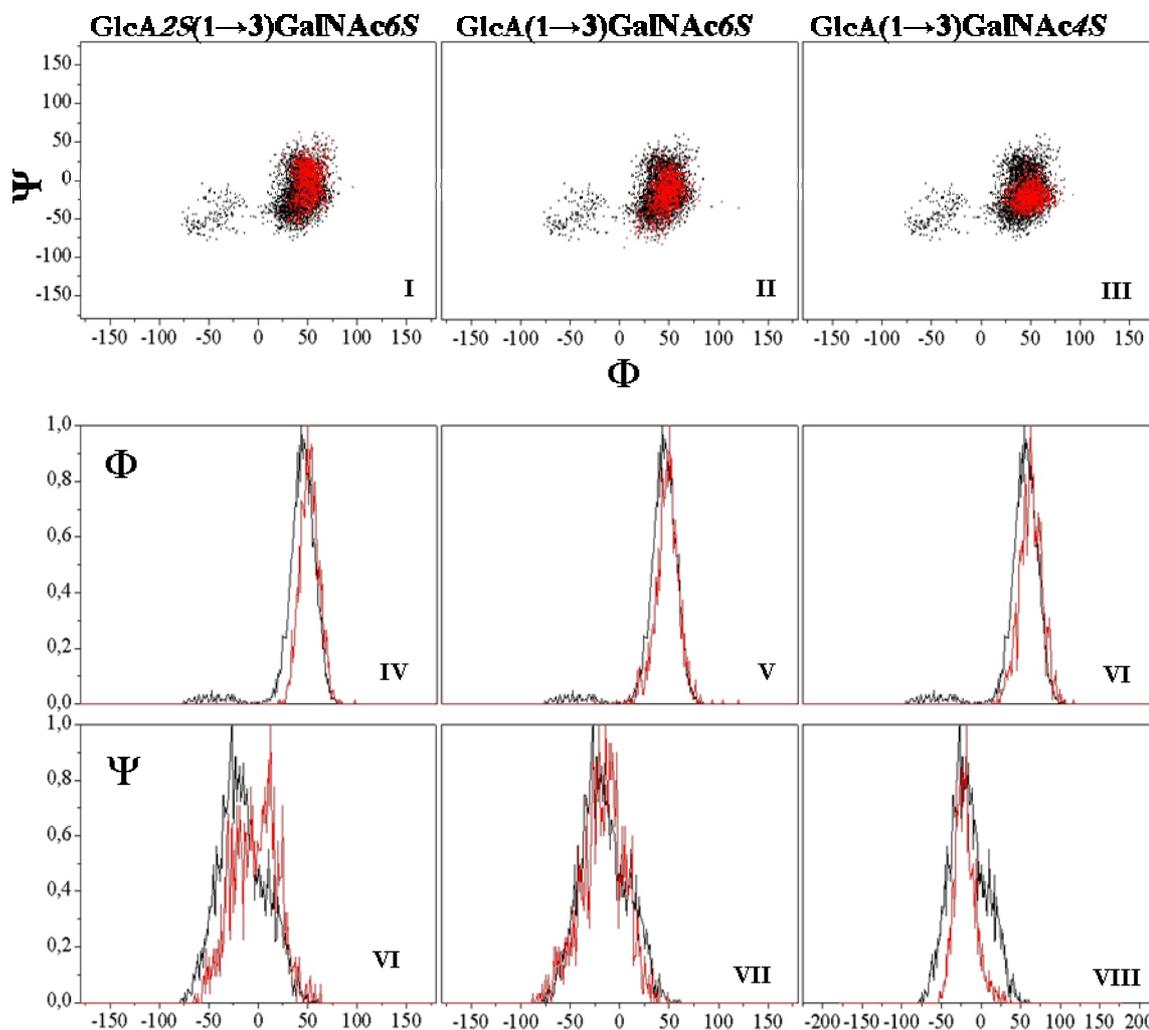


Figure 3.34: Comparison between the data of the different sulfated motifs of the dodecasaccharide containing the β -(1→3) linkage (red points and lines) and those of the unsulfate motif (black points and lines)

Plots in the figure 3.34 showed that the presence of a sulfate group only in position 6 did not caused particular changes in the behaviour of the linkage β -(1→3)(Fig. 3.34, II), but when there was a sulfate group also on the GlcA unit, the preference for positive Ψ values was enhanced (Fig. 3.34, I).

Among the three different sulfated motifs around the β -(1→3) linkage, the one that was more affected was that with the sulfate group in position 4 of the GalNAc unit (Fig. 3.20, III). In this case, a reduced fluctuation for the Ψ values was highlighted, with a great preference for negative values (Fig. 3.34, III).

For all the different sulfated disaccharides containing a β -(1→4) linkage, a similar behaviour was detected, as indicated by the plots in figure 3.35. In particular the presence of a sulfate group

affected consistently Ψ distribution, which was shifted to more positive values with respect to the distribution of the unsulfate motifs.

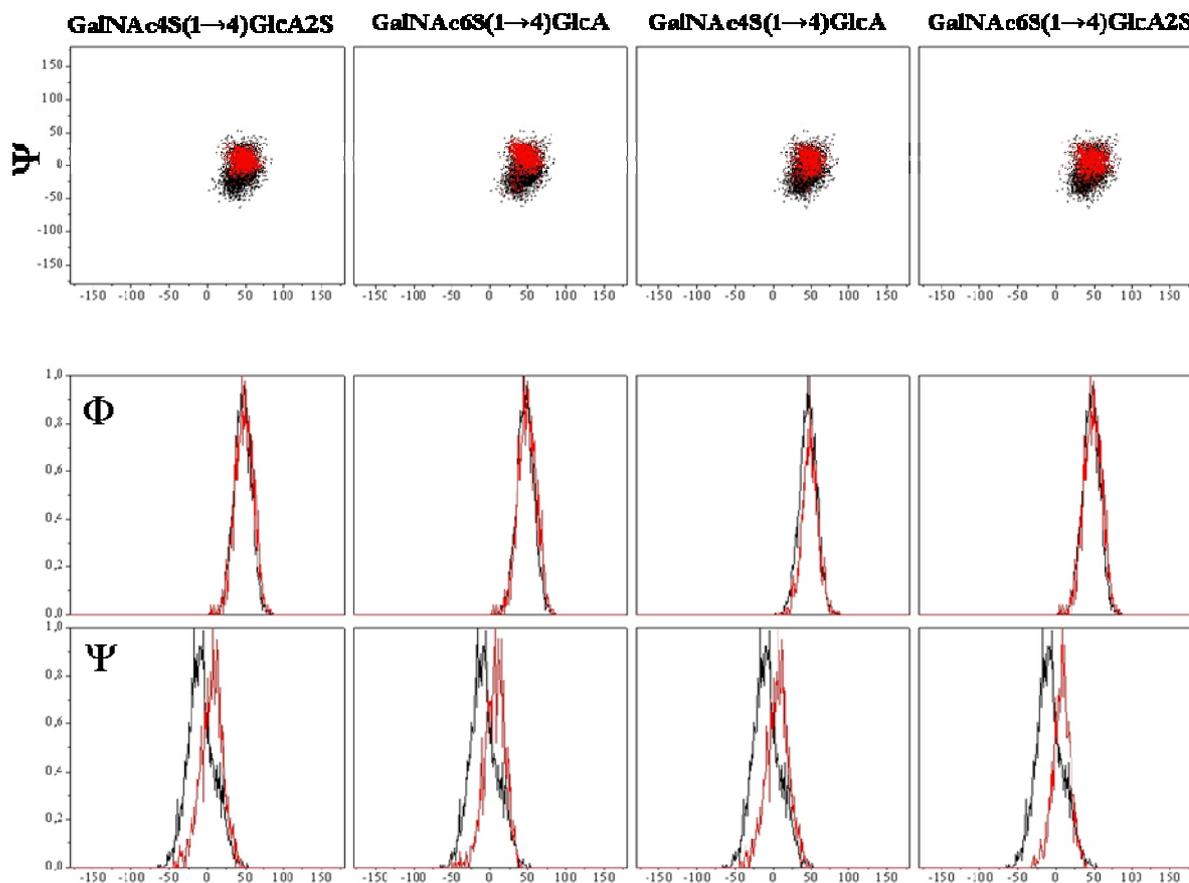


Figure 3.35: Comparison between the data of the different sulfate motifs of the dodecasaccharide containing the β -(1→4) linkage (red points and lines) and those of the unsulfate motif (black points and lines)

3.5 Molecular mechanics and dynamics on oligosaccharides of keratan sulfate

Keratan sulfate is a sulfated polylactosamine structure, so it has a backbone commonly found in O-glycans linked to glycoproteins³⁰.

The conformational behaviour of lactosamine has been investigated³¹, but the information available about sulfated structures are poor. These studies on unsulfated and sulfated KS oligosaccharides, so, have been performed in order to obtain a collection of new conformational data and to amplify the information about this GAG, that with respect to the others is less known.

³⁰ Brockhausen, I., Schochter, H., Stanley, P. Essential of Glycobiology, II Edition Eds: Varki, A., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. 2009 Chap.9:115-127.

³¹ Shim, G., Shin, J., Kim, Y. 2004 *Bull. Korean. Chem. Soc.* 25(2):198-202.

Molecular mechanics studies were performed on unsulfated and sulfated disaccharides to establish the conformational space accessible by the two β glycosidic linkages, additionally the effect of the sulfate groups on the torsional behaviour of the junctions was investigated as well. Molecular dynamics simulations in implicit solvent on polylactosamine and KS oligosaccharides, then, were used to understand how both the linkages behaved one reciprocally to the others.

3.5.1 Molecular mechanics on unsulfate and sulfate oligosaccharides

The first information about the behaviour of both β linkages of KS were collected through molecular mechanics studies on the disaccharides listed in the tables (Tab. 3.16). The corresponding unsulfated structures were analyzed as well, and their data are reported too.

Table 6: Dihedral angles and energy values corresponding to the minima of the flexible maps of both the β linkages of KS.

Linkage	Disaccharide sequence	Angle	Values (degree)	Energy (kJ/mol)
β -(1 \rightarrow 3)	Gal-(1 \rightarrow 4)-GlcNAc	Φ	59.3	135.423
		Ψ	-2.1	
	Gal-(1 \rightarrow 4)-GlcNAc6S	Φ	59.1	133.404
		Ψ	-1.9	
	Gal6S-(1 \rightarrow 4)-GlcNAc6S	Φ	59	141.712
		Ψ	-2.2	
β -(1 \rightarrow 4)	GlcNAc-(1 \rightarrow 3)-Gal	Φ	65.9	129.065
		Ψ	49.4	
		Φ	58.6	128.270
		Ψ	-26.5	
	GlcNAc6S-(1 \rightarrow 3)-Gal	Φ	46.8	146.634
		Ψ	11.9	
		Φ	39.7	146.478
		Ψ	-43.3	
	GlcNAc6S-(1 \rightarrow 3)-Gal6S	Φ	38.8	154.554
		Ψ	-52.4	
		Φ	45.4	155.740
		Ψ	15.4	

The conditions and the parameters used were the same applied for HA and CS disaccharides (see sections 3.1.3 and 3.4.1).

The flexible maps are reported in Figure 3.36 and the corresponding minimum energy and both Φ and Ψ values are listed in Table 3.16. In all the cases for linkages a conformation in agreement with *exo*-anomeric effect was detected.

All the flexible maps corresponding to the β -(1 \rightarrow 4) linkage were characterized by a similar shape and one minimum, and looking at the Φ and Ψ values no variations were detected, also when there were one or two sulfate groups.

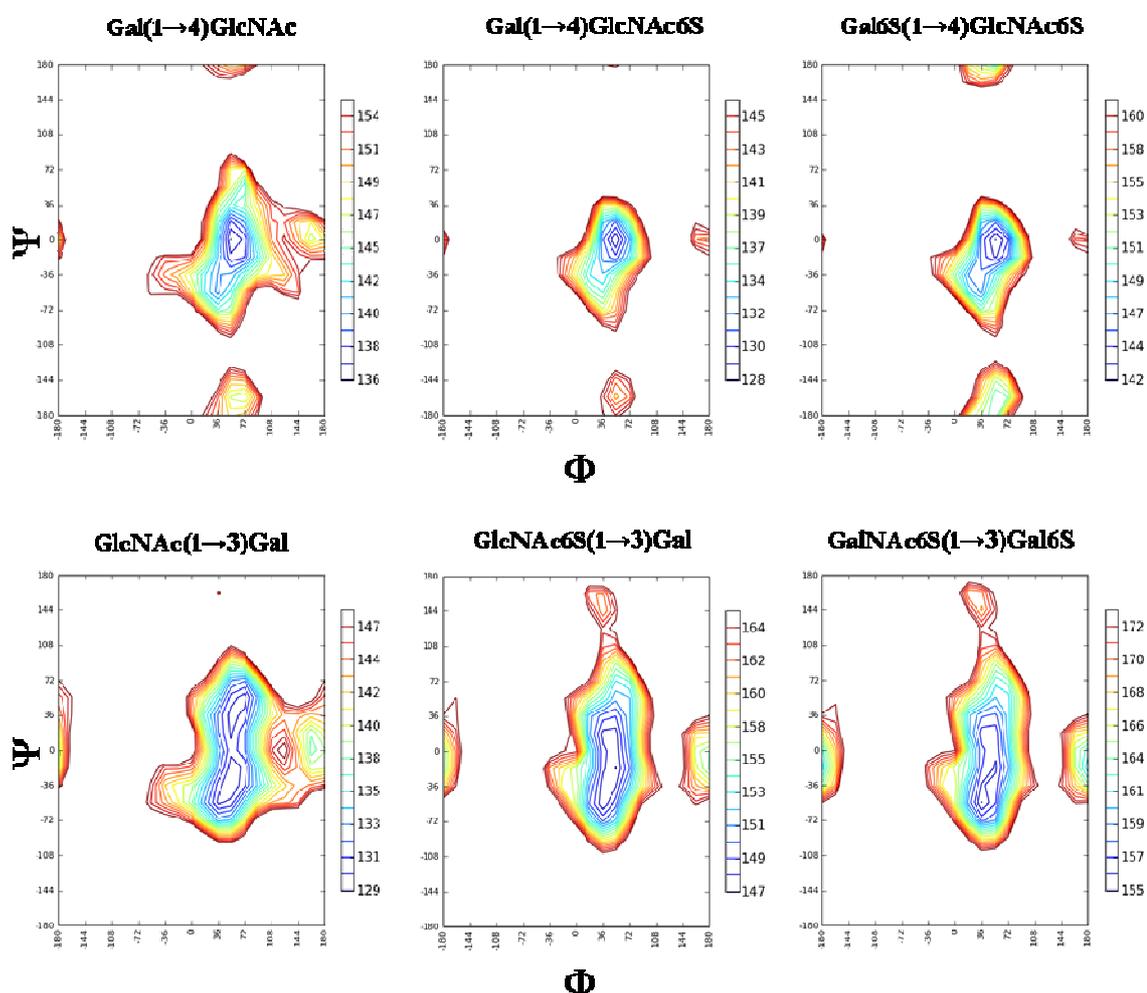


Figure 3.36: Flexible maps of the six KS disaccharides investigated through molecular mechanics calculations.

The flexible maps corresponding to the disaccharides containing the β -(1 \rightarrow 3) linkage, instead, showed a variation in correspondence of the number of the sulfate groups present. In all the cases two minima were detected: when the charged groups were absent the two minima were very different while, when there was a sulfate group on the GlcNAc unit, the difference visually decreased. The difference increased when there were two sulfate groups and a net preference for the conformation with negative Ψ values was highlighted.

3.5.2 Molecular dynamics in implicit solvent on unsulfated and sulfated oligosaccharides

The KS oligosaccharides used for the molecular dynamics simulations in implicit solvent had the same sequence but showed a different sulfation pattern. Their sequences are reported in Figure 3.37.

The unsulfated molecule was used as reference to understand the effect of the presence of one or two sulfate groups.

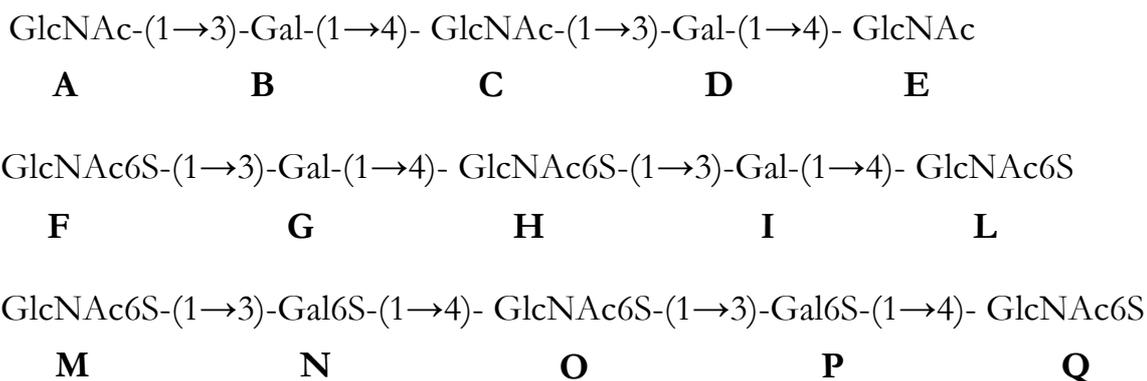


Figure 3.37: KS pentasaccharides sequences and labels used for to indicate the residues

The oligosaccharides were built up and the glycosidic torsions were set as the Φ and Ψ values of the lowest minima of the flexible maps (Tab 3.13). Three simulations of 6 ns were carried out with Macromodel 8.0¹³ using AMBER as force field and applying the GBSA solvent model. The temperature simulation was set at 288 K because the validation of the computational data was performed with the data collected at that temperature by NMR on the Keratan sulfate of the lesser spotted dogfish (see 2.3.3).

The analysis of the three MD simulations indicated that all the sugars maintained the classical 4C_1 , the *N*-Acetyl groups were in the typical *anti* disposition and the rotamer preferred by the CH₂OH group depended by the sulfation pattern. In particular, in the unsulfated molecules all the sugars preferred the *gt* rotamer; in both sulfated pentasaccharides the Gal units did not show a great preference between the *gt* and *gg* rotamers, while among the three GlcNAc units, the terminal one preferred the *gt* one and the other two the *gg* one.

All the glycosidic linkages assumed a disposition in accordance with the *exo*-anomeric effect as indicated by the scattered plots in Figure 3.38 and by the averaged Φ and Ψ values reported in Table 3.17.

Table 3.7: Averaged values of both the F and Y angle of all the three KS oligosaccharides

Angle	Average value	Angle	Average value	Angle	Average value
Φ_A	44.07 ± 13.77	Φ_F	40.12 ± 15.1	Φ_M	40.44 ± 14.97
Ψ_A	3.27 ± 34.73	Ψ_F	8.47 ± 24.18	Ψ_M	9.85 ± 23.54
Φ_B	48.36 ± 12.87	Φ_G	52.48 ± 12.06	Φ_N	47.13 ± 16.09
Ψ_B	-8.42 ± 15.97	Ψ_G	-3.84 ± 15.30	Ψ_N	-12.77 ± 24.59
Φ_C	43.44 ± 14.17	Φ_H	42.21 ± 14.59	Φ_O	41.72 ± 14.88
Ψ_C	4.62 ± 34.96	Ψ_H	5.65 ± 31.05	Ψ_O	13.49 ± 25.30
Φ_D	47.32 ± 12.32	Φ_I	51.26 ± 12.17	Φ_P	52.43 ± 11.65
Ψ_D	-7.12 ± 14.49	Ψ_I	-5.7 ± 14.75	Ψ_P	-3.90 ± 14.78

Conformational space covered during the MD simulation by β -(1 \rightarrow 4) linkages was restricted and very similar to the one predicted by molecular mechanics calculations.

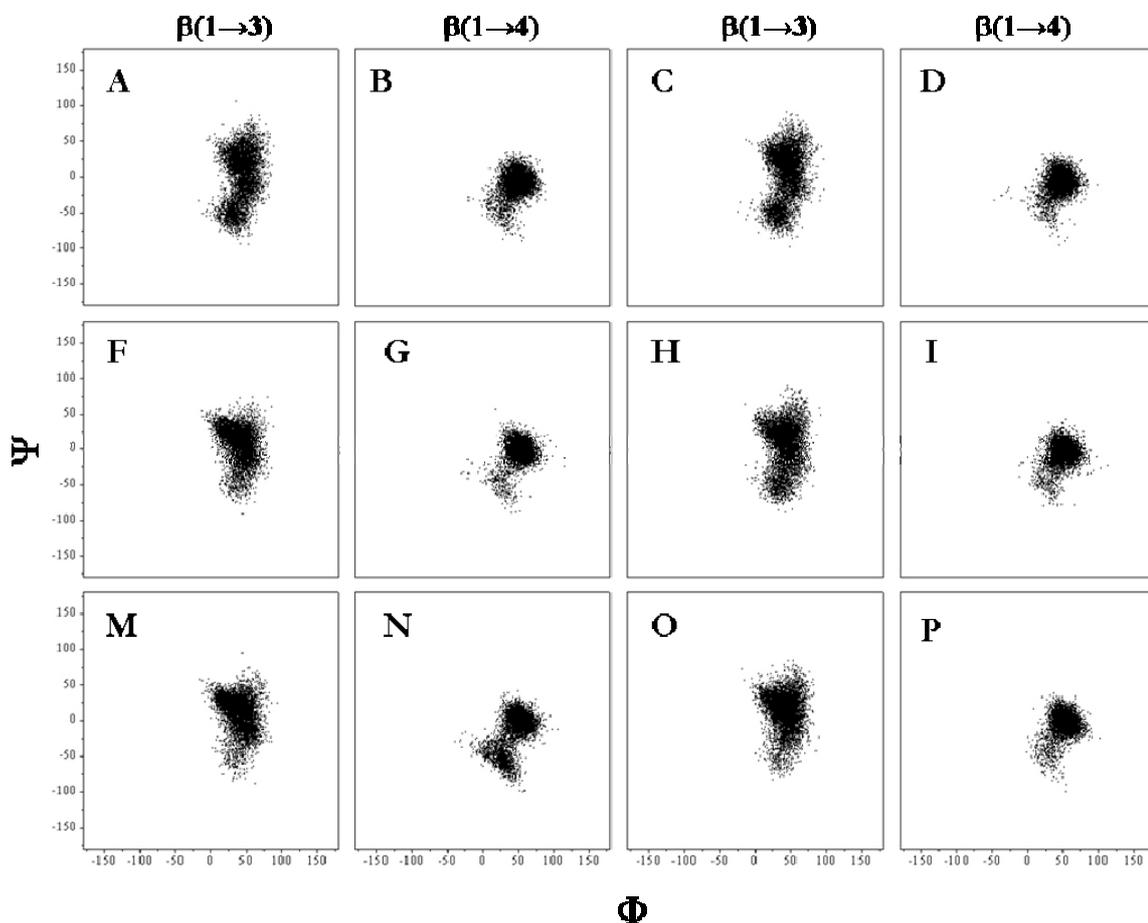


Figure 3.38: Scattered plots of all the glycosidic junctions of the three KS oligosaccharides.

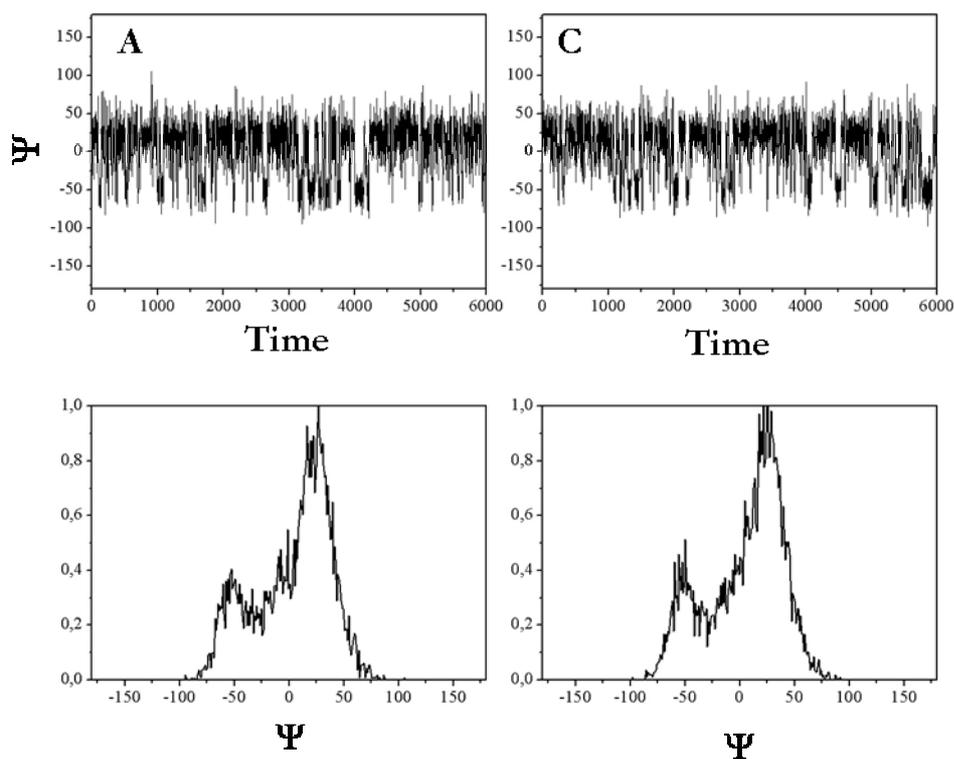
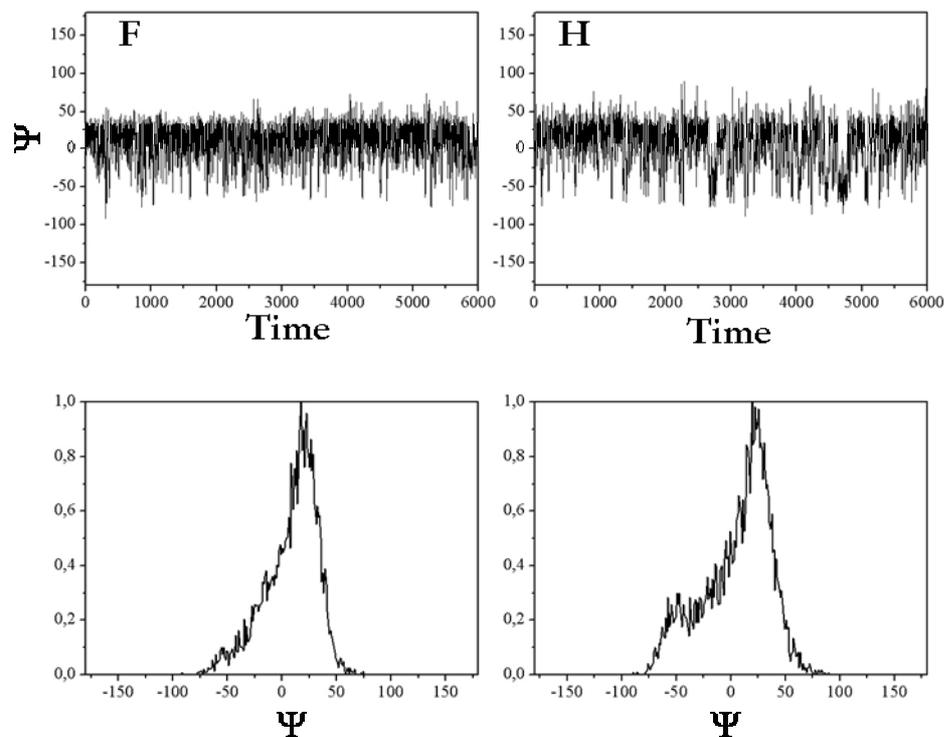


Figure 3.39: Trajectory of the dihedral angle Ψ of the β -(1 \rightarrow 3) linkages of the unsulfated oligosaccharide

The linkages β -(1 \rightarrow 3) showed a higher fluctuations on the dihedral angle Ψ , as indicated by the higher RMSD values associated to the averaged angle value and by the trajectories of the Ψ angles reported in Figures 3.39, 3.40 and 3.41, with the corresponding projections.



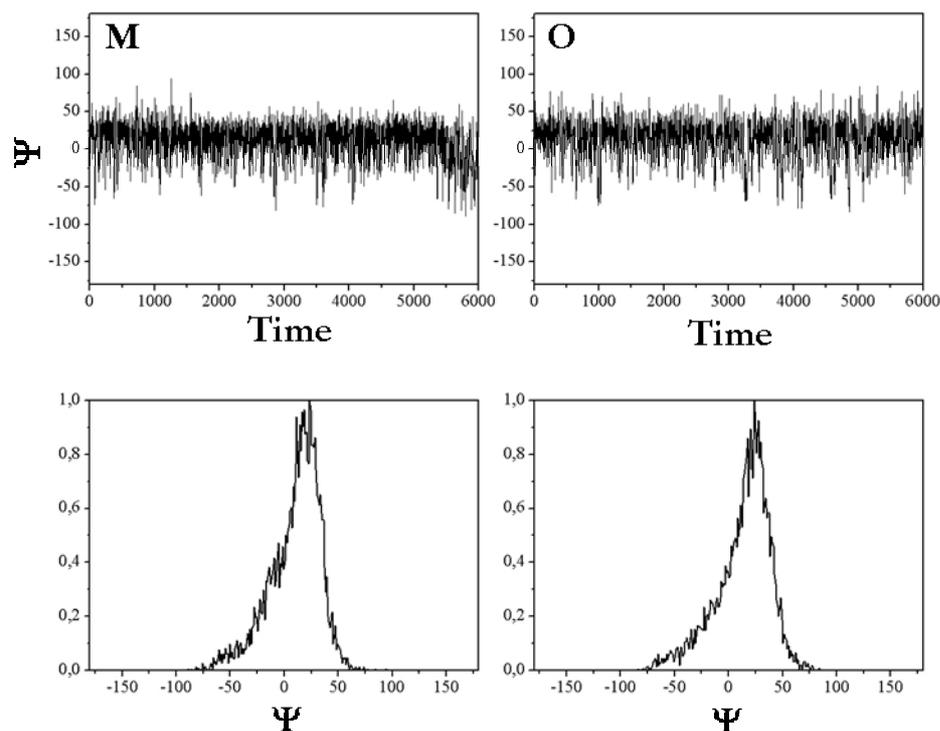


Figure 3.40: Trajectory of the dihedral angle Ψ of the β -(1 \rightarrow 3) linkages of both the sulfated oligosaccharides

Looking at the scattered plots of the β -(1 \rightarrow 3) in Figure 3.38, two clusters of conformations were individuated: the main cluster was characterized by positive Ψ values, while the other by negative ones. The entity of the second cluster decreased with the growing number of sulfate groups as indicated by the projections reported in Figures 3.39, 3.40 and 3.41. Moreover, for the linkage β -(1 \rightarrow 3), through the MD simulation a behaviour different from the one predicted by molecular mechanics calculation was highlighted, because the main cluster of conformations adopted a disposition characterized by a positive Ψ (Fig 3.38) and not a negative one, as indicated by the flexible maps in Figure 3.36.

The disposition of the sulfate groups was established on the basis of the dihedral angle $C_{x-1}-C_x-O-S$ value; in both the sulfated molecules for the Gal and GlcNAc units a value around 180° was estimated and a disposition staggered with respect both H6 protons was found (Fig. 3.41).

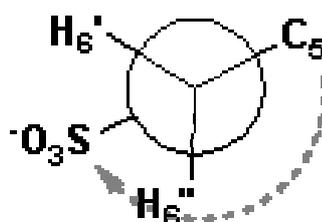


Figure 3.41: Newman projection of the sulfate dihedral angle

The estimation of the inter-residual proton distances of both the simulations on sulfated molecules confirmed the data collected through the analysis of the NOESY spectrum (Fig 3.42).

The analysis of all the MD data on the different KS pentasaccharides led to conclude that the sulfate groups influenced first of all the glycosidic torsions of the polyglucosamine backbone, with a higher effect on the β -(1 \rightarrow 4) linkage. The NOE pattern, the *N*-Acetyl disposition and the CH₂OH preferred rotamer, instead, showed a similar behaviour in all the structures considered, sulfated and unsulfated ones.

3.6 Conclusions

Molecular modelling studies carried out on glycosaminoglycans oligosaccharides combined with NMR data, allowed a deep description of the local solution conformation of such complex and heterogeneous structures, and also of the role of intramolecular hydrogen bonds and of water molecules.

These studies confirmed the structural similarity between HA and unsulfated CS, because for both structures a conformation compatible with a 3-fold left handed was estimated. On the other hand, however, the conformational studies highlighted also some difference between the two chains, in particular they indicated that the axial disposition of the OH group in the hexosamine residue in the CS backbone caused a higher flexibility across the β -(1 \rightarrow 3) linkage and the absence of a organized network of H-Bonds across both the β linkages, as found in hyaluronic acid.

The studies carried out on sulfated molecules suggested that charged groups, indicated that the sulfated were not so influential into the conformational behaviour of the molecules. In particular it was established that sulfate groups affected, as expected, the glycosidic torsions reducing the their fluctuation, but they did not influence the H-bond network, the disposition of *N*-Acetyl groups or the C₅-C₆ rotamer distribution.

3.7 Material and Methods

3.7.1 Isolation of oligosaccharides

Isolation of the deca-saccharide HA10

Commercial hyaluronic acid (Hyaluronic acid sodium salt from *Streptococcus equi*, Biochemica, 53747), was partially depolymerised using testicular hyaluronidase producing saturated oligosaccharide of various length. The enzymatic digestion was performed using the conditions reported². The polymer was dissolved in 0.25 M NaNO₃, 5 mM Na₂HPO₃, pH 4 to have a final concentration of 5 mg/mL, and treated with HAse (1mg of enzyme for 10mg of HA, Hyaluronidase from bovine testes, type I-S, Sigma, H3506) at 37°C for 5 hours. The enzyme was then denaturized boiling the solution for 15 minutes.

The resulting mix of oligosaccharide was then fractionated by SEC using a AcA202 ultrogel (115 x 1.5 cm, flow 0.2 ml/min using as aqueous eluent 0.25 M acetic acid, and 0.28 M pyridine). The fractions were pooled on the basis of the chromatogram profile, reduced under-vacuum, and desalted on Sephacryl G-10 (95 x 1.5 cm, flow 0.23 ml/min, H₂O as eluent). The oligosaccharide was identified and its purity estimated by MALDI spectrometry. The deca-saccharide was obtained with a final yield of .

All the chromatographic separations were monitored online with a refractive index refractometer (K-2310 Knauer).

Isolation of the hexasaccharide from defructosylated esopolysaccharide of *E. coli* K4

The defructosylated esopolysaccharide of *E. coli* K4, obtained as reported¹¹, was partially depolymerised with testicular hyaluronidase producing saturated oligosaccharide of various length. The enzymatic digestion was carried out applying the analogous conditions reported above for the hyaluronic acid, but modifying the amount of enzyme and the time of digestion considering the specificity of the enzyme with respect to the substrate.

The polysaccharide was dissolved in the same buffer reported above and treated with HAse (2 mg of enzyme for 10 mg of product) at 37°C for 60 hours. The enzyme was then denaturated boiling the solution for 15 minutes.

The resulting mix of oligosaccharide at first desalted by SEC on a Sephadex G10 (90 x 1.5 cm, flow 0.2 ml/min, H₂O as eluent) and then it was fractionated using a AcA202 ultrogel.

The fractions were collected and treated as reported above. Then each sample was proton-exchanged with a dowex resin in order to turn every carboxylic group in its neutral form.

Again, the oligosaccharide was identified and its purity estimated by MALDI spectrometry. The hexasaccharide was obtained with a final yield of .

3.7.2 NMR spectroscopy

The ¹H, and ¹H-¹³C NMR experiments of the oligosaccharides in isotropic conditions, were carried out on a Bruker DRX-600 equipped with a reverse probe, spectra were calibrated with respect to the internal acetone ($\delta_H = 2.225\text{ppm}$; $\delta_C = 31.45\text{ppm}$).

For all the homonuclear spectra, experiments were measured with data sets of 2048 x 512 points, a mixing time of 200 and 120 ms was employed for NOESY and TOCSY, respectively. Each data matrix was zero-filled in both dimensions to give a matrix of 4096 x 2048 points, and was resolution-enhanced in both dimensions by a shifted sine-bell function before Fourier transformation.

The heteronuclear experiments were measured using a data set of 2048 x 512 points or 2048 x 25, while the pulse sequence was optimized for a 140 Hz coupling constant. During processing, matrix was extended to 4096 x 1024 points by forward linear prediction extrapolation.

All NMR spectra were acquired, transformed, and analyzed with the Topspin 2.1 program.

NMR spectroscopy HA10

Isotropic condition: the decasaccharide (15mg) was dissolved in PBS 10 mM pH 7 in D₂O, and transferred in an NMR tube.

On the sample, COSY, TOCSY, and NOESY experiments were performed as homonuclear spectra, while HSQC and coupled-HSQC as heteronuclear. In particular for the HSQC 16 scans were acquired while for the coupled-HSQC 40 scans were acquired.

Anisotropic condition: the disaccharide was dissolved in a mixture of PBS 10 mM pH 7 in D₂O, pentaethylene glycol mono-octyl ether (C₈E₅), and *n*-octanol prepared as reported⁵. The weight percentage for the ratio C₈E₅/ water was of 3%, and the molar ratio of CmEn to *n*-alkyl alcohol was 0.87.

The resulting quadrupolar splitting of the mixture in the ²H-NMR spectrum was of 14.8 Hz

The coupled-HSQC experiments on the sample were performed using a data set of 8192 x 256 points acquiring 140 scans for each t1 value. During processing, matrix was extended to 4096 x 1024 points by forward linear prediction extrapolation.

For both the isotropic, and anisotropic conditions three coupled-HSQC experiments were performed in order to mediate the results.

NMR spectroscopy hexasaccharide

On the sample dissolved in D₂O, COSY, TOCSY, and T-ROESY³² were recorded as homonuclear spectra, while HSQC and HMBC as heteronuclear ones. The T-ROESY was performed instead of the NOESY, due to the oligosaccharide nature of the molecule. T-ROESY experiment was recorded with a data set of 2048 x 512 points and a mixing time of 200 ms. The cross peak intensities were corrected by their corresponding offset effects in accordance to the effective field for each proton resonance of the spectrum. The spin-lock field was attenuated four times (ca 8000 Hz) with respect to the one employed for hard pulses (ca 33000 Hz).

The HSQC and HMBC experiments were measured in the ¹H-detected mode by single-quantum coherence with proton decoupling in the ¹³C domain, using a data set of 2048 x 256 points, and acquiring for each t1 value respectively 80 and 100 scans. During processing, matrix was extended to 4096 x 1024 points by forward linear prediction extrapolation.

³² T.L., Hwang, A.J., Shaka, 1992 *J. Am. Chem. Soc.* 114, 3157-3158.

In all homonuclear spectra for the experiments in 10 % of D₂O, water suppression was achieved by using excitation sculpting with gradients³³. ¹H-NMR spectra were recorded with 32 K data points varying the temperature between 288 and 310K with increments of 2 K. NOESY spectra were measured using a data sets of 4096 x 512 points with mixing times between 200 ms. The data matrix was zero-filled in the F1 dimension to give a matrix of 4096 x 2048 points, and resolution was enhanced in both dimensions by a sine-bell shifted Qsine function before Fourier transformation.

Temperature coefficients ($\Delta\delta/\Delta T$) were determined calculating the angular coefficient of the interpolation line of the points defined by the $\Delta\delta$ and ΔT values of each proton.

3.7.3 Refinement of RDC data

The RDC experimental data were analyzed while the alignment tensor was determined with Mspin software 1.0¹⁷.

The 1750 molecular structures, obtained through a MD simulation in explicit solvent, were loaded in the maestro file format. Experimental RDC values were loaded in a data block built up listing at first the coupled atoms, and then the corresponding measured value as requested by the software. (For the table of RDC values see section 3.1.5).

The computation algorithm selected for the RDC calculation was TRAMITE⁶, assuming that the vectors system of the alignment tensor has the same orientation of the inertia vector .

3.7.4 Molecular mechanics calculations

Molecular mechanics calculations were performed on the unsulfated and sulfated disaccharides using the AMBER force field as implemented in the Macromodel¹³ 8.0 software package. Flexible maps were calculated using the DRIV utility. The bulk of the solvent was simulated using a constant dielectric of 80. For each disaccharide both Φ and Ψ dihedral angles (defined as H₁-C₁-O-C_{aglycon} and H_{aglycon}-C_{aglycon}-O-C₁ respectively) were varied incrementally using a grid step of 18 degrees. The corresponding flexible maps were drawn as 2D contours plots using the graphical tools of Macromodel 9.0.

³³ Hwang, T.-L., and Shaka, A.J. 1995 *J. Magn. Reson. Series A* 112, 275–279.

3.7.4 Molecular dynamics calculation

3.7.4.1 MD in explicit water

Molecular dynamic calculations were performed in explicit water using the Glycam06¹⁶ force field and the AMBER 9 software package¹⁵.

All the oligosaccharides were built up using the carbohydrate builder utility of the glycam website³⁴, in order to have a file .pdb with the appropriate atom types as requested by the AMBER 9 software and then the torsional angles were set to the values obtained through the molecular mechanics calculations.

The sulfate groups were put on the molecules using the graphical software SYBYL³⁵ and the partial charges of the corresponding sulfated oligosaccharides were calculated with Gaussian03²⁸. In particular, the partial charges were estimated through an *ab initio* approach using an Hartree-Fock (HF) calculation with a 6-31G* basis set, a full convergence criteria for the self consistent field calculation and a Mulliken population analysis. Applying the RESP methodology, then, the electrostatic potential of the minimum conformation was calculated and the partial charges were fit to the molecular electrostatic potential³⁶.

The input files, were generated using the antechamber at first (.prepin, .frcmod) and then xleap (.prmtop and .inpcrd) modules of the software package, all the minimization and molecular dynamic calculations were performed using the SANDER module.

The molecules were solvated with TIP3P water molecules in a truncated octahedric box, the net negative charge was neutralized adding sodium cations as counterions.

In all the cases the system underwent, at first, a minimization with positional restraints, in particular the solute remained fixed using a force constant of 500 kcal/mol and the water molecules and the counterions were free to move. The minimization was carried out under periodic boundary conditions and applying 500 steps of the steepest descent energy minimization followed by 500 cycles of conjugate gradient minimization. A second minimization was then performed leading the entire system to relax. In this case a total of 2500 steps of minimization were performed, just 1500 cycles were of conjugate gradient of minimization. The minimized system, then, was heated up from 0 K to the work temperature of 298 K through a molecular dynamics simulation of 100 ps. The MD simulation was performed applying periodic boundary conditions and the SHAKE protocol, constraining the all bonds involving hydrogen to their equilibrium values. The Langevin equilibration scheme was selected for the temperature regulation of the system and 50,000 molecular dynamics steps were run with a time step of 2 fs

³⁴ Website: www.glycam.com

³⁵ Tripos Associates, St. Luis, MO.

³⁶ Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman P.A. 1993 *J. Phys. Chem.* 97:10269-10280

using a weak force constant to restrain the solute motions. An equilibration of the whole system was then performed through a new MD simulation of 100 ps with both volume and pressure constant. After this equilibration step the system underwent the real molecular dynamic simulation of 8 ns applying all the parameters used for the equilibration passage.

In the case of HA10 a cut off of 11Å was applied, while for the other molecules one of 14 Å.

Ptraaj was the utility used for analyzing and processing trajectories and coordinate files created from the MD simulations. Maestro 9.0 was used as well, after the appropriate conversion of the output files.

3.7.4.2 MD in implicit solvent

Molecular dynamics simulations in implicit solvent were carried out using AMBER force field implemented in Macromodel 8.0 installed under the Red Hat 8 operating system. Simulations were performed starting from structures with glycosidic torsions set as the optimized values, calculated through molecular mechanics calculations. Water was approximated using the GB/SA solvent model. Structures underwent initially an equilibration steps of 150 ps and successively were kept in a thermal bath at 288 K for 6000 ps; a dynamic time-step of 1.5 fs together with the SHAKE for Hydrogen-bond atoms were applied. The coordinates were saved every ps of simulation leading to the collection of 6000 structures.

The analysis of the simulation data was performed using the tools of Maestro 9.0 software package and ORIGIN 7.5 software.

3.7.5 Helix parameters estimation

The possible single helical forms for the HA deca-saccharide and for the hexa-saccharide of unsulfated CS were established using the program POLYS¹⁸. The different single helical forms were extrapolated starting from the average values of the dihedral angles Φ and Ψ of the internal junctions β -(1 \rightarrow 3) and β -(1 \rightarrow 4) of the molecules. POLYS slightly modified the input values in order to obtain the nearest integral n-fold helical structure. In both the cases two helical structures were constructed using the (Φ , Ψ) optimized values for a three-fold and a four-fold left-handed helices.