Carbohydrate Polymers 277 (2022) 118885

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Review

Liquid-state NMR spectroscopy for complex carbohydrate structural analysis: A hitchhiker's guide

Immacolata Speciale [a], Anna Notaro [a], Pilar Garcia-Vello [b], Flaviana Di Lorenzo [a], Samantha Armiento [b], Antonio Molinaro [b], Roberta Marchetti [b], Alba Silipo [b], Cristina De Castro [a,*]

[a] Department of Agricultural Sciences, University of Naples, 80055 Portici, Italy
[b] Department of Chemical Sciences, University of Naples, 80126 Naples, Italy

ARTICLE INFO

Keywords:
Carbohydrates
Glycans
NMR
Spectra interpretation
Spectra processing
Chemical shifts analysis

ABSTRACT

Structural determination of carbohydrates is mostly performed by liquid-state NMR, and it is a demanding task because the NMR signals of these biomolecules explore a rather narrow range of chemical shifts, with the result that the resonances of each monosaccharide unit heavily overlap with those of others, thus muddling their punctual identification.

However, the full attribution of the NMR chemical shifts brings great advantages: it discloses the nature of the constituents, the way they are interconnected, in some cases their absolute configuration, and it paves the way to other and more sophisticated analyses.

The purpose of this review is to provide a practical guide into this challenging subject. It will drive through the strategy used to assign the NMR data, pinpointing the core information disclosed from each NMR experiment, and suggesting useful tricks for their interpretation, along with other resources pivotal during the study of these biomolecules.

1. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful technique used to investigate both synthetic and natural compounds in solution and, especially, to obtain information at atomic and molecular level by observing the behaviour of the atomic nuclei in a magnetic field. It has the advantage of being a non-destructive technique, therefore the material can be recovered after the analysis and used for further investigation. For these reasons, NMR is one of the most used techniques to characterize molecules, including oligo- and polysaccharides, as for the purpose of this review.

Glycans are the most abundant compounds in nature and include an heterogeneous ensemble of molecules that can be only composed of carbohydrates, as the polysaccharides, or that have a oligo- or polysaccharide moiety covalently linked to other class of molecules, as lipids or proteins, as it happens for the lipopolysaccharides (LPS) of Gram-negative bacteria (Cavalier-Smith, 2006), the lipoteichoic acids of Gram-positive bacteria (Rohde, 2019), the glycoproteins in the S-layers of the Archaea (Eichler, 2013), or the proteoglycans of the extracellular matrix of all animal tissues (Theocharis et al., 2016).

The NMR study of these molecules is hampered by two main factors. The first is given by the extreme diversity of the monosaccharide constituents, often arising from subtle differences, as for the case of D-glucose and D-mannose that differ for the configuration of their second carbon atom (C-2), or for the presence of one or more deoxy-positions, as for rhamnose and abequose, or for the replacement of one or more

Abbreviations: CSDB, Carbohydrate Structure Database; DEPT, Distortionless Enhancement by Polarization Transfer; DQF-COSY, Double-Quantum Filtered COSY; EM, exponential multiplication; FID, free induction decay; HMQC, Heteronuclear Multiple Quantum Correlation; GM, Gaussian multiplication; HSQC, Heteronuclear Single Quantum Coherence; LB, line broadening; LP, linear prediction; NOESY, Nuclear Overhauser Effect SpectroscopY; ROESY, Rotating-frame Overhauser Effect SpectroscopY; TOCSY, TOtal Correlation SpectroscopY; T-ROESY, transverse ROESY; TD, time domain.

* Corresponding author.
E-mail addresses: immacolata.speciale@unina.it (I. Speciale), anna.notaro@unina.it (A. Notaro), pilar.garciadelvellomoreno@unina.it (P. Garcia-Vello), flaviana.dilorenzo@unina.it (F. Di Lorenzo), samantha.armiento@unina.it (S. Armiento), molinaro@unina.it (A. Molinaro), roberta.marchetti@unina.it (R. Marchetti), alba.silipo@unina.it (A. Silipo), decastro@unina.it (C. De Castro).

https://doi.org/10.1016/j.carbpol.2021.118885
Received 29 August 2021; Received in revised form 23 October 2021; Accepted 9 November 2021
Available online 13 November 2021

hydroxyl group with an amino function, or for the oxidation of one of the primary carbons of the units as occurs for glucuronic and neuraminic acid, that have a carboxylic function at C-6 and C-1, respectively (Di Lorenzo et al., 2021). The second bottleneck is given by the fact that the spread of the proton chemical shifts occurs in a rather narrow range of values, with the results that the proton resonances of each monosaccharide unit of the glycan heavily coincide with those of other units, thus challenging their punctual identification. This problem is solved by analysing the carbon chemical shifts of the sample, assuming that enough material is available for the study, due to the low abundance of this nucleus along with its minor instrumental sensitivity when compared to that of the proton.

Therefore, this review will not enter into more sophisticated aspects of the NMR, as its use to detect the interaction with other molecules (Di Carluccio et al., 2021; Gimeno et al., 2020), or to establish their conformation (Widmalm, 2021) or to latest development regarding the acquisition and/or the processing of the spectra (Kupče & Claridge, 2018; Pedersen et al., 2021).

The scope of this review is to recap the strategies that best solve the bottlenecks associated with carbohydrate NMR analysis and that lead to the determination of the structure of any glycan, which means to address the following features: the nature of each unit, that is their stereochemistry, branching pattern, anomeric configuration, and sequence in the chain. In doing so, we will focus on the experiments that are most used for the purpose, and additionally we will discuss some introductory concepts about the derived NMR spectra, including the artifacts that might be contained therein, and give some indication about how to properly present the data.

2. General information

The tetrasaccharide **1** (Fig. 1a,b) used as tutorial was available from previous studies, and all NMR experiments were performed as reported (Speciale, Laugieri, et al., 2020). Briefly, the full set of NMR spectra were recorded in D_2O (sample concentration 2 mg/ml) at 310 K on a Bruker DRX-600 MHz (^{1}H: 600 MHz, and ^{13}C: 150 MHz) instrument equipped with a cryoprobe, and chemical shifts are referred to internal acetone (^{1}H 2.225 and ^{13}C 31.45 ppm). ^{1}H—^{1}H homonuclear experiments (COSY, DQF-COSY, TOCSY, T-ROESY) were recorded using 512 free induction decays (FIDs) of 2048 complex data points, setting 24 scans per FID for all experiments, a mixing time of 100 ms was applied for TOCSY and HSQC-TOCSY, and 300 ms for T-ROESY spectra acquisitions.

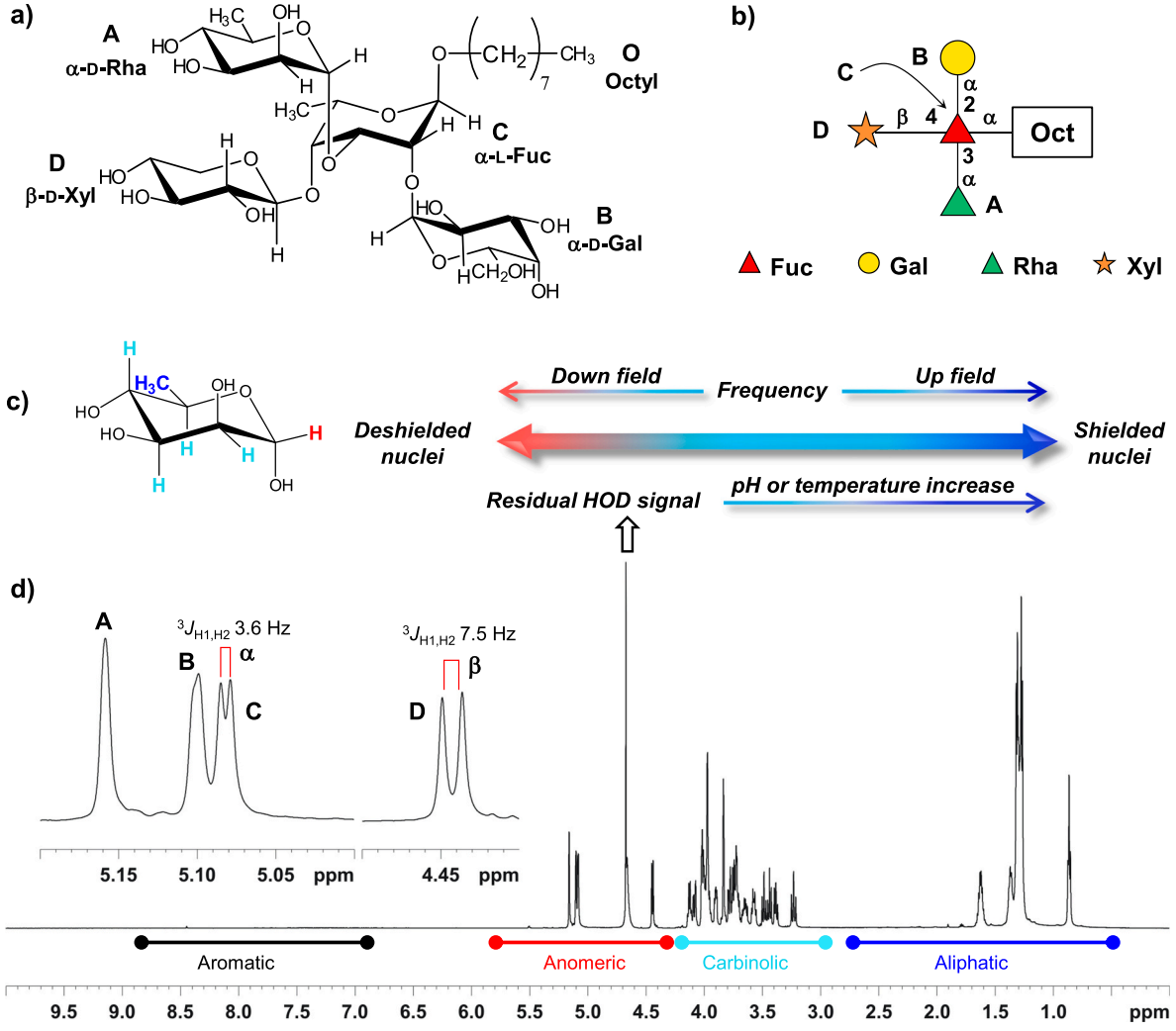

Fig. 1. a) structure of the tetrasaccharide **1** along with the indication of the labels used during NMR attribution. Notably, the carbinolic proton signals not discussed along the text have been omitted to reduce the crowding of the image. b) **1** drawn according to the formalism of the Glycan Symbolic Nomenclature. c) cartoon indicating the down (or up) field regions of the spectrum relevant for monosaccharide residues. D-Rha is given as example and the colour of its protons follows their classification and location in the different regions of the spectrum: anomeric, carbinolic and aliphatic (see also d). Moreover, the cartoon indicates in which direction the residual water signal moves upon the increase of pH or temperature. d) (600 MHz, 310 K, D_2O) ^{1}H NMR spectrum of **1:** the principal regions of the spectrum are commented by using the same colour code used in panel c, and the anomeric region is enlarged to show the different shapes of α and β anomeric signals.

^{1}H—^{13}C heteronuclear experiments (HSQC, HMBC and HSQC-TOCSY) were acquired with 512 FIDs of 2048 complex points with 40 (HSQC) or 80 (HMBC and HSQC-TOCSY) scans per FID. The following sequences from the Bruker library were used: presaturated proton, zgpr; DQF-COSY, cosydfphpr; TOCSY, mlevphpr; T-ROESY, troesyphpr; HSQC, hsqcedetgpsp; HMBC, hmbcgplpndqf; and HSQC-TOCSY, hsqcetgpml.

Regarding the sucrose NMR spectra, the original FIDs are from the entry BMSE00119 of the Biological Magnetic Resonance Data Bank (Ulrich et al., 2007), a public repository that includes the NMR raw data of several molecules. The sucrose spectra were measured at 500 MHz and calibrated on DSS.

All spectra have been processed with Topspin 3.6.1 software, freely available from Bruker for academic users. The same software has been used to prepare the NMR figures presented.

3. Sample preparation

As for the purpose of this review, this section will provide few hints regarding the preparation of the sample that in some cases are the key to solve the structure of the polymer.

Liquid-state NMR technique is based on the application of a magnetic field to a sample containing the molecules dissolved in a suitable deuterated solvent, used to lock the field so that it is stabile throughout the duration of the experiments. Among the several deuterated solvents, deuterium oxide (D_2O) is the one widely used for carbohydrate analysis, while some others (as d_6-DMSO) rarely occur. Regrettably, the major drawback of D_2O is its so-called residual signal (HOD). This signal reflects the presence of the ^{1}H proton isotope that is always present, although in a small percentage, in the deuterated solvent. The residual water signal is very intense, and it may cover those nearby with the results that some information can be lost or overlooked. It occurs at about 4.7 ppm at 300 K, namely in the same region of the anomeric signals of the carbohydrates that are among the most diagnostic signals of these molecules.

Then, measuring the NMR spectra at various temperatures is a good practise to mitigate this problem: by increasing the experiment temperature of 10 K, the residual HOD signal shifts upfield of about 0.1 ppm (Fig. 1c), on the contrary, the decreasing of the temperature moves the signal downfield (Gottlieb et al., 1997). Notably, the variation of the chemical shift of HOD with temperature is more pronounced than that observed for the proton signals of the glycans, which are poorly affected if not at all.

A similar effect can be obtained by changing the pH: compared to the neutral solution, at alkaline pH the residual solvent signal moves upfield (Fig. 1c), while the contrary happens at acid pH; clearly, the entity of the shift depends on the final pH reached. Importantly, any pH variation may impact on the chemical shifts of the sample, as result of the change of the ionization status of some groups as the carboxylic, amino and phosphate groups. Moreover, a change in the pH can modulate the resolution of the sample by improving for instance the solubility of sample. More importantly, any of the changes mentioned above, alone or in combination, will produce a spectrum with a different profile, and its full assignment will likely require the acquisition and interpretation of a new set of spectra.

4. Power and limits of 1D ^{1}H NMR analysis

The one-dimensional (1D NMR) spectroscopy is the first step undertaken for the structural characterization of any glycan, and the most relevant and technically accessible monodimensional spectrum is the (^{1}H) proton spectrum. Other nuclei (as ^{13}C, ^{31}P or ^{15}N) are also worth of direct investigation, even though the recording of their spectra has become less frequent for several reasons. First, the widespread use of reverse probe makes the measurement of some of them a challenge. Indeed, these probes are optimized to detect the proton, therefore their performance on other nuclei, as ^{13}C and ^{15}N, is rather poor. Second, some nuclei (as ^{31}P) do not occur often in glycans, therefore their measurement is not performed as routine.

With regard to the proton spectrum, it furnishes key information regarding the nature of the sample, as illustrated for the tetrasaccharide **1** (Fig. 1a,b). In general, the proton spectrum can be divided into four main regions (Fig. 1d): the first at high (or up) fields (2.7–1.0 ppm) or the aliphatic region, is relevant for the detection of deoxysugars. The methyl group of 6-deoxyresidues, as rhamnose or fucose, can be found at about 1.1–1.3 ppm, while the remaining part of this range reports the protons of deoxy position other than carbon 6 (C-6), as for instance the two diastereotopic protons of the methylene group at C-3 of Kdo (3-deoxy-2-keto-D-*manno*-octulosonic acid), the hallmark of bacterial lipopolysaccharides (Marchetti et al., 2021). The next region (4.4–3.0 ppm) appears as the most crowded part of the spectrum because it reports the carbinolic protons of the monosaccharide residues, and for this reason its assignment necessitates the combined study of a discrete set of 2D NMR spectra. Thus, the information contained in this part of the spectrum are of no immediate reading, and this region is generally considered as the "fingerprint" of a specific glycan.

Then, the range at 5.6–4.4 ppm is considered as the most informative of the whole spectrum. It is referred to as the "anomeric region" since it reports the anomeric protons of any aldose unit, even though it may contain also other types on non-anomeric signals, as discussed below (Section 7). Finally, the remaining down field part the proton spectrum is less relevant for carbohydrate analysis; in particular, the 8–7 ppm range is diagnostic of aromatic signals (Fig. 1d) and it may have some importance for glycosides with an aromatic aglycon or for sugar-nucleotides in biosynthetic studies.

Importantly, this division has not to be strictly considered because exceptions may occur owing to the influence of specific substituents (such as phosphate, sulfate, acetyl groups), that may shift the geminal proton - generally a ring proton - up to the anomeric region of the spectrum (Section 7).

Therefore, applying these guidelines to **1** (Fig. 1), it emerges that the sample does not contain an aromatic component, and that it is composed by four aldose units because of the presence of four anomeric signals (inset in Fig. 1d) in the corresponding region, hereafter labelled with a capital bold letter in decreasing order of chemical shift. Generally, for sugars in the pyranose form, the anomeric protons above 4.7 ppm indicate the α configuration of the anomeric centre, otherwise they are β configured, even though some exception may occur as often happens with *manno* configured residues. Importantly, the signal of each anomeric proton is split due to the coupling with the neighboring proton (i.e. H-2), and the entity of this splitting (the "coupling constant", denoted as $^{3}J_{H1,H2}$ and expressed in Hz) depends on the relative orientation between the two protons, being thus indicative of the anomeric configuration of the sugar. Accordingly, in sugars with the axial orientation of H-2, as galactose or xylose (Fig. 1) or glucose (Fig. 2a), β anomers have the *trans*-diaxial orientation of the H-1/H-2 protons so that they appear as doublets with a coupling constant value of ~7–9 Hz. On the contrary, when these same monosaccharides have the α configuration (as **B** and **C,** Fig. 1), their anomeric protons form a dihedral angle of ~60° with H-2 (equatorial–axial arrangement), which translates into a small $^{3}J_{H1,H2}$ coupling value of ~2–4 Hz. In contrast with the previous cases, the anomeric configuration of monosaccharides with H-2 in equatorial position, as mannose and rhamnose, is harder to define. For these residues, the H-1/H-2 dihedral is close to ~60° in both β and α anomers so that the measured $^{3}J_{H1,H2}$ values very similar and both close to 1 Hz (α configuration above 1 Hz and β below 1 Hz), so that the anomeric configuration of these units is usually inferred by comparing their ^{13}C chemical shift to that of the methylglycosides taken as reference (Section 8.4), and/or by measuring their $^{1}J_{C1,H1}$ values (Sections 5.5 and 5.6).

Regarding monosaccharides in the furanose form, the H-1 chemical shift is less indicative of the stereochemistry of the anomeric centre since both α and β anomers are generally found above 5 ppm. Likewise, the

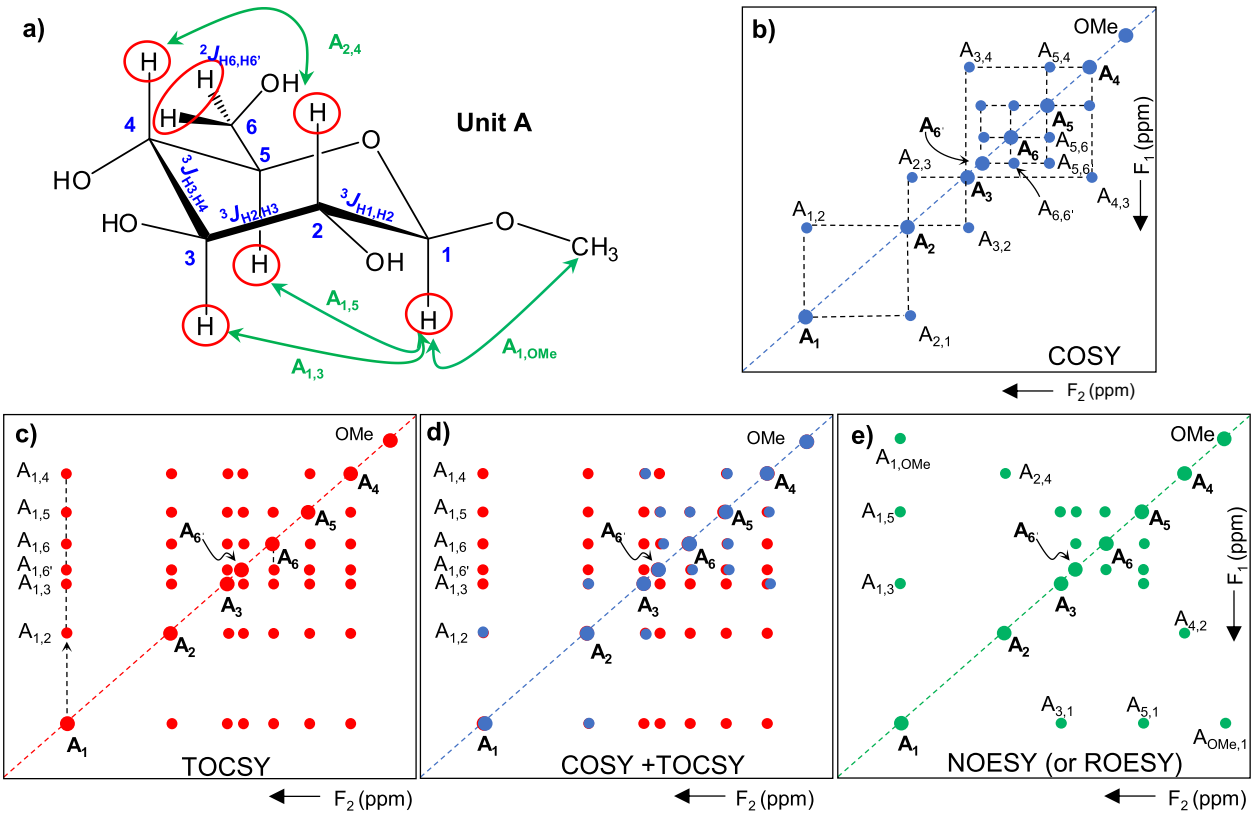

Fig. 2. a) Methyl-β-D-glucopyranoside unit (labelled **A**) with indication of the information gathered by homonuclear $^{1}H—^{1}H$ NMR experiments (COSY, TOCSY, and NOESY). Some of the ^{3}J couplings are indicated along the linkage that joins the two coupled protons, while the subscript indicates the identity of the two protons; the red circle around the protons indicates that they should appear all interconnected in the TOCSY spectrum; the green arrows point to the main NOE effects expected. b–e) Cartoons representing different types of homonuclear spectra. b) Cartoon representing a COSY spectrum: the cross peaks interconnect nuclei related from ^{2}J or ^{3}J coupling constants. The diagonal peaks should be labelled $\mathbf{A}_{1,1}$, $\mathbf{A}_{2,2}$ etc., however they are denoted with $\mathbf{A}_{1}$ and so forth for simplicity. c) Cartoon of the TOCSY spectrum that for this unit is expected to relate all the protons (red circled in a) due to the efficient propagation of the magnetization. d) Overlap of the COSY and TOCSY spectra, an information rich way to visualize the two spectra. e) Cartoon of the NOESY spectrum with indication of the main densities expected.

$^{3}J_{H1,H2}$ values depend on the stereochemistry of the unit, so that the identification of the sugar in the furanose form is trickier than the pyranose counterparts, and it is generally afforded by comparing the experimental data with those from the literature. This comparison is generally done by matching the carbon chemical shifts of each residue of the glycan with those reported for the corresponding methylglycoside taken as reference (Bock & Pedersen, 1983) (Tables S1 and S2), paying attention to the occurrence of substituents because these influence the carbon chemical shift value observed (Section 7).

Finally, the analysis of the region at 2.7–0.9 ppm evinces the presence of several aliphatic protons, methyl groups (at ~1.3 ppm) of deoxy sugars, along with protons (of aliphatic nature or arising from sugar deoxy at positions other than C-6.

Clearly, the proton spectrum is not sufficient to define the oligosaccharide structure, which instead requires an extensive use of homo- and heteronuclear 2D spectra.

5. 2D-NMR spectra

The bidimensional (2D) NMR spectra provide different sets of information depending on the physical phenomena examined: scalar coupling (through-chemical bond) or dipolar (through-space) interactions between the spins of two (or more) nuclei, and diffusion-based mobility of the molecule in solution.

The scalar coupling generally occurs between two nuclei that are separated by one, two (geminal) or three (vicinal) bonds, and the entity (or the intensity) of the coupling depends on the relative orientation between the nuclei, being null or almost null in some cases.

The dipolar interactions, or NOE effects, occur between two or more nuclei that are close in space, and for this reason they are used to deduce information on the conformation of the molecule. As for inter-proton NOE effects, these can be observed for protons generally at less than 4 Å (Claridge, 2016c). Regarding the diffusion-based NMR spectra, these shed light on the physical property of the molecule in solution and do not disclose its fine chemical structure, and for this reason they will not be discussed in this review.

Then, either scalar coupling or dipolar interactions based 2D-NMR experiments relate two nuclei and the spectra present two frequency axes, F2 (x-axis) and F1 (y-axis), along with a third one, the intensity, always omitted because the spectra are displayed as contour plots (Fig. 2).

Homonuclear experiments relate the same nuclei, and they always display a diagonal (same F1 and F2 values) whose chemical shifts match those of the monodimensional spectrum of the compound; these densities are referred as diagonal peaks. On the contrary, all the densities outside the diagonal are named cross peaks and are those that contain the searched information.

Homonuclear (COSY, TOCSY, NOESY or ROESY or its variation T-ROESY) and heteronuclear (HSQC, HSQC-TOCSY and HMBC) experiments disclose different information, that all together allow the structural determination of the examined molecule.

Notably, a full set of 2D NMR spectra is generally necessary to establish the structure of a glycan. In some cases, the information from two different spectra – as NOESY and HMBC – can appear redundant: on the contrary, these two spectra countercheck each other, strengthening the overall interpretation.

In the following subsections, the information gathered from each of these experiments is examined, together with some warnings about the artifacts that might occur.

5.1. ^{1}H—^{1}H COSY and DQF-COSY

The homonuclear correlation experiment COSY is generally the first spectrum acquired on a sample. There exist several variants of this sequence (see (Claridge, 2016b), for a thorough presentation), and this section will focus mostly on COSY-90, or simply COSY, to introduce the formalisms used throughout this review along with other general considerations.

The basic principle of any COSY spectrum is that it relates protons that are scalarly (or through-bonds) coupled, with a coupling constant value (J) different from zero. Generally, these protons are either geminal or vicinal, namely separated from two (^{2}J) or three (^{3}J) chemical bonds (Fig. 2a), respectively, while protons separated by four linkages (or more) generally are not coupled, except when some specific geometric conditions are met, which seldom occurs in carbohydrates.

Like other homonuclear spectra, the COSY spectrum is symmetrical with respect to the diagonal, therefore its densities are divided into two different sets: those building the diagonal and those off the diagonal. Regarding the first type, these densities correspond to the trace of the proton spectrum, therefore they do not add any information. On the contrary, the symmetrical off-diagonal peaks (cross-peaks) are of relevance as they represent protons that mate with each other because coupled. Therefore, starting from the anomeric proton of unit **A** (H-1 of **A** or $\mathbf{A_1}$ for brevity, Fig. 2b) read on the F2 (or the x) axis, a straight line

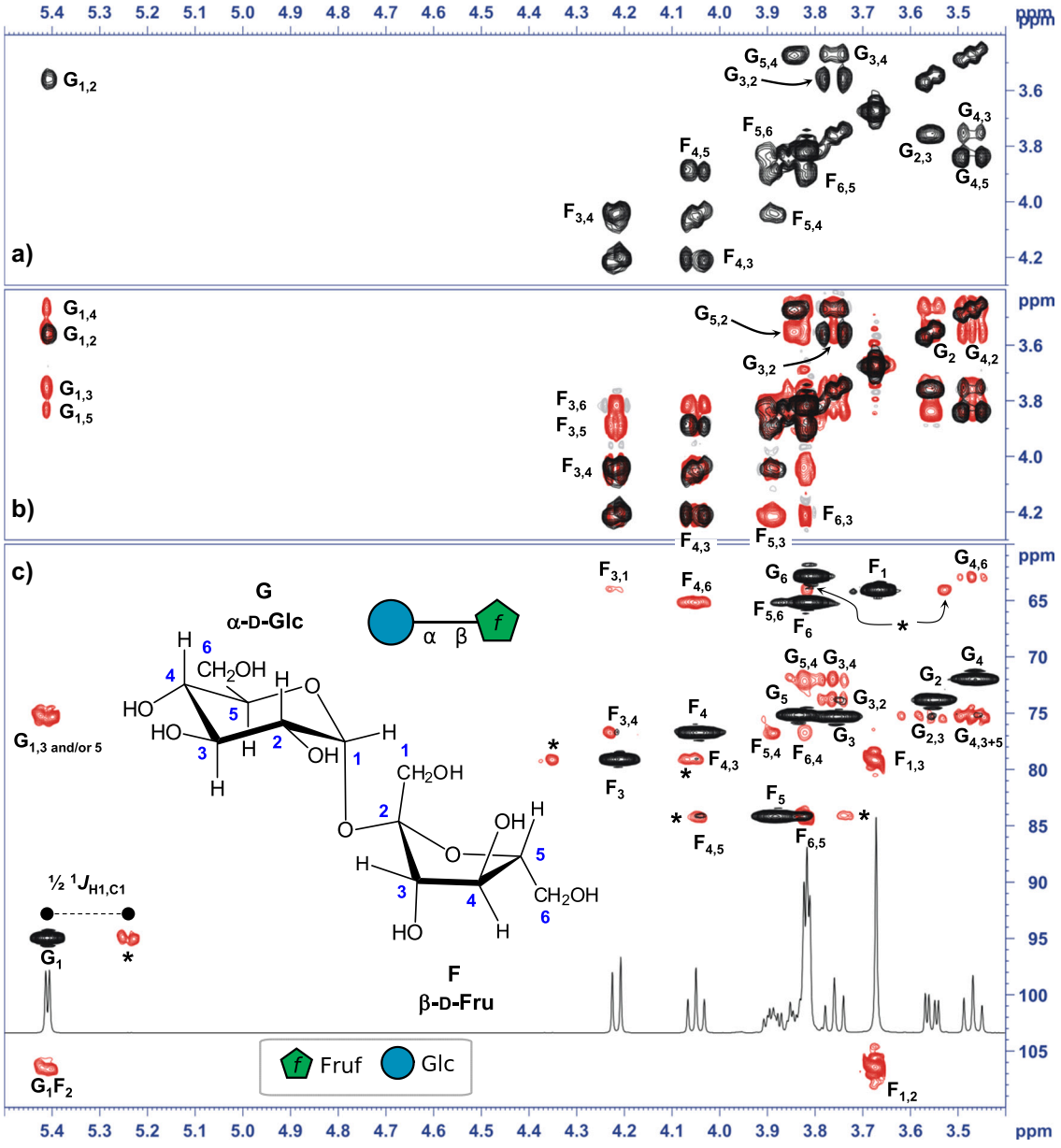

Fig. 3. Spectra measured for sucrose (BMSE00119 entry from the BMRDB database) at 500 MHz, 298 K and referenced versus TSP. The structure of sucrose is reported in panel c) along with the labels used. The densities are labelled with capital letters that refer to the sugar unit (**G** stands for glucose, **F** for fructose) and numbers that indicate the position (hydrogen or carbon) of the unit. a) Expansion of the COSY reporting the area more relevant for the assignments. b) Overlay of TOCSY (red) and COSY (black) with only some of the densities labelled to avoid crowding. c) overlay of the HSQC (black) and HMBC (red); the HMBC artifacts are indicated with an asterisk and arise from the inefficient removal of the direct proton/carbon correlation, and they can be used to measure the $^{1}J_{H1,C1}$ value. The HSQC spectrum can present the COSY-artifacts, like $\mathbf{F_{5,6}}$, $\mathbf{F_{6,5}}$ or $\mathbf{F_{4,5}}$, which in some case can overlap with the expected correlations in the HMBC.

parallel to the F1 dimension (or the y-axis) intersects the off-diagonal cross-peak that relates $\mathbf{A_1}$ to the next proton of the residue, namely $\mathbf{A_2}$. This cross-peak is labelled $\mathbf{A_{1,2}}$ (and not $\mathbf{A_{2,1}}$) because the order "1,2" reflects the x,y-coordinates of the density (Fig. 2b), and it denotes that H-1 and H-2 are scalarly coupled with a certain value of the $^3J_{H1,H2}$ coupling constant. Then, the chemical shift of $\mathbf{A_2}$ is the y-value (the F1 dimension) of the cross-peak, and it will cross the diagonal of the spectrum at the position where this proton lies in the 1D proton spectrum. On the contrary, the $\mathbf{A_{2,1}}$ cross-peak is found when a straight line parallel to the F2 axis is drawn starting from $\mathbf{A_1}$. The two cross-peaks, $\mathbf{A_{1,2}}$ and $\mathbf{A_{2,1}}$, are symmetrical with respect to the diagonal, and any difference in their shape is due to the resolution used during the acquisition of the spectrum, that is never the same for the two dimensions (see cross-peaks $\mathbf{F_{3,4}}$ and $\mathbf{F_{4,3}}$, or $\mathbf{G_{5,4}}$ and $\mathbf{G_{4,5}}$ in Fig. 3a).

The process used to identify the chemical shift (or the position) of $\mathbf{A_2}$, can be then reiterated so that starting from $\mathbf{A_2}$ it is possible to find all the others, thus completing the identification of all the – often so called - ring protons, including also the two exocyclic protons linked at C-6 or the hydroxymethylene group, H-6 and H-6′ (or $\mathbf{A_6}$ and $\mathbf{A_{6'}}$).

In a real case, the COSY spectrum of sucrose (expansion in Fig. 3a) enables the detection of the ring protons of the two units of the disaccharide. Starting from the anomeric signal of the glucose unit (labelled $\mathbf{G_1}$) at 5.4 ppm, the cross peak $\mathbf{G_{1,2}}$ defines the position of H-2 (or $\mathbf{G_2}$) from which $\mathbf{G_3}$, $\mathbf{G_4}$ and $\mathbf{G_5}$ are found. However, the identification of $\mathbf{G_6}$

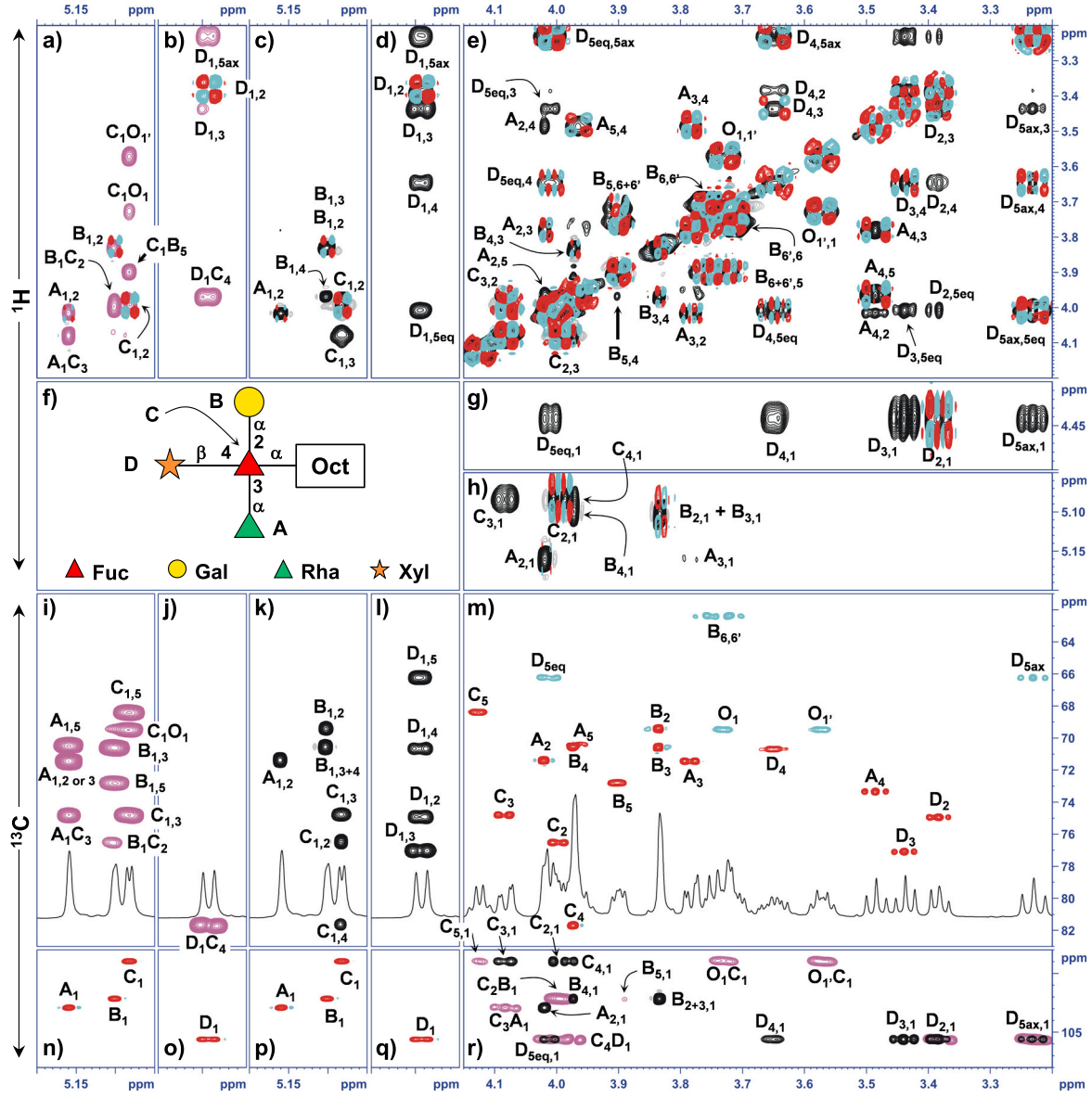

Fig. 4. Selection of NMR spectra of the tetrasaccharide **1** (panel f). Panels a–e, g, h report selected regions of ^{1}H—^{1}H homonuclear spectra, while panels i–r those of ^{1}H—^{13}C heteronuclear spectra along with the trace of the proton spectrum. In detail: a) overlay of T-ROESY (pink) and DQF-COSY (cyan/red, hereafter named only COSY) spectra detailing the anomeric region (along F1) of the **A**, **B**, **C** residues. b) overlay of T-ROESY (pink) and COSY (cyan/red) spectra detailing the anomeric region (along F1) of the **D** residue. c,d) same regions as in panels a,b), respectively, except that TOCSY (black) instead of T-ROESY is reported along with the COSY (cyan/red) spectrum. e) overlay of TOCSY (black) and COSY (cyan/red) spectra detailing the carbinolic region. f) representation of **1** according to the symbolic nomenclature of glycans with indication of the labels used for each unit. g,h) overlay of TOCSY (black) and COSY (cyan/red) detailing the anomeric region (along F2) of **D**, and **A–C**, respectively. i) enlargement of HMBC (pink) spectrum of the anomeric region (along F1) of the **A**, **B**, **C** residues. j) same as in i) except that the **D** is detailed. k,l) enlargement of HSQC-TOCSY (black) spectrum of the anomeric region (along F1) of the **A**, **B**, **C** residues (in k) and **D** (in l). m) HSQC expansion detailing the carbinolic region. n,p and o,q) HSQC regions detailing the anomeric densities of **A**, **B**, and **C**, and **D**, respectively. r) enlargement of HSQC-TOCSY (black) and HMBC (pink) spectra detailing the anomeric region (along F2). Spectra are modified from (Speciale, Laugieri, et al., 2020).

and $\mathbf{G_{6'}}$ is not straightforward, because the cross-peak connecting H-5 to any of the H-6s, is located very close to the diagonal where it blurs with it and with the cross-peaks that belong to the fructose unit, the other residue of the disaccharide.

The difficulty noted above represents the major bottleneck associated to the NMR study of carbohydrates, namely the occurrence of the chemical shifts in a narrow range with the high probability that the signals of one residue overlap over each other (as $\mathbf{G_5}$ and $\mathbf{G_6}$, Fig. 3) or with those arising from other units (as $\mathbf{G_5}$ and $\mathbf{F_6}$, Fig. 3).

Regarding the fructose unit, it must be noted that this is a ketose therefore it lacks the anomeric proton commonly used to start the NMR attribution. In this case the attribution starts from H-3 ($\mathbf{F_3}$), and the follow-up of the cross-peaks leads to the identification of $\mathbf{F_4}$, $\mathbf{F_5}$, and $\mathbf{F_6}$.

Beside the problems noted above, the COSY spectrum has some additional limitations: i) it does not discriminate vicinal from geminal protons; ii) if two signals overlap, the stepwise assignment is complicated, and it could easily lead to mistakes; iii) the diagonal peaks could hide important nearby cross-peaks causing the loss of important information.

This latter problem is partially overcome by using the double quantum filtered COSY (DQF-COSY, an example is given in Fig. 4e). In this experiment, a double quantum filter is applied during the selection of the magnetization with the result that only signals with *J*-couplings are detected, while those with no coupling – as the two H-1 of fructose – are filtered out. The use of this sequence facilitates the study of the spectrum because the diagonal is less crowded and the cross-peaks next to it are visualized better (note the $\mathbf{B_{6,6'}}$ and $\mathbf{B_{6',6}}$ cross peaks at ca. 3.75 ppm, Fig. 4e). In addition, this type of spectrum provides quali/quantitative information about the coupling constant value between two protons, as discussed with a practical example in Section 8. Another point of attention is that this experiment is less sensitive compared to its simplest version, therefore it requires a major number of scans (about a fourfold) to reach the same signal-to-noise ratio.

Hence, the DQF-COSY spectrum can mitigate to a certain extent the problem of signals overlap, even though it will hardly lead to the structural elucidation of the sample, which instead is afforded by a combination of different NMR experiments.

5.2. TOCSY

Differently from the COSY, the TOCSY spectrum detects the correlations between protons that are in a chain of spin-spin (*J* or scalarly) coupled protons and that become inter-related through a process called magnetization propagation, which is realized with the spin-lock sequence during the acquisition of the spectrum.

Accordingly, the TOCSY unveils al the nuclei that are within the same spin-system, independently from the fact that they are directly coupled to each other via a ^2J or ^3J coupling constant, or not. Taking the unit **A** as example (Fig. 2c), its spin system includes all the protons of the sugar ring (from $\mathbf{A_1}$ to both $\mathbf{A_6}$), while the spin-spin couplings occur between $\mathbf{A_1}$ and $\mathbf{A_2}$, $\mathbf{A_2}$ and $\mathbf{A_3}$, and so forth. Focusing on $\mathbf{A_1}$, the COSY spectrum will display the $\mathbf{A_{1,2}}$ (or $\mathbf{A_{2,1}}$) cross peak only (Fig. 2b), while the TOCSY spectrum will present also the $\mathbf{A_{1,3}}$ (and $\mathbf{A_{3,1}}$) correlation because of the magnetization transfer between $\mathbf{A_1}$ and $\mathbf{A_3}$ has been mediated by $\mathbf{A_2}$, since it is coupled to both. Thus, a properly set TOCSY experiment is expected to correlate $\mathbf{A_1}$ to all the other protons of the sugar ring, including both $\mathbf{A_6}$ (Fig. 2c), assuming that all the proton-proton ^3J (or ^2J) values are different from zero.

The use of TOCSY is advantageous in solving the crowded regions of the spectra. Starting from the anomeric signal (or from any other devoid of overlaps with other signals), the TOCSY trace shows all the protons that are in the same monosaccharide spin system, and this information drives the selection of the correct proton that is correlated to another in the COSY spectrum. In the common practise, the TOCSY is studied together with the COSY spectrum (examples in Figs. 2d, 3b) to maximize the information that can be achieved at glance just by looking just at one proton of the spin system, as $\mathbf{A_1}$: the chemical shifts of all the protons interconnected to it, along with the one ($\mathbf{A_{1,2}}$) that is vicinal (or geminal).

An additional advantage is that the TOCSY enables a preliminary identification of the relative stereochemistry of the residue, namely if the monosaccharide has a *manno*, or a *gluco*, or a *galacto* (or another) relative configuration, even though it cannot provide information about their absolute configuration, D or L.

Taking glucose as example (Fig. 2c), the TOCSY trace from $\mathbf{A_1}$ is expected to give six different correlations (either in the F2 or F1 dimensions), one for each proton of the monosaccharide due to the favourable proton-proton coupling constant values that exists between all protons. However, when an epimer of glucose is studied, the TOCSY pattern changes due to the presence of a coupling constant value of little entity that leads to an interruption of the magnetization propagation at the level of the different stereocentres. Then, if the residue is an epimer in position 2 of glucose, like mannose or rhamnose, the TOCSY pattern from the anomeric signal displays one intense correlation with H-2, while all the others do not appear or have an extremely low intensity (unit **A** in Fig. 4h). Similarly, if the residue is epimer at position 4, as galactose or fucose (units **B** and **C** in Fig. 4h, respectively), the TOCSY pattern from the anomeric signal generally stops at proton H-4.

Notably, the general consideration given above does not take into account that some proton signals of the unit might have at the same chemical shift. In case this happens, the number of correlations expected decreases. A pertinent example is the glucose unit of sucrose. By analysing the row passing through the $\mathbf{G_2}$ diagonal peak, it is possible to count only four cross-peaks and not six as expected (Fig. 3b), because $\mathbf{G_5}$ and the next two H-6 protons are almost coincident. The same occurs for the galactose unit of the tetrasaccharide, whose $\mathbf{B_2}$ and $\mathbf{B_3}$ are coincident leading to the observation of only two cross-peak densities in the TOCSY spectrum, instead of the three expected (Fig. 4h).

Finally, the TOCSY spectrum may present some artifacts that can be easily recognized since they have the sign opposite to that of the true TOCSY correlations. These artifacts are named ROESY-artifacts and depend on the fact that the spin-lock sequence is the employed by both sequences, with just minimal variations in the settings, so that the TOCSY spectrum may contain some of the effects expected in the ROESY, and vice versa (Section 5.4).

In summary, the advantages of the TOCSY rely on its ability to drive the selection of the correct cross-peak(s) during the study of the COSY, along with giving some preliminary information about the stereochemistry of the monosaccharide investigated.

5.3. NOESY and ROESY

NOESY is again a homonuclear experiment, but conversely from COSY and TOCSY, this experiment relies on the nuclear Overhauser effect (NOE), that said in simple terms, detects the phenomena of cross-relaxation that occurs between two nuclei. Then, the cross peaks report the dipolar (or through-space) interactions between spins, and they do not depend on the number of bonds that separate the protons, but only on their distance. The closer the nuclei are, the greater the signal intensity is, with 4 Å being the distance limit for this effect to be detected.

Then, the NOESY can relate protons within the same residue (*intra*-residue NOE effects), as the H-1/H-3 and the H-1/H-5 correlations (Fig. 2e) typical of the residues β configured at the anomeric centre (Fig. 2e) or belonging to different residues (*inter*-residue effects) just because close in space, as the methyl group in the example (Fig. 2e).

However, it is worth to note that care needs to be taken during the interpretation of the NOESY (Reynolds & Enríquez, 2002). The first problem is that this experiment may report the so-called COSY-artifacts, that are cross-peaks relating two protons with a strong scalar coupling, as it occurs for H-1/H-2 proton of a β-glucose. These artifacts are easily sorted out because the corresponding cross-peaks have an anti-phase multiplet aspect, namely they are composed of both positive and

negative densities.

Second and more important, the size of the NOE effects depends on the molecular tumbling in solution and on the field strength of the instrument. Accordingly, NOE effects are positive for fast tumbling molecules as small oligosaccharides, so that the phase (or the sign) of the cross-peaks is opposite to that of the diagonal of the spectrum. Conversely, NOE effects are negative for slow tumbling molecules as polysaccharides, and their densities are negative or in-phase with the diagonal. Hence, oligosaccharides of intermediate size – about 4–6 sugar units – roughly trace the line between positive/negative NOE effects with the result that their NOEs are very close to zero if not just zero (named zero-crossing point), even though the molecule contains several proton pairs at the right distance to give an effect. This problem can be circumvented in two different ways: by measuring the NOESY at a different field strength, or by resorting to a different pulse sequence, the ROESY.

The ROESY provides the same information of the NOESY, and the corresponding dipolar coupling is generally referred as NOE's effects even though they should be more properly named ROEs. This spectrum has the advantage that the sign of the effects does not depend on the size of the molecule: the ROE cross-peaks are positive (or never in phase with the diagonal) and there is not the risk of zero-crossing.

However, artifacts may plague the ROESY spectrum, as well. The major source of them derives from the use of the spin-lock pulse sequence, similar to that used in the TOCSY spectrum (Section 5.3), and for such reason these artifacts take the name of TOCSY-artifacts. Contrary to the ROESY densities, the TOCSY-artifacts are negative (or in phase with the diagonal) so that they can be easily found in the spectrum. However, in case they coincide with a true ROE effect, they will cancel each other, and no cross-peak will appear in the spectrum. Some additional artifacts occur due to the mixing of both TOCSY and ROE effects (Claridge, 2016c), known as transmission of the magnetization, with the result that their cross-peaks have the same sign of the true ROESY cross-peaks, even though the two protons are not close to each other.

To mitigate these issues and the potential misinterpretation of the data that follows, the transverse-ROESY or T-ROESY spectrum can be acquired as possible alternative. This sequence is able to minimize all the shortcomings arising from the spin-lock pulse, namely both the TOCSY-artifacts and those related to the transmission of the magnetization.

Last, if not properly set the NOESY and the T-ROESY spectra can display a large array of effects devoid of any physical meaning because relating protons that are not close in the molecule, as H-1 and H4 of a sugar in the pyranose form. These artifacts are due to the so-called spin-diffusion phenomenon that arises with the use of a mixing time excessively long during the acquisition of the spectra. Generally, this problem is solved by reducing the mixing time.

5.4. HSQC and HMQC

$^{1}H—^{13}C$ Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) or its multi-quantum counterpart (HMQC) are used to correlate the chemical shifts of protons (displayed on the F2 axis) to that of the carbon atom directly attached (reported on the F1 axis) utilizing the one-bond coupling $^{1}J_{CH}$, generally set to ≈145 Hz to detect both anomeric ($^{1}J_{CH}$ ≈ 160–170 Hz) and carbinolic ($^{1}J_{CH}$ ≈ 140 Hz) correlations. The two sequences provide the same information although the appearance of the cross-peaks is slightly different: the densities in the HMQC spectrum maintain the homonuclear proton couplings in F1, therefore they are less resolved compared to those of the HSQC spectrum, where instead this coupling is removed as effect of the sequence used with a high gain in resolution. Through this text, we will refer to the HSQC spectrum, although the same considerations apply to the other.

The carbon chemical shifts of carbohydrates can be divided in different regions (Fig. 3c): 10–25 ppm is the aliphatic region and it is diagnostic of the methyl groups of 6-deoxysugars and of the acetyl groups; 50–58 ppm is typical of carbon bearing an amino group; 60–70 ppm is diagnostic of the hydroxymethylene (−$\underline{C}H_2OH$) carbons, either with the free hydroxyl function (60–63 ppm) or substituted (64–70 ppm); 70–85 ppm is diagnostic of the carbinolic (−$\underline{C}HOH-$) carbons; 90–110 ppm is the region of anomeric carbons. Here, an additional classification can be made depending on the α/β configuration of the anomeric carbon and on the status of the sugar, namely if it is in the free reducing form or involved in a glycosidic linkage, and if it is in the pyranose or the furanose form.

Considering the residues in the pyranose ring and with the free reducing end, the carbon densities are found at 90–98 ppm, with those of the α anomers hardly above 95 ppm, whereas they are at lower fields when the residue is in β glycosidic linkage (Tables S1, 2).

In general, in glycans the ^{13}C values of the α anomers are at about 98–103 ppm, while those of the β anomers are at 103–106 ppm (Agrawal, 1992), these ranges do not depend on the absolute configuration of the residues (D or L), and any exception to this rough division depends essentially on the substitution pattern of the sugars.

Of note, the anomeric configuration of residues with the *manno* stereochemistry, as mannose and rhamnose, cannot be distinguished based on the anomeric ^{13}C values, because of the similarities between the α/β values. In such cases, such feature is inferred by comparing the C-3 or the C-5 values of the unit with those of the corresponding methylglycoside taken as reference (Tables S1 and S2, Section 8.4), or by observing the $^{1}J_{C1,H1}$ values. These coupling constants are predictive of the anomeric configuration of almost any type of aldopyranose, and measure about 170 or 160 Hz on average for α or β anomers, respectively. These values can be read from a ^{1}H-coupled HSQC or in the HMBC spectrum when the parameter given to filter out the direct (one-bond) correlation does not match the $^{1}J_{C1,H1}$ value (Section 5.6).

Regarding the furanose residues in the free reducing form, the anomeric carbon is found at 96–104 ppm, and this range increases sensibly (103–110 ppm) when the monosaccharide is engaged in a glycosidic linkage. In both cases, the distinction between the α/β anomeric configuration depends on the stereochemistry of the monosaccharide and it cannot be ascertained on the basis of the $^{1}J_{C1,H1}$ value, because the range covered (168–171 Hz) is narrow and it is not distinctive of any of the two forms. In this case, the comparison with the chemical shifts of the unsubstituted glycosides is the best solution to solve this issue.

The HSQC spectrum yields useful information not limited to the anomeric region, because the values of the ring carbon signals (Figs. 3c, 4m) are equally informative – if not more accurate – of the structure the glycan. First, the presence of furanose sugars can be inferred by the diagnostic densities at 80–85 ppm, that corresponds to C-4 of aldofuranose or C-5 of ketofuranose (Fig. 3c) units, respectively. Then, detailed information can be extracted from the ^{13}C values once that the full attribution of the unit has been carried out, through a process that consists in the comparison of the values found for the unit with those of the methylglycoside taken as reference, as discussed in Section 8. Importantly, the HSQC experiments detects all (and only) the carbon nuclei that possess at least one attached proton, for this reason, the signals of the carbonyl of any acyl group (except the formyl), or of the anomeric carbon of keto-sugars are not detected. The simplest way to rescue these data is by recording the HMBC spectrum (Section 5.6, Fig. 3c). Finally, an improvement of the HSQC spectrum consists into its multiplicity editing often referred as HSQC-DEPT sequence that has the advantage to present the “CH” and the “CH_3” densities in antiphase with the methylene carbons “CH_2” (Fig. 4m). The only drawback of this improvement is that the cancellation of overlapping correlations of opposite phase may arise when the carbinolic region is crowded.

Another artifact common to any HSQC involving the INEPT sequence for sensitivity enhancement occurs when two protons are connected with a strong $^{3}J_{H,H}$ coupling (Turner et al., 1999), with the result each of the two protons appears as correlated to two different carbons. In such

cases, the density with the strongest intensity belongs to the carbon directly linked to the proton, while the second density indicates the chemical shift of the carbon attached to the other proton, as the correlations $\mathbf{F_{6,5}}$ and $\mathbf{F_{5,6}}$ in Fig. 3c. This type of artifact is sometimes named COSY-type artifact and when identified, it can facilitate the assignment of the densities of the HSQC spectrum.

From the labelling viewpoint, this review will adopt the following formalism: given the unit **G** (as in Fig. 3) the density of its anomeric carbon will be labelled $\mathbf{G_1}$, that of carbon 2 as $\mathbf{G_2}$, and so forth. This notation is compact and of immediate reading compared to other possible, like G(H-1)/G(C-1) that may be hard to place in crowded area of the spectrum.

5.5. HMBC

The $^{1}H—^{13}C$ HMBC spectrum shows correlations between protons and carbons that are scalarly coupled, even though they are not directly linked to each other. Accordingly, this sequence detects proton/carbon pairs separated from two (^{2}J) or three (^{3}J) linkages (Fig. 3c), without the possibility to make a distinction between them because the size of the coupling constant (2–15 Hz) is the same in both cases. Nevertheless, this sequence is essential for the structure elucidation of polysaccharides since it allows to tie together different molecular fragments into a complete structure, thus counterchecking the results of the NOESY (or ROESY) experiment and to rescue information otherwise lost. For instance, the HMBC enables the assignment of carbon atoms with no protons attached, as the anomeric carbon of ketoses (Fig. 3c), or the carbonyl of the acyl groups.

From the experimental point of view, the set-up of the HMBC incorporates two filters necessary for the selection of the desired signals. The first is about 4–8 Hz, namely to a likely average of the possible $^{2\text{ or }3}J_{C,H}$ values. The second instead, is used to minimize (or to filter out) the response arising from the direct correlation namely from those proton/carbon pairs related by a one-bond coupling ($^{1}J_{C,H}$). In this case, the ^{1}J filter is generally set to 145 Hz, to remove the direct correlations that may affect the most crowded area of the spectrum, the carbinolic region.

The drawback of this choice is that it may not remove completely the magnetization arising from the anomeric carbons because their values $^{1}J_{C,H}$ (160–170 Hz) diverge significantly from the filter of 145 Hz used. As consequence, the densities of the anomeric carbons are detected in the HMBC spectrum, where they retain the coupling with their own proton, so that they appear as split in two densities along the F2 dimension, with the centre matching the position of the anomeric proton, while their distance (in Hz), is the $^{1}J_{C1,H1}$ value (Tvaroska & Taravel, 1995). Therefore, this effect is not a disadvantage because it can be used to evaluate the coupling constant value $^{1}J_{H1,C1}$ and to ascertain the configuration of the anomeric centre, as detailed for the **G** unit in Fig. 3c or as reported in Table S3 for the tetrasaccharide **1**.

With regard to the labelling convention used in this review, the densities in the HMBC spectrum can represent *intra*- as well as *inter*-residue correlations (Figs. 3c and 4i,r). Taking unit **G** as example (Fig. 3c), the anomeric proton can display a maximum of three *intra*-residue correlations, namely with C-2 ($\mathbf{G_{1,2}}$), C-3 ($\mathbf{G_{1,3}}$) and C-5 ($\mathbf{G_{1,5}}$). The logic of this formalism is to report the letter (**G**) used to identify the sugar unit followed by the position of the two nuclei (first the one in F2) as subscript. In this specific case, the $\mathbf{G_{1,2}}$ correlation is not detected probably because the $^{2}J_{H1,C2}$ diverges from the filter of 8 Hz used (or is null), while the density at $^{1}H/^{13}C$ 5.41/75.5 ppm is compatible with both C-3 and C-5 of the unit, since the chemical shifts of these two carbon atoms are very similar.

As for the *inter*-residue correlation, the anomeric proton of **G** (along the F2 dimension) is related to the anomeric carbon (C-2) of **F** (the nucleus in the F1 dimension), so that the corresponding density is labelled $\mathbf{G_1F_2}$ (Fig. 3c). Notably, this correlation leads also to disclosure of the chemical shift of the anomeric signal of the keto-sugar, that had not the requirement for being detected in the HSQC spectrum.

5.6. Hybrid HSQC experiments and HSQC-TOCSY

The potentiality of the HSQC experiment can be further expanded by adding other criteria for the selection of a certain magnetization component. This approach leads to the so-called "hybrid (or hyphenated) sequences" where the HSQC-TOCSY is the one largely exploited in carbohydrate structural analysis.

This hyphenated sequence gathers the advantages from both types of experiments, the HSQC and the homonuclear TOCSY, so that the chemical shifts of a certain unit are spread along the wide interval characteristic of the ^{13}C nucleus, thus circumventing the probability of overlaps that instead complicates the attribution in the carbinolic region of the homonuclear spectra. Accordingly, the analysis of the HSQC-TOCSY spectrum drives the selection of the correct carbon density within a pool of possible values.

The potentiality of this technique can be appreciated by looking at the HSQC-TOCSY correlations expected for the xylose unit **D** of the tetrasaccharide **1**. Xylose is a pentose and its carbinolic hydrogens are related by ^{2}J or ^{3}J coupling constants of about 10 Hz, a value that enables the efficient propagation of the magnetization during the TOCSY spinlock across all the protons of the spin system. Then, the HSQC part of the sequence transfers this relayed magnetization to the carbon atoms attached to these protons with the result that all the carbon atoms of the unit are detected. Accordingly, focusing on anomeric proton (Fig. 4l), the spectrum presents five densities (including the anomeric signal that is not shown), one for each carbon of the unit: $\mathbf{D_1}$ and those of all the other carbons of the sugar ring ($D_{1,2–5}$). The same situation is expected for *gluco*se, with the difference that the number of carbon atoms detected should be six. Other examples regarding the use of this spectrum are given in Section 8.

Quite often the spectrum displays less correlations than expected, and this can be due to the fact that some of the carbon chemical shifts are overlapping (see Section 8.3) and/or that the stereochemistry of the monosaccharide is other than *gluco* (as **A**–**C** in Fig. 4k) so that the magnetization is not transferred across all the ring protons during the spin-lock, as previously commented for the TOCSY spectrum (Section 5.2).

As additional implementation of the HSQC-TOCSY experiment, it is possible to: i) differentiate the direct (for instance $\mathbf{A_1}$) and relayed peaks ($\mathbf{A_{1,2}}$, $\mathbf{A_{1,3}}$) by their phase in the spectrum, or ii) to suppress the direct correlation (i.e. $\mathbf{A_1}$), or iii) to distinguish the densities of the hydroxymethylene signals ($-CH_2OH$) from all the others because detected with an opposite phase. As warning, all the time that certain densities are differentiated from the others through the reversal of their phase, it must be considered that some cancellation may occur in case of overlap leading to a potential loss of information.

Of note, the use of a long mixing time during the spinlock (100 ms) is the best choice to detect all the signals in a *gluco* configured unit, while the use of a shorter mixing time (20 ms) limits the magnetization transfer with the result that only the carbon attached to the vicinal proton is detected, like exploited to analyse the ribitol moieties of the teichoic acids from *Staphylococcus aureus* by Gerlach et al. (2018).

Finally, even though the addition of the TOCSY transfer in the HSQC sequence is probably the most used extension, other hybrid schemes have been used in the "glycan polymer" field, by adding COSY, NOESY or ROESY spectra as additional components, although they appear less used in the common practise.

6. Spectra processing

As rule of thumb, no NMR experiment is complete until it is properly processed and presented. Indeed, the correct processing is crucial because it deeply impacts the accuracy of the information contained therein. Then, the study of any spectrum can start only after that it is properly transformed, and calibrated. Then, the way spectra are traced is also important for the presentation and the understanding of the results.

The considerations done hereafter assume that the acquisition windows of the spectra were sufficient to sample of the signals, so that none of them occurred as folded, or wrapped back in the spectrum. More details about practical aspects can be found elsewhere (Claridge, 2016a).

The next sections will refer to some commands that are used in the Topspin program, and for this reason they have no acronym.

6.1. Spectra transformation

When any experiment finishes, the instrument returns an FID that indicates how the frequency of each nucleus has decayed over time, namely how the nuclei re-align to the static field after that the perturbation, a pulse or a series of them, is over. Data in this form are all but useful to extract structural information, therefore they are converted from the time domain (TD) to the frequency domain by applying the Fourier transform, a mathematical approach that decomposes the FID curve into its constituent frequencies, namely the NMR spectrum. Of note, Fourier transform produces both real and imaginary data, and only half of the output, namely the real part, is taken to produce the spectrum in the absorptive mode.

At this stage, the sequence of operations depends on the experience of the operator, and the one given hereafter reflects that common practise adopted in our laboratories.

The first step when setting the processing parameters generally consists in (at least) doubling the number of points used during the acquisition of the FID, in the so-called *zero-filling* process. This manipulation adds zero-points to the end of those acquired during the experiment, the TD, so that it enables the recovery of the points that are lost by effect of the Fourier transformation mathematical process.

Then, the spectrum needs to be phased to remove any dispersive component that twists the signals (Fig. 5a) away from their pure absorptive form (Fig. 5b). This process is performed by adjusting the zero-order (PHC0) and the first-order (PHC1) phase parameters: the first, PHC0, affects in the same way all the signals of the spectrum, so it is worth to operate this initial correction by putting in-phase a signal at one edge of the spectrum (or to set one manually as pivot-signal). Then, the other parameter, PHC1, is used to bring in-phase the other signals; the entity of the correction increases linearly being zero on the signal at one edge of the spectrum (or chosen as pivot-point) and it reaches the maximum for those that are at the opposite edge (or far from the pivot).

The same phase-correction criteria apply for the bidimensional spectra that require a phase-sensitive transformation (all those reported in this review, except the HMBC spectrum that is in the magnitude mode). In this case, the phasing has to be performed in both dimensions, which is done by working on each dimension per time and by extracting a couple of rows along the F2 dimension (or columns if phasing the F1 dimension), in turn phased by the same approach shown for the monodimensional spectrum. Fig. 5e reports a HSQC spectrum wrongly phased exclusively in the F2 dimension, as visible from the appearance of positive and negative densities with a strong tailing parallel to the F2

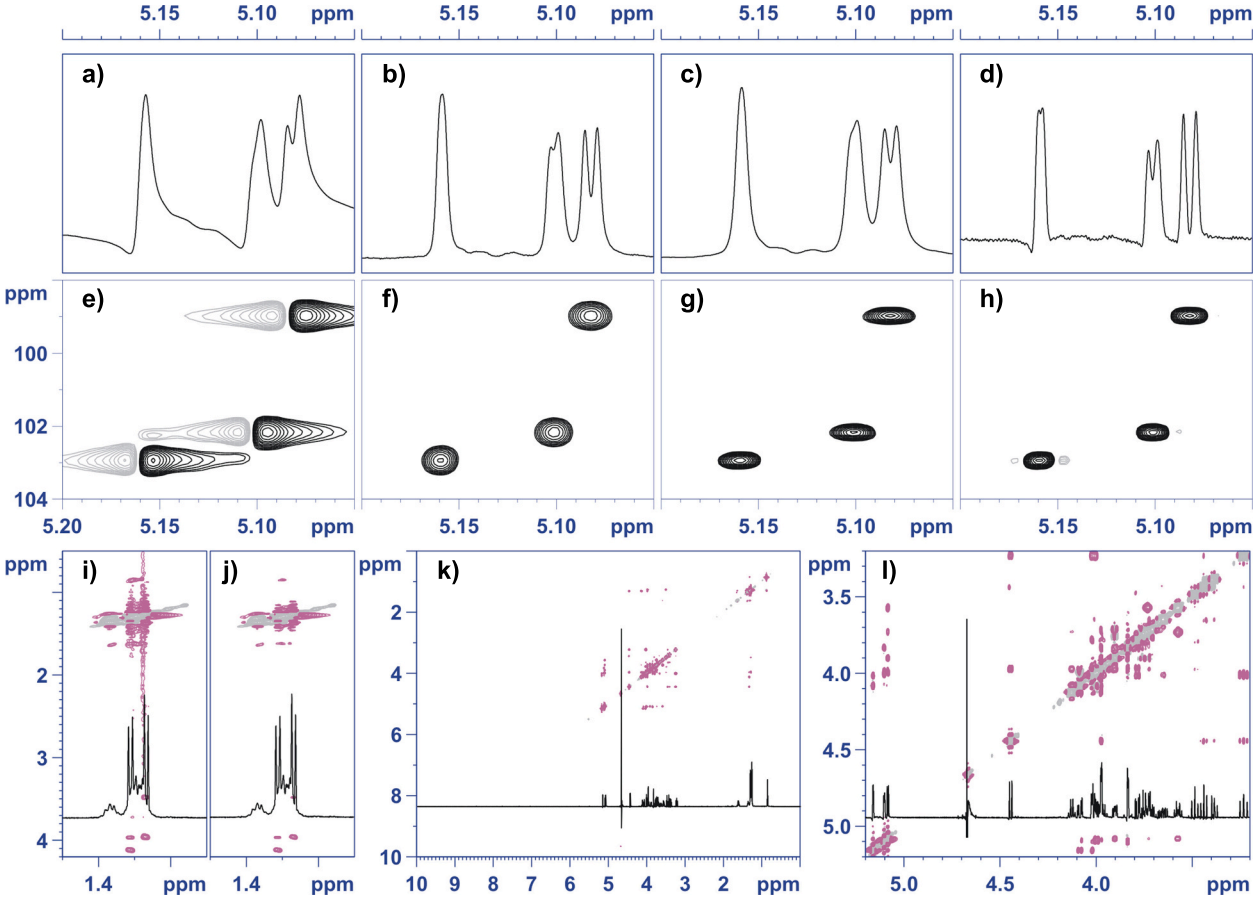

Fig. 5. Spectra measured for the tetrasaccharide **1** (Speciale, Laugieri, et al., 2020). Expansion of the proton (panels a–d), HSQC (panels e–h), and T-ROESY (panels i–l) spectra. a,e) examples of wrongly phased spectra; b,f) the two previous spectra after phasing, with no changes in other transformation parameters; c) improvement of "b" by application of exponential multiplier window function: lb. = 1; d) improvement of "b" by application of Gaussian multiplier window function, lb. = −1.60; gb = 0.52; g) HSQC spectrum transformed as in "f", windows functions: F2, QSINE = 2; F1, SINE = 2, along with forward Linear Prediction (LPfr) in F1 (NCOEF = 50; LPBIN = 0); h) the same as in "g" except the window function in F2 is QSINE = 4. i,j) region of the T-ROESY spectrum detailing the high field region, without (panel i) or with (panel j) baseline correction along F1. k) T-ROESY spectrum reported in full size. l) expansion of the T-ROESY spectrum detailing the anomeric – carbinolic region.

axis, opposite to the defined round-shape acquired as result of the correction (Fig. 5f).

Then, upon the proper phasing, the transformation of the FID can be optimized by "playing" with the so-called window functions, or by reconstructing key portion of the FID by Linear Prediction (LP), or by resorting to both possibilities.

Window functions are nothing else than mathematical functions that manipulate the FID by giving a different weight to the points that compose it: the overall effect of this procedure is that they can suppress – for instance – the first part of the FID, or the last or both.

As for the monodimensional spectra, the window functions mostly used are the exponential multiplication (EM) and the Lorentz-Gauss transformation or Gaussian multiplication (GM).

The effect of EM is to improve the signal-to-noise ratio because it resizes the FID by multiplying it with an exponential function that decays to zero after a certain time. In this way, the last part of the FID is forced to zero with the result that the instrumental noise contained therein is removed. The effect is cosmetic, and it yields to an increase in sensitivity (or an improved signal-to-noise ratio): the baseline appears smoother but the price to pay is the reduced resolution, namely the line broadening that goes along with it, as visible in Fig. 5c where the valley of the doublet at 5.08 ppm appears less pronounced compared to that of the spectrum with no manipulation (Fig. 4b). The behaviour of the EM is modulated by one parameter, LB (line broadening) that is always a positive number, and that for routine proton spectra is about 0.3–0.5 Hz, even though the common practise relies on finding the best compromise between sensitivity and decreased resolution by visual inspection of the result.

Contrary to EM, the GM window function combines two exponential curves which have the effect to convert the original Lorentzian lineshape of the NMR signal into a Gaussian curve. Accordingly, the GM function is modulated by two parameters, LB and GB. The first, LB, has the same purpose seen for the EM window function, with the only difference that in this case it takes a negative value; GB instead is always positive, and it indicates a fraction of the total acquisition time. The choice of these two values is done by trial and error while pondering the effect on the spectrum by visual inspection. Application of GM leads to a remarkable improvement of the resolution, as visible in Fig. 5d where the valley between to the two sides of the doublet at 5.08 ppm reaches the baseline. This improvement enables the accurate reading of the coupling constants also for signals where this value is extremely narrow, as that at 5.15 ppm ($^3J_{H1,H2}$ = 1.2 Hz). However, the use of GM has its toll that consists in the decrease of the signal-to-noise ratio, as visible in the baseline of Fig. 5d along with the introduction of some distortion in the shape of the peaks that now have a mild wiggle at the left part.

From the practical viewpoint, the ProcPars tab of the Topspin program presents a window function (WDW) section with a pull-down menu in which this parameter can be set to *no*, meaning that no function is used, or to EM (exponential), GM (gaussian), SINE, QSINE, and others. Then, the remaining part of the WDW section enables the user to set the values for the selected window function.

Although the use of window functions is important for 1D spectra, their application to the two dimensions of the 2D spectra is crucial to enhance their quality.

Regarding the kind of windows functions, those used more often derive from the trigonometric SINE function, which consists of one half of a sine wave, so that it forces to zero the beginning and the end of the FID, while reaching a maximum at half of the acquisition window. This function can be used as it is, or its zero starting point can be shifted by operating on the SSB parameter, or it can be replaced by its variant, the squared SINE function or QSINE, for which the same shifting mentioned for SINE applies.

The SSB parameter affects how the SINE (or the QSINE) function operates and it is an integer value $\geq$0, able to impact the sensibility and the resolution of the spectrum: a low value increases the sensibility reducing the resolution, the contrary happens by increasing its value with the risk of introducing some distortions, leaving to the operator the choice of the best compromise (check Fig. 5g,h).

The application of the window functions to improve the 2D spectrum is only part of the story since other manipulation can be applied together with or even after them.

Apart from zero filling, the FID can be manipulated by linear prediction, a mathematical procedure used to replace the missing data points because able to extrapolate their behaviour from a selected portion of the FID. There are two types of LP - backward and forward - depending on the location of the data to be predicted. In detail, the backward LP is used when there is a corruption of the first few data points of the FID due to instrumental imperfections, while the forward LP (LPfr) reconstructs the end of the FID when it is truncated because the acquisition time was too short or poorly sampled. To enable LP, the ME_mod in the "fourier transform" section of the ProcPars of Topspin, must be set to LPfr/LPfc for forward LP (real and complex data, respectively) or to LPbr/LPbc for backward LP (real and complex data, respectively).

As for the 2D NMR spectra of carbohydrates, they mostly require the use of LPfr along the indirect dimension, F1, because the one more plagued from resolution issues. LPfr is modulated by two parameters, NCOEF and LPBIN. NCOEF is the number of coefficients used to reconstruct the FID, and it must be set to value >0 to perform the linear prediction, while LPBIN indicates the number of points used for the prediction, that when 0 matches that set for TD. Of note, the size of the TD can be eventually decreased by operating on the TDeff parameter, so that the last part of the FID is cut off, with the overall effect of a reduction of the background noise of the spectrum.

Then, the application of zero-filling and LP produces a new FID that is in turn transformed with the window functions of choice, and the final effect of all these of the various manipulations is generally judged by eye-inspection. Indeed, after the correct phasing, the HSQC spectrum in Fig. 5f displays the expected round-like shape of the densities, which are further sharpened by effect of the LPfr along the F1 (Fig. 5g). Here, the two panels are transformed with the same windows functions (QSINE = 2 in F2, and SINE = 2 in F1) and the same size of points (SI) in the ProcPars menu. An additional improvement is then afforded by manipulating the FID as in Fig. 5h through a slight change of the window function in F2 that changes QSINE from 2 to 4: this narrows further the densities in the F2 dimension leading to an additional improvement in resolution at the cost of some minor artifacts (or distortion) that take the form of tiny antiphase densities along F2 at the sides of anomeric densities at $^1H/^{13}C$ 5.16/102.9 and 5.10/102.1 ppm.

Once the transformation process has been optimized, the quality of the spectrum can be further improved through a baseline correction, which subtracts a polynomial curve to either the F2 or the F1 dimension of the spectrum or to both. The baseline correction can be applied by using the "abs" command (abs1 or abs2, depending on which dimension is applied the correction, F1 or F2, respectively), while the degree of the polynomial equation is determined by the ABSG parameter, generally set to 5 in the program by default, and the type of correction depends on the BC_mod choice that indicates the type of function used. This last correction is particularly useful for spectra presenting noise along the F1 dimension, as reported in Fig. 5i,j.

Last, it should be noted that all the manipulations listed above are almost mandatory to improve the spectra recorded for samples poorly abundant, for which the recording of the NMR spectra is always a challenge.

Then, the take home message about FID processing is that this stage of the work is worth of all the time dedicated to it since it really contributes to the improvement of the data and to the accuracy of the information gathered from them.

6.2. Spectra calibration

In general, spectra need to be calibrated so that the chemical shifts of

the product under investigation can be readily compared to those of another taken as reference – for instance the free monosaccharides and the methylglycosides (Tables S1 and S2) – or to establish the identity (or the difference) between different products, or to follow their temperature (as well as their pH) dependency.

Traditionally, spectra of organic compounds are calibrated by using tetramethylsilane (TMS) as reference and setting both its ^{1}H and ^{13}C signals to 0 ppm. However, TMS is not soluble in water therefore it has been replaced with sodium trimethylsilylpropanesulfonate (DSS) or trimethylsilylpropanoic acid (TSP or TMSP); in both cases, the chemical shifts of the trimethylsilyl group are relatively insensitive to pH and set to ^{1}H/^{13}C 0.0/−1.6 ppm, respectively. The main drawback of DSS (or TSP) is that they need to be removed from the sample by chromatography or other purification means. For this reason, their use has been dropped in favour of volatile molecules, as acetone (^{1}H/^{13}C 2.225/31.45 ppm) that have the advantage of being easily removed, while their signals occur in areas generally devoid of carbohydrate-related signals.

6.3. Spectra presentation

The analysis of the spectra is an interactive process conducted on the computer screen, therefore this section will deal primarily with the presentation of the spectra that is required for reporting needs or for publications.

First, the conventions presented in Section 5 to label the densities of 2D spectra will be summarized here again. Briefly, any 2D NMR spectrum can be considered as a graph with x,y coordinates, that correspond to the F2 and F1 dimensions, respectively. Therefore, each density will be defined by its F2/F1 values and instead of reporting the chemical shifts values, the labels will indicate the structural location of the nucleus. Taking the HSQC spectrum as example (Fig. 3c), the density of the anomeric signal of **G** with will be named $\mathbf{G_1}$: its position along the F2 (or x) axis is the proton chemical shift while that along the F1 axis is the ^{13}C value. Regarding the COSY (or the TOCSY and the T-ROESY, Fig. 4a–d), the density $\mathbf{A_{1,2}}$ indicates the anomeric chemical shift of the unit along F2 and the chemical shift of H-2 along the F1 dimension. As for the *inter*-residue correlations found in the HMBC and in the T-ROESY (or NOESY) spectra, the notation will list the two nuclei according to their F2,F1 order. For instance, in the T-ROESY spectrum (Fig. 4b) the density $\mathbf{D_1C_4}$ indicates H-1 of **D** along the F2 dimension of the spectrum, and H-4 of **C** along F1. The same label, $\mathbf{D_1C_4}$, is used in the HMBC spectrum (Fig. 4j), with the difference that here "$\mathbf{C_4}$" indicates the C-4 of the **C** unit.

Then, another point is about which area of the spectrum is worth to display in order to present all the information relevant for the attribution without sacrificing clarity. Presentation of the full spectrum is not the best choice because the densities will appear in a restricted area of the figure where it will be hard to label and to see them all (Fig. 5k). Therefore, a first level of improvement is achieved by getting rid of the regions devoid of signals and by reporting the anomeric and the carbinolic regions (Fig. 5l); at this stage, the full spectrum can be given as supporting material (Figs. S1–6). As final improvement, the expansion in Fig. 5l can be broken in the slices that are more informative (Fig. 4a,b) that in this case are those detailing the NOE effects found in the anomeric region of the tetrasaccharide **1**. The slice in Fig. 4a shows clearly which densities belong to **B** and which to **C**, and it enables their labelling, while the same would have been difficult in Fig. 5l, and not achievable at all Fig. 5k.

Last, spectra presentation benefits from having a couple of them overlap and distinguished by adopting different colours, as in Figs. 3 and 4.

Then, the take home message for this section is that presentation of the NMR spectra has to privilege clarity to enable colleagues to follow the assignments and to learn from the work of others.

7. Interpretation of the chemical shifts

The target of the NMR analysis is to assign the proton and carbon chemical shifts of the residues that compose the glycan, and to define how they are interconnected. Still, even when the full set of spectra is available, the interpretation of the NMR data is prone to errors and for this reason it is worth to support the attributions by checking the NMR structural data available literature or through a careful analysis of the chemical shifts found.

Regarding with the finding of the NMR structural data, this can be achieved by querying some public repositories, and those more relevant for this purpose are ECODAB, dedicated to the *Escherichia coli* O-antigenic structures (Stenutz et al., 2006), Glycosciences.de (Böhm et al., 2018), and the Carbohydrate Structure Database (CSDB) (Toukach & Egorova, 2015), probably the most comprehensive for the repertoire of tools offered, and also because it covers prokaryotes, plants and fungi while being regularly updated. Accordingly, the interrogation of these databases answers to the question about the novelty of the glycan found and, in case the database returns a structural match, it makes possible the comparison of the chemical shifts of the glycan with those of reported for hit found, so that the identity between the two polysaccharides can be established. Of note, for spectra measured in the same conditions (temperature, pH, and concentration to a lesser extent), the match has to consider both the proton and the carbon chemical shifts of each unit of the polysaccharide, otherwise it has to limit to the ^{13}C values only, because these are less sensitive to the sample preparation and acquisition conditions used.

In case the glycan found is new and it is not possible to resort to literature data to support the conclusion taken, it is still possible to analyse the chemical shift found to rule out possible inconsistencies.

This analysis considers the variations that some substituents induce for the nuclei at the site of attachment or next to it. Such variations refer to the difference between the value found and that of a reference, that depending on the case considered can be the monosaccharide in the free reducing form or the corresponding methylglycoside (Tables S1, 2).

With regard to the nature of the substituents, glycans can be decorated with a rather large plethora of groups as recently reviewed (Di Lorenzo et al., 2021), and the substituents most commonly found are either alkyl or acyl groups.

Then, the remaining of this section will discuss how these substituents influence the chemical shift of a certain unit, apart from giving relevant information about the so-called glycosylation shift.

7.1. Alkyl substitution

The alkyl group mostly found in glycans is the methyl group: it can be linked as ether to a hydroxyl function or as ester to the carboxylic group of an uronic acid, like in pectins. Monosaccharides can be considered as a sort of alkyl group, and they will be discussed in the next subsection.

About the methyl group, when linked as ether (^{13}C 57–60 ppm), it has a profound effect on the chemical shift of the carbon atom bearing the modified hydroxyl function. This carbon atom is referred to as C_α and its ^{13}C value is shifted of about 6–10 ppm at lower fields compared to the value reported for the not substituted unit. On contrary, the carbon atoms next to C_α, named C_β, experience a mild effect in the opposite direction of about 0–3 ppm (Fig. 6a). Proton chemical shifts are not markedly influenced, or the entity of the variation is of no predictive value.

As example of the alkylation effect is the variation of the chemical shift of the anomeric carbon when the monosaccharide is transformed in the corresponding methylglycoside (Tables S1, 2). Indeed, the C-1 value of the methylglycoside is always 6–10 ppm above that of the free monosaccharide, while the C-2 value is always few ppm less (~2 ppm). In this case, the monosaccharide in the free reducing form has been taken as reference to quantify the entity of the chemical shift variation.

In the common practise, the monosaccharide unit under

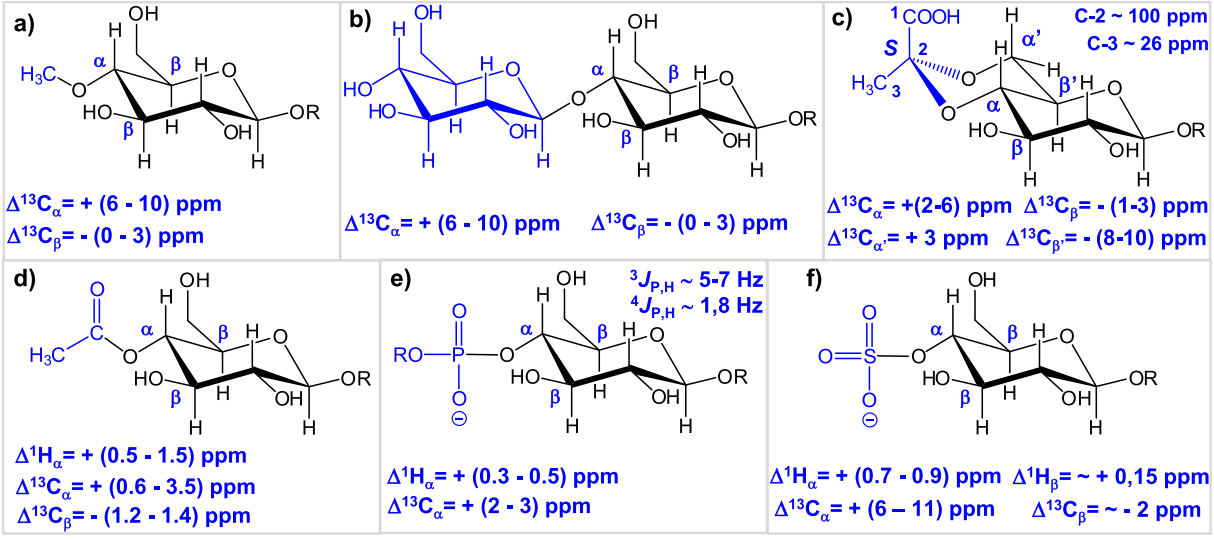

Fig. 6. Type of substituents mostly found in polysaccharides. The monosaccharide unit that is substituted is drawn in black, while the substituent is blue. Each of these groups influences the chemical shifts of the nuclei in the immediate proximity, and the effect is classified as α-effect when the nucleus is geminal to the substituent, or β-effect when the nucleus vicinal. For each group, the entity of the α/β-effect, along with some other properties are reported within the corresponding panel. a) Methyl substituent taken as model of any alkyl substituent; b) glycosyl substituent; c) pyruvic acid linked as 4,6-cyclic ketal with the *S* configuration at its C-2 position, otherwise also named as 1-*S*-carboxyethylidene group; d) acetyl group taken as model for any acyl substituent; e) phosphate group linked as diester; f) sulfate ester.

investigation is always part of a glycan therefore it is not in the free reducing form, and for this reason the ^{13}C chemical shift comparison has to take into account those of the methylglycoside (if available) and not those of the free monosaccharide.

When the methyl group is linked as ester to an uronic acid, it does not induce any relevant variation to the proton or carbon chemical shifts of the sugar, with the exception of C-6 that undergoes a shift at high (or up) fields of about 4 ppm, namely it is found at ca. 171 ppm, as reported for the esterified galacturonic unit of pectin oligosaccharides (Mort et al., 2013). Moreover, the occurrence of the methyl group can be inferred from the presence of a signal at about ^{13}C 54–57 ppm, and by the observation that by varying the pH of the sample, the chemical shift of the proton or of the carbon signals of the uronic acid do not change consistently, on the contrary from what happens for the same unit when it is not esterified (note the variation of C-5 and C-6 with the pH reported for the glucuronic acid in Table S2).

7.2. Glycosyl substituents

Sugars themselves can be considered as substituents of another sugar unit (the one substituted), and from the spectroscopical viewpoint, the consideration done for the ether-linked methyl group apply for carbohydrate substituents, with some extensions.

Taking the methylglycoside as reference, it has been observed that an aldose monosaccharide substituent, either in the pyranose or in the furanose form, shifts of about 6–10 ppm to lower fields the carbon of the residue to which it is attached (Fig. 6b). This is the so-called as α-glycosylation effect, and it goes together with the β-glycosylation effect, that instead is the variation of about 0–3 ppm up-field observed for the carbon atoms adjacent to the one that is substituted (Fig. 6b). With regard to the α-glycosylation effect observed for keto-sugars, it is weaker compared to that of the aldose counterparts, and it is hardly above 4 ppm as demonstrated by the C-1 of the glucose unit of sucrose, that is at 95.0 ppm (Fig. 3c), about 2 ppm above the value reported for the not substituted monosaccharide (92.9 ppm) taken as reference (Table S2). As consequence, the determination of the site of attachment of ketoses is sometimes complicate to achieve because of the lack of a pronounced α-glycosylation effect.

Then, the rules set above for the interpretation of the chemical shift variation are those mostly used in the common practise, and they are sufficient to explain the variations found on a qualitative basis.

However, this analysis can be made quantitative by taking into account that the substituent and the substituted units influence the chemical shift of each other in a reciprocal fashion, and that the entity of such variations is governed from stereochemical factors, namely from the nature of both units involved and from the type and location of the glycosidic linkage (Lipkind et al., 1988; Shashkov et al., 1988).

Notably, these rules from Lipkind et al. (1988) and Shashkov et al. (1988) have been formulated to predict the magnitudes of the glycosylation effects, and their application can go beyond this scope since they can be used for the determination of the absolute configuration of one of the sugars, assuming that that of the other unit is known. The discussion of these rules and of their application is beyond the scope of this review, and readers interested in deepening this aspect are encouraged to source to the original articles (Lipkind et al., 1988; Shashkov et al., 1988).

7.3. Pyruvate as ketal substituent

Pyruvic acid, or pyruvate when dissociated, is a three carbon atoms organic keto-acid that decorates many bacterial polysaccharides and few eukaryotic glycans (Hager et al., 2019) by forming a cyclic acetal, named 1-carboxy-ethylidene, realized by the keto function upon the linkage with two hydroxyl functions of the monosaccharide. Most of the times, pyruvate is placed across the 4,6 positions of the sugar residue therefore closing a 6-membered ring (Fig. 6c), or it can join the 2,3-or the 3,4-positions leading to a 5-membered cycle. The size of the ring can be deduced from the C-2 value of this molecule, that is at ~100 ppm for six-membered cycles, or at ~110 ppm when it forms a five membered ring.

The first insight into the presence of this substituent is inferred by the finding in the proton spectrum of a sharp singlet at about 1.3–2.0 ppm, that is methyl group. Then, the exact location of pyruvate can be deduced by analysing the carbon chemical shifts of the monosaccharide since the positions blocked as cyclic acetal are shifted to lower field (α-effect), while those adjacent experience an opposite (β-) effect. Notably, the transformation of the keto function at C-2 of pyruvate into an acetal generates a chiral centre that can be either *R* or *S* configurated,

and the comparison of the $^{1}H/^{13}C$ values of the pyruvate methyl group with those reported in literature for model compounds (Garegg et al., 1980; Gorin et al., 1982), is sufficient to address this parameter in most of the cases.

The six-membered 4,6-ketal rings are those better covered by literature (Garegg et al., 1980). For these derivatives, the ^{13}C signal of the methyl group is quite diagnostic of the configuration of the new stereocenter: when in the equatorial position (C-2 in the *S* configuration) the methyl group resonates at ~26 ppm (Fig. 6c), and when in the axial position (C-2 in the *R* configuration) it is found at ~17 ppm. Regarding the ^{13}C chemical shift variations at the site of attachment (Fig. 6c), it must be noted that there exist two α-effects (α- and α′) and two β-effects (β- and β′). Generally, the two α-effects are positive and of modest entity, ~6 ppm when the monosaccharide is a mannose or a glucose, or ~2 ppm is the unit is a galactose; the α′-effect is the same for all and ~3 ppm. As for the β-effect, the ^{13}C signal of the C-3 of the monosaccharide is shifted upfield of about 1–3 ppm, while the β′-effect experienced from the C-5 is more intense and it is about 8–10 ppm (Jansson et al., 1993).

As for the 3,4 and 2,3 ketals, the ^{13}C shifts have been studied in detail only for some 3,4 derivatives (Gorin et al., 1982). Here, the definition of the stereochemistry of the C-2 of the ketal is less straightforward: the ^{13}C value of the C-2 and of the C-3 (the methyl group) are about 107 and 24 ppm, respectively, independently from the *R/S* configuration of C-2. On the contrary, the proton chemical shift of the methyl group enables the differentiation between the two isomers, since in the *R* configuration, the chemical shift of this group measured in deuterated water is 1.94 ppm, otherwise it is 2.04 ppm (Gorin et al., 1982).

For all the cases where the existing literature is of no support, the configuration of this chiral centre should be determined by observing the NOE effects between the methyl group and the neighboring protons of the sugar unit, once that its position is determined by analysing the carbon chemical shift values of the monosaccharide unit.

7.4. Acyl substitution

The acyl group with the major occurrence in carbohydrate is the acetyl group and for this reason the consideration done hereafter will take this group into consideration even though they can be extended to any acyl substituent.

First, the acetyl group can be present as ester (Fig. 6d) or as amide and it can be recognized in the proton spectrum due to the appearance of its methyl group as a singlet at 2.0–2.2 ppm when ester linked, or at 1.9–2.0 when present as amide, while the ^{13}C values associated are 21–23 ppm and 173–175 ppm for the methyl and the carbonyl group, respectively.

The acetyl group affects both the proton and carbon chemical shifts of the nuclei geminal to its location (α-acylation effect), as well as those of the vicinal carbon atoms (β-acylation effect) (Fig. 6d).

Regarding the α-acylation effect, the signal of the geminal proton undergoes a strong downfield shift that can reach up to 1.5 ppm, so that it is not rare to find any of these protons in the anomeric region of the spectrum. Differently, the shift induced on the hydroxymethylene protons is of minor entity and it is not greater than 0.5 ppm, so that these protons are still found in the carbinolic region of the spectrum (Agrawal, 1992), although at the border with the anomeric region. As for the carbon signals, the acetyl group induces a downfield α-acylation shift of 0.6–3.5 ppm, while the β-acylation effect is opposite and it is about 1.2–1.4 ppm (Agrawal, 1992).

Hence, the variation of chemical shifts observed for both proton and carbon nuclei, along with the presence of the intense signal of the methyl group, contribute to ascertain the presence and the location of this substituent. The HMBC spectrum can be used as additional confirm, since it generally displays a correlation between the carbonyl group and the proton of the monosaccharide geminal to this substituent.

7.5. Phosphate substitution

The phosphate group occurs in different types of carbohydrates, where it is primarily linked as diester (Fig. 6e) as in the repeating unit of several capsular polysaccharides and in the teichoic acids from Gram-positive bacteria. Regarding Gram-negative bacteria, it is a key element of the lipopolysaccharide, the major constituent of the external leaflet of the outer membrane. Here, the phosphate group is a conserved motif of the lipid A moiety of the lipopolysaccharide (Di Lorenzo et al., 2021), a β-(1 → 6)-glucosamine disaccharide phosphorylated at the α-anomeric position of the glucosamine at the reducing end, and at O-4 of the other unit. Most of the times, these two phosphates are in the form of monoesters, otherwise they can be further substituted with other groups as reviewed elsewhere (Molinaro et al., 2015). Other phosphate groups may occur in the core-region of the lipopolysaccharide and their presence and location depends on the bacterial species considered.

Compared to the chemical shift variation induced from the deshielding of the acyl (Section 7.4) and the sulfate groups (Section 7.6), the effect of a phosphate is less pronounced. Indeed, the proton geminal to the phosphorylation site undergoes a downfield shift of about 0.3–0.5 ppm, while that of the corresponding carbon is about 2–3 ppm in the same direction.

However, the great advantage of this group is that and the ^{31}P isotope, the major in terms of natural abundance, is NMR active so that it can be directly measured by adapting the sequences normally set for ^{13}C.

Apart from the possibility opened for ^{31}P NMR spectroscopy, the presence of this group can be already inferred in the proton spectrum, because it is scalarly coupled to the proton geminal to the site of attachment with a coupling constant $^{3}J_{P,H}$ ~ 5–7 Hz, and to the vicinal proton with a $^{4}J_{P,H}$ ~ 1.8 Hz, so that it impacts on the multiplicity observed for these protons. Finally, the chemical shift of ^{31}P in phosphates may change by varying the pH and for this reason it is recommended to measure this nucleus in a pH buffered solution.

7.6. Sulfate substitution

Sulfate is an important substituent for a large number of polysaccharides, such as carrageenans and glycosaminoglycans (except hyaluronan) where it modulates their mode of interaction with their cellular receptors. On the contrary, the occurrence of this inorganic substituent in bacterial glycans is rather sparse, and it is mainly relegated to bacteria from marine sources as denoted by the query “sulfuric acid” in the composition menu of CSDB (Toukach & Egorova, 2015).

The sulfate group mostly occurs as ester, except in heparin where it can be also amide-bound.

As the acyl group, this substituent has strong deshielding properties that influences the chemical shifts of both ^{1}H and ^{13}C nuclei next to the site of attachment (Fig. 6f). The effect on the geminal proton (or the α-effect) depends upon the nature of the hydroxyl group to whom it is attached. In case of the hydroxymethyl group, the two protons undergo a downfield shift of about 0.4–0.5 ppm, whereas the shifts observed for carbinolic protons is much larger, 0.7–0.9 ppm (Fig. 6f). Then, the proton adjacent to the sulfated position is shifted to downfield as well, although the entity of this β-effect is minor compared to the α-effect, and it is about 0.25 ppm if the sulfate occurs on a primary hydroxyl group, or 0.15 ppm if it is on a secondary hydroxyl function (Agrawal, 1992).

As for the ^{13}C values, the sulfate causes a pronounced downfield shift of 6–11 ppm to the signal of the carbon directly bearing the substituent, while the vicinal carbons experience a mild upfield shift of about 2 ppm. Importantly, the variation of the proton chemical shifts alone is not sufficient to claim the presence of a sulfate since a comparable variation occurs upon acylation, while the observation of the carbon chemical can make the difference and be used to infer (or confute) the presence this group.

8. Practical example: the tetrasaccharide 1 (structure in Figs. 1, 4, chemical shifts in Table S3)

For this oligosaccharide, the interpretation of the NMR data will be done as if the structure of the molecule were unknown to underline how to proceed with the study of the spectra, and what is the power of this technique along its limits.

Generally, a first idea of the complexity of the product is gained by looking at the anomeric region of the polysaccharide and by counting the anomeric signals present. The proton spectrum (Fig. 1d) gives a first indication; however, it is necessary to confirm this first evaluation by analysing the HSQC spectrum to be sure that the number of the signals in the anomeric region does not result from the overlap of two or more protons. Then, it is also worth to inspect the HMBC spectrum (Fig. 4r) to rule out the possibility that there is one (or more) ketose unit, because this type of sugar has no anomeric proton so that its anomeric carbon is not detected in the HSQC spectrum.

Regarding **1**, it displays 4 proton signals in the anomeric region (Fig. 1d), each correlated to one carbon, while the HMBC does not show densities of anomeric carbons diverse from the four already detected in the HSQC. Therefore, **1** is a tetrasaccharide and the study of the 2D NMR spectra continues with the evaluation of all the proton and carbon chemical shifts of each unit, with the final aim to deduce their stereochemistry and how they are linked to each other. The starting point of the NMR analysis depends on the "feelings" of the operator, and generally the choice falls on the unit with more cross peaks in the TOCSY spectrum since the attribution is simplified by the prior knowledge of where its signals should be found.

8.1. *D unit, a terminal β-xylose*

Therefore, starting from the anomeric signal labelled as **D**, the proton chemical shift is typical of a residue in the β configuration (4.44 ppm, Table S3), which is confirmed by the value of the $^{3}J_{H1,H2}$ coupling constant value (7.5 Hz). Then, the TOCSY spectrum (Fig. 4d,g) reports 6 different correlations with one coincident with that from the DQF-COSY (hereafter simply mentioned as COSY), Therefore, the correlations common to the COSY and TOCSY spectra define the position of $\mathbf{D_2}$, while the total number of TOCSY correlations suggests that this unit is *gluco* configured. The recognition of **D** as xylose, and not glucose, takes advantage of the HSQC-TOCSY spectrum that displays four densities in addition to the anomeric density (Fig. 4l,r). In particular, the one at 66.2 is a $-CH_2-$ because in the HSQC its phase is opposite to the others (Fig. 4m), and it correlates with the protons at 4.01 and 3.23 ppm, both detected in the TOCSY spectrum of this units (Fig. 4g). Moreover, this carbon signal (66.2 ppm) is shifted downfield when compared to the value of not substituted sugars (~62 ppm) because its hydroxyl function is linked to another carbon atom. Taking together this first set of information, it is possible to infer that **D** is a xylose unit (Fig. 1a), with the hydroxyl function at C-5 involved in the pyranose cyclization of the unit. Then, the COSY correlation path of **D** is rather straightforward to follow (Fig. 4e). Indeed, the $\mathbf{D_{2,3}}$ or $\mathbf{D_{3,2}}$ cross-peaks are visible although rather close to the diagonal of the spectrum, and $\mathbf{D_3}$ in turn has a clear correlation with $\mathbf{D_4}$, which then identifies the two H-5 protons, $\mathbf{D_{5eq}}$ (4.01 ppm) and $\mathbf{D_{5ax}}$ (3.23 ppm). Finally, the two $\mathbf{D_5}$ protons are interrelated by the correlations generated by their geminal coupling, $\mathbf{D_{5eq,5ax}}$ or $\mathbf{D_{5ax,5eq}}$. It must be noted that the two $\mathbf{D_5}$ have very diverse chemical shifts: this is expected since they are diastereotopic protons and the large difference denotes the fact that one is blocked in the equatorial position while the other is axial. Of note, the chemical shift of axial protons is generally always lesser than their equatorial counterpart.

Another relevant point regards the shape of the COSY cross-peaks, which is made of several densities of opposite phase: this pattern is due to the fact the information of the coupling constants of each proton is maintained from the pulse sequence used (DQF-COSY). In principle, these *J* values can be read on the cross peaks by measuring the distance between their components.

Taking the $\mathbf{D_{5ax}}$ proton as example (Fig. 7), the following considerations apply to the pattern observed along the F2 dimension; the F1 dimension is not discussed because its resolution during acquisition – lower than that in F2 – causes the loss of the fine structure expected for the cross-peak.

Then, looking at $\mathbf{D_{5ax,4}}$ it is possible to note that it is composed from two sets of four densities each that align along a horizontal line; these two sets are symmetrical except for their sign (or colour) that is inverted. Then, the considerations done for one set apply to the other, as well as to the other cross-peaks of the COSY spectrum.

First, $\mathbf{D_{5ax}}$ is coupled to $\mathbf{D_4}$ and $\mathbf{D_{5eq}}$ and the position of the four densities in each row is indicative of the value of these two coupling constants. Indeed, the cross-peak $\mathbf{D_{5ax,4}}$ has the so-called "active coupling" with $\mathbf{D_4}$ that corresponds to the distance between the first two densities of opposite sign (or the last two because of the symmetry of the cross-peak), while the passive coupling indicates the $^{2}J_{5ax,5eq}$ coupling constant, and it correspond to the distance between the two in-phase densities (Fig. 7). Then, the role of active/passive coupling is exchanged in the other cross peak, $\mathbf{D_{5ax,5eq}}$, and to a trained eye, this swap of phases can be used to follow the connections between different cross-peaks when the spectrum is crowded.

Regarding the estimation of the coupling constants from a DQF-COSY, it must be noted that the values might not be very accurate and may diverge consistently from those read on the proton spectrum (Fig. 7, top trace). The main problem is that when the value of the two coupling constants is very similar, the two inner densities may cancel each other, and when the cancellation is partial (like in $\mathbf{D_{5ax,4}}$ and $\mathbf{D_{5ax,5eq}}$) what is left leads to not accurate values. In this case, the apparent values read on the COSY spectrum as $^{3}J_{5ax,4} = 7.0$ and a $^{2}J_{5ax,5eq}$ of 14 Hz, while the correct value is that read on the proton spectrum, 11.0 Hz for both. Other factors can complicate the shape of the cross-peak and therefore the evaluation of the coupling constants. One is the roof effect, that in this sample occurs between $\mathbf{D_2}$ and $\mathbf{D_3}$, because close in terms of chemical shifts and coupled with a large coupling constant; the final effect is that the triplet symmetry expected for both signals is altered so that the three lines do not respect the 1:2:1 proportion. Accordingly, the inner part of the $\mathbf{D_{2,3}}$ (or the $\mathbf{D_{3,2}}$) correlation is not symmetrical as that found for $\mathbf{D_{5ax}}$, and it cannot be used to estimate any coupling constant value. Another example of complication is when the two coupling constants perfectly cancel each other, as visible for the $\mathbf{A_{4,3}}$ and the $\mathbf{A_{4,5}}$ correlation that are void in the middle (Fig. 7).

For all the reasons above, the use of the DQF-COSY for the exact estimation of the coupling constant is discouraged, unless the spectra are acquired with a high resolution (along F2). In the present example, the coupling constant values can be used in a qualitative way: all the values appreciated on the COSY spectrum are rather large (above 7 Hz), therefore they indicate a *trans*-diaxal disposition of all the protons (except $\mathbf{D_{5eq}}$), which is consistent with the hypothesis derived from the TOCSY spectrum that the **D** is a xylose, along with the occurrence of the H-1/H3 and H-1/H-5 NOE effects in the T-ROESY spectrum (Fig. 4b).

Once that all the protons of **D** are defined, the analysis of the HSQC enables the finding of its carbon chemical shifts, making possible the comparison of these values (except the anomeric carbon) with those of the corresponding glycoside taken as reference (Table S1 or (Bock & Pedersen, 1983). Accordingly, it possible to find that all the values ($\mathbf{D_2}$ 74.9 ppm, $\mathbf{D_3}$ 77.0 ppm, $\mathbf{D_4}$ 70.7 ppm, $\mathbf{D_5}$ 66.2 ppm, Table S3) are rather close to those of the reference (C-2 74.0 ppm, C-3 76.9 ppm, C-4 70.4 ppm, C-5 66.3 ppm), indicating that this unit is not substituted to any of its hydroxyl function, namely that it has a terminal location. Finally, the HMBC spectrum (Fig. 4j) discloses that $\mathbf{D_1}$ is correlated to the carbon at 81.6 ppm, namely to another sugar unit since this value is not within those found. This carbon signal is related to **C** as readily inferred by the HSQC-TOCSY spectrum (Fig. 4k).

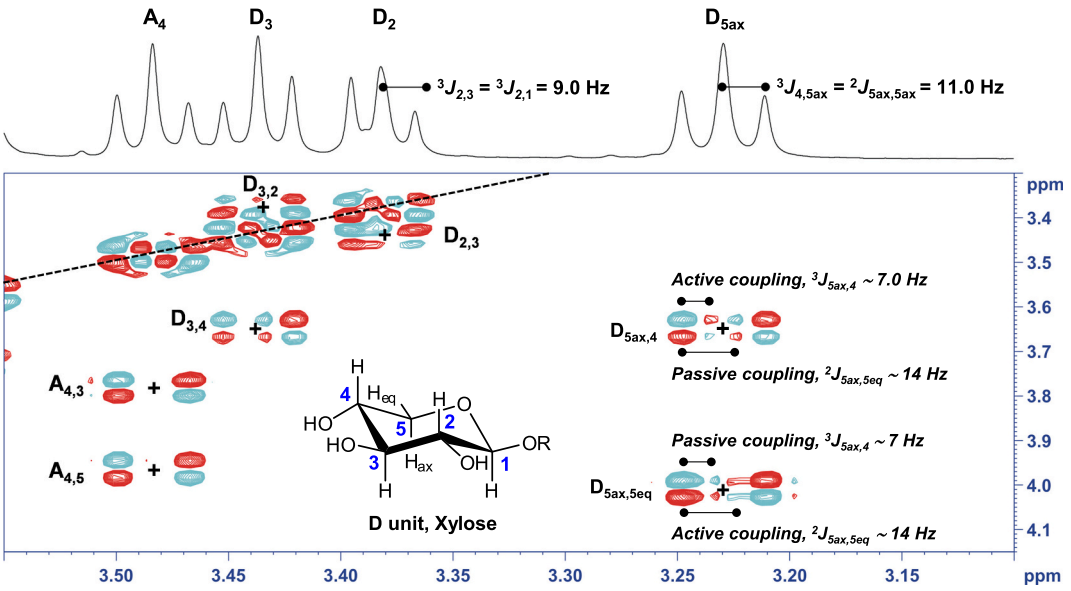

Fig. 7. Expansion of the DQF-COSY spectrum of **1** detailing some of its cross-peaks, with indication of the so-called passive and active couplings along with their distance in Hz. The proton spectrum is reported on the top of the DQF-COSY spectrum together with the coupling constant values read on it. The "+" denotes the centre of the cross-peak, while the broken line indicates the diagonal of the spectrum.

8.2. *Analysis of C unit, a fully branched α-fucose, and of O, its aglycon*

The inspection of the TOCSY from the $\mathbf{C_1}$ signal shows only two correlations along the F1 with one coincident with the COSY cross peak, $\mathbf{C_{1,2}}$ (Fig. 4c) while the number increases to three along F2 (Fig. 4h) because of the improved resolution along this dimension. Then, the fact that $\mathbf{C_1}$ correlates with other three protons is confirmed by looking at the HSQC-TOCSY trace along F1 which reports three different carbon atom densities (Fig. 4k) in addition to that of the anomeric carbon (Fig. S5). Then, based on the information that $\mathbf{C_1}$ correlates with three protons, it is possible to ascribe to this unit the *galacto* configuration. The analysis of the COSY spectrum enables the finding of $\mathbf{C_2}$ and $\mathbf{C_3}$ at 3.99 and 4.08 ppm, respectively (correlations $\mathbf{C_{2,3}}$ and $\mathbf{C_{3,2}}$ in Fig. 4e), while the position of $\mathbf{C_4}$ (3.97 ppm) is deduced from the TOCSY trace along the F2 dimension (Fig. 4h), since this is the only correlation left unattributed between the three found. The carbon chemical shifts of the C-2, C-3 and C-4 positions are defined from the HSQC spectrum, with the help of the HSQC-TOCSY to select the correct $\mathbf{C_4}$ value (81.6 ppm) between the three possible that cross with the proton at ^{1}H 3.97 ppm: 81.6, 70.5 and 70.4 ppm (Fig. 4m).

The HMBC and the COSY spectra enable the complete the attribution of **C**. The first reports a correlation between the proton at 4.13 ppm and the C-1 of **C** (Fig. 4r). The carbon signal of this same proton is at 68.3 ppm (Fig. 4m), namely it is not shifted by glycosylation (see Section 7.2) and for this reason it has to be considered as an *intra*-residue correlation of **C**, namely it is $\mathbf{C_5}$; the shape of this proton as quartet is consistent with the finding in the COSY spectrum that it correlates to a methyl group at 1.31 ppm (Fig. S1). Notably, the position of H-5 is confirmed by the presence of a NOE cross-peak relating H-5 to H-4 (Fig. S3). Taken together, this data complete the attribution of **C** and identify this sugar as a fucose unit, with the α configuration of the anomeric centre due to the $^3J_{H1,H2}$ value (3.6 Hz, Fig. 1). Interestingly, the HMBC spectrum shows for $\mathbf{C_1}$ two *intra*-residue correlations ($\mathbf{C_{1,3}}$ and $\mathbf{C_{1,5}}$, Fig. 4i) and one *inter*-residue with a carbon signal at 69.4 ppm, a hydroxymethylene group assigned to C-1 of the unit **O**. The downfield shift of this hydromethylene group is consistent with being glycosylated; then, the position of the corresponding protons (3.73 and 3.57 ppm) has been deduced from the HMBC correlations, $\mathbf{O_1C_1}$ and $\mathbf{O_{1'}C_1}$ (Fig. 4r), and the analysis of the HSQC-TOCSY spectrum (Fig. S5) shows that $\mathbf{O_1}$ and $\mathbf{O_{1'}}$ protons correlate with a set of carbon signals at high field (32–15 ppm) with the corresponding protons located in the aliphatic region of the proton spectrum. The integration of these protons integrated (Fig. S7), has enabled the evaluation of the length of the lipophylic group. In detail, the integral of the signals at 1.4–1.2 ppm (0.2090 arbitrary units) has been corrected by subtracting the contribution of two methyl groups (approximately 0.0335 each as that of the methyl group at 0.85 ppm), deriving from **C** (mentioned above) and **A** (explained later in the text). The resulting number (0.142) is then divided by the area of one -CH_2- aliphatic group (0.0273, taken from the signal at 1.62 ppm) to give a ratio of 5.20, approximated to 5. As result, the aglycon **O** is a *n*-octanol residue as inferred by counting the number of proton/carbon atoms: the first (69.4 ppm) linked to **C**, followed by six methylene groups (five at 1.4–1.2 ppm and the other at 1.62 ppm) and the methyl group at 0.86 ppm.

Finally, the carbon chemical shifts of $\mathbf{C_2}$ (76.5 ppm) and $\mathbf{C_3}$ (74.7 ppm) are shifted downfield compared to the values of the glycoside used as reference (69.0 and 70.6 ppm, respectively), leading to the conclusion that C is substituted to all the available positions. Indeed, the HMBC spectrum reports two key *inter*-residue correlations: $\mathbf{B_1C_2}$ and $\mathbf{A_1C_3}$ (Fig. 4i), along with their counterparts: $\mathbf{C_2B_1}$ and $\mathbf{C_3A_1}$ (Fig. 4r).

8.3. *B unit, a terminal α-galactose*

As for **B**, the trace of the TOCSY spectrum from the anomeric proton displays only two correlations, with one coincident with the COSY correlation $\mathbf{B_{1,2}}$ (along F1, Fig. 4c) or $\mathbf{B_{2,1}}$ (along F2, Fig. 4h), while the HSQC-TOCSY spectrum from $\mathbf{B_1}$ (Fig. 4k) has only two carbon densities apart from that of the anomeric carbon (Fig. S6). Then, this pattern indicates that this monosaccharide might have the H-3 in the equatorial position, since the magnetization transfer stopped after this position. However, this hypothesis is not correct, and it has to be dismissed because the signal at ^{1}H 3.83 ppm includes two protons as denoted by the presence of two densities in the HSQC spectrum (Fig. 4m), both compatible with **B**, because found in the HSQC-TOCSY from its anomeric proton (Fig. 4k). Based on this finding, the number of TOCSY correlations from the anomeric proton increases to four, suggesting that this unit has the *galacto* configuration, with $\mathbf{B_2}$ and $\mathbf{B_3}$ coincident at 3.83 ppm, and $\mathbf{B_4}$ at 3.97 ppm, as indicated by the corresponding $\mathbf{B_{3,4}}$ (or $\mathbf{B_{4,3}}$) COSY cross-peak. Of note, these two correlations appear weak because their intensity is modulated by the $^3J_{H3,H4}$ coupling constant

that in a *galacto* unit is very small. The equatorial placement of H-4 results in coupling constant with H-5 even smaller than that with H-3, so that the $\mathbf{A_{4,5}}$ correlation is not detected in the COSY, while it may appear in the TOCSY – like in this case – since this spectrum is more sensitive of the other. The attribution of $\mathbf{B_5}$ at 3.90 ppm has been then confirmed from the fact that the corresponding carbon (72.7 ppm, Fig. 4m) matches with one of the *intra*-residue correlations detected in the HMBC spectrum, labelled $\mathbf{B_{1,5}}$ (Fig. 4i). Finally, $\mathbf{B_5}$ is connected to two hydroxymethylene protons at 3.75 and 3.72 ppm by the corresponding correlation in the COSY spectrum (Fig. 4e).

Taking all the information together, **B** is a galactose residue, α configurated at the anomeric centre ($^3J_{H1,H2} = 2.8$ Hz, read on the spectrum in Fig. 5d), and not further substituted because of the similarity between its carbon chemical shifts and those reported for the methylglycoside taken as reference (Table S2 or (Bock & Pedersen, 1983). Notably, this comparison enabled to assign the C-2 and C-3 values of this unit to the densities at 69.4 and 70.6 ppm, respectively.

8.4. Residue A, a terminal α-rhamnose

The anomeric proton of **A** has pattern different from the other units of the oligosaccharide. It has one main correlation in the TOCSY spectrum, coincident with that of the DQF-COSY and assigned to $\mathbf{A_{1,2}}$ (Fig. 4c) or $\mathbf{A_{2,1}}$ (Fig. 4h) and another very weak visible along the F2 dimension of the spectrum (Fig. 4h), while the TOCSY pattern from $\mathbf{A_2}$ (4.02 ppm) indicates the position of all the protons of the unit (Fig. S2), which includes a methyl group at 1.27 ppm, assigned to $\mathbf{A_6}$.

This pattern is typical of *manno* configured monosaccharides and the presence of a methyl group, identifies this residue a rhamnose. For this unit, the reading of the COSY spectrum is straightforward and leads to the finding of all the proton chemical shifts (Fig. 4e), while the HSQC spectrum determines the chemical shifts of the corresponding carbon atoms. Here, it should be noted that the position of $\mathbf{A_5}$ almost overlaps with $\mathbf{B_4}$, however the selection of the right density is supported by the HSQC-TOCSY spectrum, read in correspondence of the $\mathbf{A_4}$ or $\mathbf{A_3}$ densities (Fig. S5). The choice of these signals is advantageous because they do not overlap with any other, so that the correlations that can be seen from them belong to **A** unit alone. The right selection of $\mathbf{A_5}$ density is further confirmed by the finding of the $\mathbf{A_{1,5}}$ correlation in the HMBC spectrum (Fig. 4i).

Then, the carbon chemical shifts values of $\mathbf{A_2}$ (71.4 ppm), $\mathbf{A_3}$ (71.4 ppm), $\mathbf{A_4}$ (73.3 ppm), and $\mathbf{A_5}$ (70.4 ppm) compared with those reported for the α- and the β-methylrhamnosides (Table S2) suggest that the unit is α configured at the anomeric centre. In particular, the C-3 and C-5 values are closer to those of the α-glycoside (71.3 and 69.4 ppm, respectively) that those of the β-glycoside (73.0 and 73.6 ppm, respectively), in agreement with the small value of the $^3J_{H1,H2}$ coupling constant above 1 Hz (1.2 Hz) measured for this unit by enhancing the resolution of the proton spectrum with a GM function (Fig. 5d).

8.5. Final remarks

The analysis of the full set of spectra of the tetrasaccharide **1** has enabled the identification of its structure as reported in Figs. 1, 4, and the scope of this subsection is to briefly remark some points that were not mentioned hitherto.

First, the assignment of a set of spectra is complete only when all the spectra are interpreted and all (or almost all) the densities assigned and found consistent with the structure proposed. In this frame, the interpretation of the T-ROESY spectrum may appear redundant since the same information are gathered from the HMBC, still this analysis is worth the effort because it counterchecks the results of the other. Of note, this HMBC spectrum reported all the key correlations, but this was a fortunate case because it does not always happen.

Regarding the T-ROESY spectrum of the tetrasaccharide, it has been fully assigned (including the carbinolic region), and the NOEs detected (Fig. 4a,b) were consistent with the HMBC data; this spectrum has not been discussed to limit the length of the manuscript. Of note, the NOE effects reflect the spatial proximity between two protons and for this reason they do not necessarily indicate where the sugar is substituted, so that their interpretation needs care. This warning is well exemplified by the correlation $\mathbf{C_1B_5}$ (Fig. 4a). The residue **C** is linked to the octyl unit and certainly not to O-5 of **B**, for a couple of reasons: 1) this hydroxyl function is involved in the cyclization of the unit as pyranose so it cannot link another sugar, and 2) the unit **B** is terminal, so it does not have any substituent. Then, the reason of this NOE grounds on the conformation adopted from the oligosaccharide that keeps the anomeric proton of **C** close enough to H-5 of **B** for this NOE effect to be seen. The only way to understand this and eventually other unexpected NOE effects necessitates a thorough molecular modelling study of the molecule, which for an extended version of this oligosaccharide is reported (De Castro et al., 2018). As for the HMBC spectrum, the $^1J_{C1H1}$ have been read from the spectrum (Table S3) and found consistent with the anomeric configurations of the residues.

Second, this NMR analysis has not considered the absolute configuration of the different units. This parameter can be in principle afforded by NMR if the configuration of one of the units is known. Otherwise, it is possible to assign an arbitrary configuration to one of the units and to deduce all the others relatively to that. The approaches at this stage are mostly two. The first consists in the comparison of the experimental *inter*-residues NOE effects (or of the distances that can be derived from them) with those calculated from molecular modelling of the molecule, as recently performed for a peculiar polymer alternating glycosidic and peptide linkages (Speciale, Di Lorenzo, et al., 2020). The other approach instead relies on the quantification of the α- and β-glycosylation effects of the two units joined by the glycosidic linkage, as mentioned in Section 7.2 and based on the method developed from Shashkov et al. (1988) and Lipkind et al. (1988)

However, both these approaches should be considered as a last resort when it is not possible to determine the absolute configuration through established experimental methods, since the interpretation of the results is not always crystal clear, and it may lead to erroneous conclusions.

Then, the bottom line of any NMR assignation is that the attributions are further strengthened when integrated with the chemical analyses specific for: 1) the determination of the absolute configuration of the sugars, namely by measuring the optical rotation of the monosaccharide isolated in the pure form from the sample or by the less demanding gas chromatographic approaches (De Castro et al., 2010; Gerwig et al., 1978; Leontein et al., 1978), and 2) by determining their linkage pattern by the consolidated approaches (De Castro et al., 2010; Sims et al., 2018).

Third, the approach used to assign the NMR signals of **1** is of general applicability and it can be employed for any type of glycan, as those from natural sources as well as those afforded by chemical manipulation. Independently from the source of the material, polysaccharides can present several challenges and the most common are summarized hereafter along with some possible solutions.

The simplest case is that of regular polysaccharide, namely a polymer that consists of multiple copies of the same repeating unit assembled up to reach a certain polymerization degree. In this case, the NMR spectra display only the signals of the repeating unit (for instance three sugars if this is the size of the repeat) since all the repeating units are equivalent for NMR. Then, spectra analysis determines the nature of the repeat, and when this information crossed with the molecular weight of the glycan, obtained by other approaches, it can disclose the average degree of polymerization of the polysaccharide. The only potential risk for the analysis of such polysaccharides is given by their tendency to present broad signals, and the strategy to address this potential problem is discussed at the end of this section.

Differently from the case above, many polysaccharides do not possess a regular repeating unit, as pectins or yeast mannans and many others. Still, a thorough NMR analysis can unveil many important

features of the sample, i.e., the nature of the monosaccharide residues, how they are linked and which residue they are linked by and to. Key to complete the picture of an unregular polysaccharide is the definition of the ratio between the different units, or the different motifs that have been found, which can be done by integration of the appropriate NMR signals.

To this end, the integration can be performed by using the proton spectrum if the signals of interest do not overlap with others, or it can be done by using the HSQC spectrum and integrating – or measuring the volume – of the densities. The HSQC has the unquestionable advantage that the cross peaks are more spread due to the large interval of frequencies of ^{13}C nucleus. For this reason, HSQC integration is finding its application in metabolomic (Puig-Castellví et al., 2018) and it can be applied in the study of non-regular glycans. For instance, the HSQC has been used in this quantitative mode to estimate the many different structural motifs of heparin from different sources, either natural or semisynthetic (Guerrini et al., 2005; Mauri et al., 2017).

Similarly, the proportions between the different ribitol-phosphates motifs that build the teichoic acid of a selected strain of *Staphylococcus aureus* along some of its mutant have been established by comparing the densities of the carbinolic atoms of different ribitol units (Gerlach et al., 2018).

The last and probably most serious issue posed from polysaccharides is given by the broadening of their NMR signals which in the worst cases determines their total disappearance in the spectrum. This problem arises because an inverse correlation between the molecular weight of any molecule and the associated tumbling rate exists, so that the slowest is the tumbling the worst is the resolution observed in liquid-state NMR. The MW of polysaccharide is in the range of several kDa, which negatively impact on the resolution of their spectra; this latter is often worsened by the augmented viscosity of the solution due to self-aggregation processes, as for hyaluronic acid and pectins.

Yet, in most of the cases it is still possible to measure NMR spectra of reasonable quality by a careful preparation of the sample and by increasing the temperature used when measuring the NMR experiment. As for the preparation of the sample (see Section 3), often the pH variation (either toward acidic or alkaline pH) can be sufficient to improve the resolution whereas other tricks consist in the removal of cations, especially bivalent cations, through the usage of resins able to exchange ions or with the addition of deuterated EDTA.

When any of these remedies fails or leads to unsatisfactory results, the last resort is to depolymerize the sample to decrease its molecular weight. Enzymatic depolymerization is generally the best choice due to the regioselectivity of hydrolases (Stone et al., 2008), or – to a reduced extent – of the lytic polysaccharide monooxigenases (Frandsen & Lo Leggio, 2016). Regretfully, the usage of enzymes cannot be always pursued because it is not always possible to find the right the enzymes for a given polysaccharide.

It is also possible to opt for a selective chemical degradation (Knirel et al., 2019), as the following given as example: nitrous acid degradation of sugar presenting a free amino group (Conrad, 2001); Smith degradation, a three step procedure based on the oxidative cleavage of vicinal diols (Abdel-Akher et al., 1952), the β-elimination of 4-substituted uronic acids promoted by temperature and strong alkaline pH (BeMiller & Kumari, 1972) and the so-called peeling reaction: a recursive degradation of glycans whose backbone presents 1,3-linkages and the reducing end in the free form (Ponder & Richards, 1997).

Moreover, it is possible to depolymerize a polysaccharide by a mild acid hydrolysis to preferentially cleave the glycosidic linkages that are more labile, such as furanose rings or deoxysugars. Despite this approach is generally applicable, it might not be sufficiently selective with the risk to afford complex mixtures of oligosaccharides.

Hence, scientists facing the challenge of the NMR study of a polysaccharide have many arrows in their bow to improve the quality of their NMR spectra and the choice of the approach to follow depends very much by searching the literature in the field and in part by the personal experience of the NMR operator.

9. Conclusions

This review highlights some technical aspects of the NMR experiments recording along with some artifacts that may trouble the spectra, and it gives direction about how data should be presented and – more importantly – interpreted to make a reasonable guess about the nature of the glycan.

When spectra are to be submitted for a scientific publication, we encourage the authors to follow this short checklist.

- experimental conditions for acquisition should be detailed enough to enable other laboratories to replicate the same experiments. Conditions include: NMR field strength, solvent used, pH, temperature, concentration of the sample, calibration signal, number of scans, number of points in acquisition, type of sequence(s) used.
- data presentation: clarity above all. The size of the figures should be as large as possible compatibly with the space allowed from the journal. Full size spectra can be given as supporting, while the main part of the manuscript should report the area of the spectra that are more informative. Spectra can be expanded or fragmented in slices to focus the attention of the readers on the most relevant signals used for the attribution process (as in Figs. 3 and 4). Labelling of the peaks or of the densities (possibly all) should be simple and made of few letters/numbers, so that the signals remain visible. Finally, the spectra should be overlapped in order to maximize the information that can be gathered. The readers can refer to the presentation strategy in place for Figs. 3 and 4.
- NMR chemical shift reporting. This should be done essentially in a table (as Table S3), where the identity of each sugar unit, inclusive of its anomeric configuration and linkage pattern, should be reported along with the label used during the assignment. The same label should be placed in the structure of the glycan next to the corresponding monosaccharide unit, and it should be used to mark the corresponding peak/density of the spectrum together with the indication of the position of nucleus under consideration. Proton NMR chemical shifts should be reported with two decimals, while carbon chemical shifts with only one.
- consistency check: this can be done in several steps. First, spectra must be fully assigned, and all the densities (true signal or artifact) should be explained. Then, the information from one spectrum should be consistent with those from another. For instance, the HMBC correlations should fit with those from the NOESY/ROESY spectra, or the COSY chemical shifts of a certain residue should be in the pool of those detected in the TOCSY spectrum. The carbon chemical shifts of each residue should be compared to those of the corresponding methylglycosides taken as reference (Tables S1, 2) and any difference should be consistent with the substitution pattern found for the unit. Finally, once the attribution process is finished, the results can be eventually compared with those reported in literature (and not the way around) or by consulting those of suitable databases, as CSDB already mentioned in this review (Toukach & Egorova, 2015).
- what should be avoided: 1) the usage of spectra with low signal-to-noise ratio and/or unproperly transformed; 2) the presentation of unlabelled (or poorly labelled) spectra, because this may raise doubts on the consistency of the interpretation process; 3) the identification of a unit based on the matching of a spare number of chemical shift (s). The chemical shift matching must involve all the positions of the sugar, and this happens only when the structural environment of the unit under exam is very similar to that of the reference.

Concluding, we hope that this tutorial review might help scientists engaged in the challenging task of carbohydrate/glycans NMR analysis and/or that it may attract others in this area, facilitating their learning

curve by the adoption of the guidelines and of the resources here collected and that are necessary to tackle this problem.

CRediT authorship contribution statement

C.D.C. conceived and wrote the review, all the other authors contributed with particular parts and by reading, editing and approving.

Declaration of competing interest

The authors declare no competing interest.

Acknowledgments

C.D.C and A.M. gratefully acknowledge H2020 MSCA ITN 2018 SWEETCROSSTALK grant No 814102P. A.S. gratefully acknowledges PRIN 2017 (2017XZ2ZBK, 2019–2022), and H2020-MSCA-ITN-2020-GLYTUNES– grant agreement 956758. P.G-V. fellowship has been supported by the Train2Target project granted from the European Union's Horizon 2020 framework program for research and innovation (Project #721484). This research was carried out also in the frame of Programme STAR, financially supported by UNINA and Compagnia di San Paolo as acknowledged by F.D.L, from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 851356 to R.M.

Appendix A. Supplementary data

The supplementary material includes: the ^{13}C chemical shifts of pentose (Table S1) and hexose (Table S2) units, and the $^{1}H/^{13}C$ values of the tetrasaccharide **1** (Table S3), and it can be found online at doi: https://doi.org/10.1016/j.carbpol.2021.118885.

References

Abdel-Akher, M., Hamilton, J. K., Montgomery, R., & Smith, F. (1952). A new procedure for the determination of the fine structure of polysaccharides. *Journal of the American Chemical Society, 74*(19), 4970–4971.

Agrawal, P. K. (1992). NMR spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry, 31*(10), 3307–3330.

BeMiller, J. N., & Kumari, G. V. (1972). Beta-elimination in uronic acids: Evidence for an ElcB mechanism. *Carbohydrate Research, 25*(2), 419–428.

Bock, K., & Pedersen, C. (1983). Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. In R. S. Tipson, & D. Horton (Eds.), *Advances in carbohydrate chemistry and biochemistry* (pp. 27–66). Academic Press.

Böhm, M., Bohne-Lang, A., Frank, M., Loss, A., Rojas-Macias, M. A., & Lütteke, T. (2018). Glycosciences.DB: An annotated data collection linking glycomics and proteomics data (2018 update). *Nucleic Acids Research, 47*(D1), D1195–D1201.

Cavalier-Smith, T. (2006). Rooting the tree of life by transition analyses. *Biology Direct, 1*, 19.

Claridge, T. D. W. (2016a). Chapter 3 - Practical aspects of high-resolution NMR. In T. D. W. Claridge (Ed.), *High-resolution NMR techniques in organic chemistry* (3rd ed., pp. 61–132). Boston: Elsevier.

Claridge, T. D. W. (2016b). Chapter 6 - Correlations through the chemical bond I: Homonuclear shift correlation. In T. D. W. Claridge (Ed.), *High-resolution NMR techniques in organic chemistry* (3rd ed., pp. 203–241). Boston: Elsevier.

Claridge, T. D. W. (2016c). Chapter 9 - Correlations through space: The nuclear overhauser effect. In T. D. W. Claridge (Ed.), *High-resolution NMR techniques in organic chemistry* (3rd ed., pp. 315–380). Boston: Elsevier.

Conrad, H. E. (2001). *Nitrous acid degradation of glycosaminoglycans*.

De Castro, C., Klose, T., Speciale, I., Lanzetta, R., Molinaro, A., Van Etten, J. L., & Rossmann, M. G. (2018). In , *115*. *Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches* (pp. E44–E52).

De Castro, C., Parrilli, M., Holst, O., & Molinaro, A. (2010). Chapter five - microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. In M. Fukuda (Ed.), *Methods in enzymology* (pp. 89–115). Academic Press.

Di Carluccio, C., Forgione, M. C., Martini, S., Berti, F., Molinaro, A., Marchetti, R., & Silipo, A. (2021). Investigation of protein-ligand complexes by ligand-based NMR methods. *Carbohydrate Research, 503*, Article 108313.

Di Lorenzo, F., Duda, K. A., Lanzetta, R., Silipo, A., De Castro, C., & Molinaro, A. (2021). A journey from structure to function of bacterial lipopolysaccharides. *Chemical Reviews*. https://doi.org/10.1021/acs.chemrev.0c01321. In press.

Eichler, J. (2013). Extreme sweetness: Protein glycosylation in archaea. *Nature Reviews Microbiology, 11*(3), 151–156.

Frandsen, K. E. H., & Lo Leggio, L. (2016). Lytic polysaccharide monooxygenases: A crystallographer's view on a new class of biomass-degrading enzymes. *IUCrJ, 3*(Pt 6), 448–467.

Garegg, P. J., Jansson, P.-E., Lindberg, B., Lindh, F., Lönngren, J., Kvarnström, I., & Nimmich, W. (1980). Configuration of the acetal carbon atom of pyruvic acid acetals in some bacterial polysaccharides. *Carbohydrate Research, 78*(1), 127–132.

Gerlach, D., Guo, Y., De Castro, C., Kim, S.-H., Schlatterer, K., Xu, F.-F., & Peschel, A. (2018). Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity. *Nature, 563*(7733), 705–709.

Gerwig, G. J., Kamerling, J. P., & Vliegenthart, J. F. G. (1978). Determination of the d and l configuration of neutral monosaccharides by high-resolution capillary g.L.C. *Carbohydrate Research, 62*(2), 349–357.

Gimeno, A., Valverde, P., Ardá, A., & Jiménez-Barbero, J. (2020). Glycan structures and their interactions with proteins. A NMR view. *Current Opinion in Structural Biology, 62*, 22–30.

Gorin, P. A. J., Mazurek, M., Duarte, H. S., Iacomini, M., & Duarte, J. H. (1982). Properties of 13C-n.M.R. Spectra of O-(1-carboxyethylidene) derivatives of methyl β-d-galactopyranoside: Models for determination of pyruvic acid acetal structures in polysaccharides. *Carbohydrate Research, 100*(1), 1–15.

Gottlieb, H. E., Kotlyar, V., & Nudelman, A. (1997). NMR chemical shifts of common laboratory solvents as trace impurities. *The Journal of Organic Chemistry, 62*(21), 7512–7515.

Guerrini, M., Naggi, A., Guglieri, S., Santarsiero, R., & Torri, G. (2005). Complex glycosaminoglycans: Profiling substitution patterns by two-dimensional nuclear magnetic resonance spectroscopy. *Analytical Biochemistry, 337*(1), 35–47.

Hager, F. F., Sützl, L., Stefanović, C., Blaukopf, M., & Schäffer, C. (2019). Pyruvate substitutions on glycoconjugates. *International Journal of Molecular Sciences, 20*(19), 4929.

Jansson, P.-E., Lindberg, J., & Widmalm, G. (1993). Syntheses and NMR studies of pyruvic acid 4,6-acetals of some methyl hexopyranoside. *Acta Chemica Scandinavica, 47*, 711–715.

Knirel, Y. A., Naumenko, O. I., Senchenkova, S.y. N., & Perepelov, A. V. (2019). Chemical methods for selective cleavage of glycosidic bonds in the structural analysis of bacterial polysaccharides. *Russian Chemical Reviews, 88*, 19.

Kupče, Ē., & Claridge, T. D. W. (2018). Molecular structure from a single NMR supersequence. *Chemical Communications, 54*(52), 7139–7142.

Leontein, K., Lindberg, B., & Lönngren, J. (1978). Assignment of absolute configuration of sugars by g.L.C. Of their acetylated glycosides formed from chiral alcohols. *Carbohydrate Research, 62*(2), 359–362.

Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V., & Kochetkov, N. K. (1988). A computer-assisted structural analysis of regular polysaccharides on the basis of 13C-n.M.R. Data. *Carbohydrate Research, 175*(1), 59–75.

Marchetti, R., Forgione, R. E., Fabregat, F. N., Di Carluccio, C., Molinaro, A., & Silipo, A. (2021). Solving the structural puzzle of bacterial glycome. *Current Opinion in Structural Biology, 68*, 74–83.

Mauri, L., Boccardi, G., Torri, G., Karfunkle, M., Macchi, E., Muzi, L., & Guerrini, M. (2017). Qualification of HSQC methods for quantitative composition of heparin and low molecular weight heparins. *Journal of Pharmaceutical and Biomedical Analysis, 136*, 92–105.

Molinaro, A., Holst, O., Di Lorenzo, F., Callaghan, M., Nurisso, A., D'Errico, G., & Martín-Santamaría, S. (2015). Chemistry of lipid A: At the heart of innate immunity. *Chemistry, A European Journal, 21*(2), 500–519.

Mort, A., Bell-Eunice, G., & Wu, X. (2013). Characterization of a methyl-esterified tetragalacturonide fragment isolated from a commercial pectin with a medium degree of methyl-esterification. *Carbohydrate Research, 380*, 108–111.

Pedersen, C. P., Prestel, A., & Teilum, K. (2021). Software for reconstruction of nonuniformly sampled NMR data. *Magnetic Resonance in Chemistry, 59*(3), 315–323.

Ponder, G. R., & Richards, G. N. (1997). Arabinogalactan from Western larch, part III: Alkaline degradation revisited, with novel conclusions on molecular structure. *Carbohydrate Polymers, 34*(4), 251–261.

Puig-Castellví, F., Pérez, Y., Piña, B., Tauler, R., & Alfonso, I. (2018). Comparative analysis of 1H NMR and 1H–13C HSQC NMR metabolomics to understand the effects of medium composition in yeast growth. *Analytical Chemistry, 90*(21), 12422–12430.

Reynolds, W. F., & Enríquez, R. G. (2002). Choosing the best pulse sequences, acquisition parameters, postacquisition processing strategies, and probes for natural product structure elucidation by NMR spectroscopy. *Journal of Natural Products, 65*(2), 221–244.

Rohde, M. (2019). The Gram-positive bacterial cell wall. *Microbiology Spectrum, 7*(3).

Shashkov, A. S., Lipkind, G. M., Knirel, Y. A., & Kochetkov, N. K. (1988). Stereochemical factors determining the effects of glycosylation on the 13C chemical shifts in carbohydrates. *Magnetic Resonance in Chemistry, 26*(9), 735–747.

Sims, I. M., Carnachan, S. M., Bell, T. J., & Hinkley, S. F. R. (2018). Methylation analysis of polysaccharides: Technical advice. *Carbohydrate Polymers, 188*, 1–7.

Speciale, I., Di Lorenzo, F., Gargiulo, V., Erbs, G., Newman, M.-A., Molinaro, A., & De Castro, C. (2020). In , *59*. *Biopolymer skeleton produced by Rhizobium radiobacter: Stoichiometric alternation of glycosidic and amidic bonds in the lipopolysaccharide O-antigen* (pp. 6368–6374) (16).

Speciale, I., Laugieri, M. E., Noel, E., Lin, S., Lowary, T. L., Molinaro, A., & De Castro, C. (2020). Chlorovirus PBCV-1 protein A064R has three of the transferase activities necessary to synthesize its capsid protein N-linked glycans. *Proceedings of the National Academy of Sciences, 117*(46), 28735–28742.

Stenutz, R., Weintraub, A., & Widmalm, G. (2006). The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiology Reviews, 30*(3), 382–403.

Stone, B. A., Svensson, B., Collins, M. E., & Rastall, R. A. (2008). Polysaccharide degradation. In B. O. Fraser-Reid, K. Tatsuta, & J. Thiem (Eds.), *Glycoscience: Chemistry and chemical biology* (pp. 2325–2375). Berlin, Heidelberg: Springer, Berlin Heidelberg.

Theocharis, A. D., Skandalis, S. S., Gialeli, C., & Karamanos, N. K. (2016). Extracellular matrix structure. *Advanced Drug Delivery Reviews, 97*, 4–27.

Toukach, P. V., & Egorova, K. S. (2015). Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic Acids Research, 44*(D1), D1229–D1236.

Turner, C. J., Connolly, P. J., & Stern, A. S. (1999). Artifacts in sensitivity-enhanced HSQC. *Journal of Magnetic Resonance, 137*(1), 281–284.

Tvaroska, I., & Taravel, F. R. (1995). Carbon-proton coupling constants in the conformational analysis of sugar molecules. In D. Horton (Ed.), *Advances in Carbohydrate Chemistry and Biochemistry* (pp. 15–61). Academic Press.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., & Markley, J. L. (2007). BioMagResBank. *Nucleic Acids Research, 36*(suppl_1), D402–D408.

Widmalm, G. (2021). 1.14 - General NMR spectroscopy of carbohydrates and conformational analysis in solution☆. In J. J. Barchi (Ed.), *Comprehensive glycoscience* (2nd ed., pp. 340–373). Oxford: Elsevier.