Polysaccharide Sequencing

Petra Mischnick*

Chemical compounds whose names begin with the prefix "oligo" or "poly" broaden questions of structure to include aspects of sequencing. At the constitutional level, natural products that serve as skeletal compounds or energy depots usually demand little more than orderly building plans with which, for example, the requirements for mechanical rigidity or osmotic inactivity can be fulfilled. Instances of this are the polysaccharides cellulose and starch, which are composed solely of glucose. A different situation appears if phenomena such as molecular recognition or information storage come into play. Here the different monomers serve as an alphabet with which exact instructions can be written by means of an inexhaustible combination of letters, or as modular systems for distinctive macromolecular structures that recognize receptors or substrates with high selectivity. Prime examples of this are nucleic acids and proteins whose sequencing has long been established. On the other hand, work published recently in *Science* by Sasisekharan and co-workers under the title "Sequencing Complex Polysaccharides" has attracted attention.[1] There are several reasons for this: The polyfunctionality of carbohydrates allows more diverse coupling possibilities than is the case with nucleotides and amino acids, and includes branching such as additional substitutions. Furthermore, the biological significance of carbohydrates for, for example, cell recognition and adhesion, inflammatory processes, or the regulation of enzymatic reactions, and hence also the analytical challenge, has been recognized only relatively recently.

What makes sequence analysis of oligo- and polysaccharides so difficult? To start with, one is almost always confronted with structural diversity. N- and O-glycans of glycoproteins differ from each other according to species, tissue, cell type, and protein. In the biosynthesis of the glycosaminoglycans heparin and heparan sulfate the situation is further complicated by polymer-like, enzymatic modification of (D-GlcAβ1,4-D-GlcNAcα1,4-)$_n$, which leads to the partial isomerization of the repeating units to (L-IdoAα1,4-D-GlcNAcα1,4-)$_n$, to N-deacetylation/N-sulfation, and to different O-sulfation patterns (Figure 1).[2] Likewise, the polyfunctionality of carbohydrates also hampers stepwise chemical degradation from one chain terminus, as is the case for the Edman degradation of peptides. Moreover, amplification of certain structures, as with the polymerase chain reaction (PCR) in the case of nucleic acids, cannot be carried out with polysaccharides.

Initial advances were made with the development of "soft" ionization methods in mass spectrometry, first with FAB, and then later with MALDI-MS and ESI-MS (FAB = fast atom bombardment, MALDI = matrix assisted laser desorption/ionization, ESI = electrospray ionization). In principle, these methods make it possible to determine the molecular mass as well as the order of the sugar building blocks, which may be identified from the sequential cleavage of glycoside bonds.[3] However, because of the large number of possible stereoisomers, it is not possible to decide which hexose is involved, or which deoxy or amino sugar, as long as the building block composition does not impose any restrictions. Furthermore, branching-specific fragment ions, which provide further structural information, are formed by high-energy collision activation.[4] With all these ionization methods carbohydrates bring up the rear since they are ionized in much lower yield than peptides or oligonucleotides. Fundamental to the successful sequencing strategy of Sasisekharan et al.[1] was therefore the procedure of Juhasz and Biemann[5] in which an increase in the sensitivity of MALDI-TOF-MS into the 100-fmol range was achieved by complexation of the sulfated heparin-like oligosaccharides with basic proteins of defined molecular mass, and sulfate loss could be avoided.

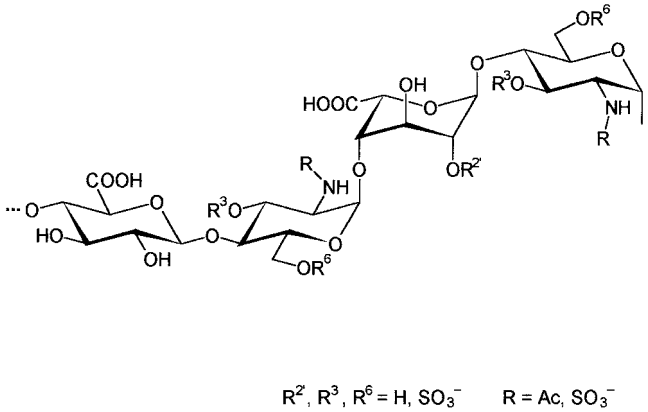

Figure 1. Possible section of a heparin or a heparan sulfate molecule (D-GlcAβ1,4-D-GlcNHRα1,4-L-IdoAα1,4-D-GlcNHRα1,4-). GlcA = glucuronic acid, $GlcNH_2$ = 2-amino-2-deoxyglucose, IdoA = iduronic acid.

Because of their manifold biological functions and high structural diversification, O- and N-glycans and heparin/heparan sulfate are very frequently both topic and stimulus for powerful, sensitive, and practical carbohydrate analysis. Glycans from glycoproteins have been sequenced by their fragmentation behavior under CID conditions (CID = collision-induced fragmentation) by means of MALDI-MS[6] as well as through the use of metastable ions (post source decay, PSD-MALDI-MS),[7] or after stepwise enzymatic degradation with exoglycosidases.[9–11] It has also been possible to obtain new information on the complexity of saccharides from human milk by this strategy.[12]

Which are now the key steps in the new sequence strategy described above? The authors investigated hexa- to decasaccharides with heparin-like structures. The starting point for

[*] Prof. Dr. P. Mischnick
Technische Universität Braunschweig
Institut für Lebensmittelchemie
Schleinitzstrasse 20, 38106 Braunschweig (Germany)
Fax: (+49) 531-391-7230
E-mail: p.mischnick@tu-bs.de

  0570-0833/00/3907-1222 $ 17.50+.50/0

Table 1. Sequencing example from ref. [1]

	Example 1	n[a]
a)	MALDI-MS: $M = 2230 \rightarrow$octasaccharide-$S_{11}Ac_0$	
b)	exhaustive enzymatic degradation to disaccharides with heparinases: CE $\rightarrow \Delta U_{2S}H_{NS,6S}$, $\Delta UH_{NS,6S}$ (3:1)[b]	32
c_1)	heparinase I: MALDI-MS $\rightarrow$tetrasaccharide-S_5, disaccharide-S_3	4
c_2)	heparinase II: MALDI-MS $\rightarrow$hexasaccharide-$S_9 \rightarrow \Delta U_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}GH_{NS,6S}$[b]	1
	Example 2	
a)	MALDI-MS: $\rightarrow$decasaccharide-$S_{13}Ac_1$	
b)	exhaustive enzymatic degradation to disaccharides with heparinases: CE$\rightarrow \Delta U_{2S}H_{NS,6S}$, $\Delta UH_{NAc,6S}$, $\Delta UH_{NS,3S,6S}$(3:1:1)	320
c_1)	heparinase I ("mild"): $\rightarrow$MALDI-MS	52
	heparinase I ("exhaustive"): $\rightarrow$MALDI-MS	28
d)	labeling of the decasaccharide at the reducing terminus	
e_1)	heparinase II: $\rightarrow$MALDI-MS	6
e_2)	exhaustive degradation by deamination: $\rightarrow$MALDI-MS$\rightarrow$localization of H_{NAc}	6
f)	successive treatment with exoenzymes$\rightarrow \Delta U_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}IH_{NAc,6S}GH_{NS,3S,6S}$	1

[a] n=calculated number of sequences which are in agreement with the experimental data. [b] U = uronic acid (I or G), ΔU = 4,5-unsaturated uronic acid from enzymatic cleavage, I = iduronic acid, G = glucuronic, H = hexosamine, index: 2S sulfated at O2, NS: sulfated at N, etc.

their deliberations are the possible sequences that can be constructed from the disaccharide building blocks (Figure 1), which consist of an uronic acid and a hexosamine. When the possible sulfation and N-acetylation patterns are taken into consideration, there are 32 building blocks which are coded for the computer-aided calculations. This is absolutely necessary since even for one octasaccharide constructed from these units there are theoretically more than a million possible structures (32^4). Exhaustive degradation with heparinases gives the building blocks, the composition of which may be analyzed by capillary electrophoresis (CE) and comparison with standard substances. Determination of the molecular mass by MALDI-MS supplies information on the possible combinations of the coded disaccharide units. A clear limitation is now possible on the basis of the variation in masses in relation to the degree of sulfation and acetylation together with knowledge gained from the building block analysis. By selective chemical or enzymatic partial degradation new fragments are obtained in each case, whose molecular masses are also determined by MALDI-MS. In this way the possibilities of suitable precursors are restricted stepwise. Commercially available exoenzymes have been used for identifying the configuration of the uronic acid at the nonreducing terminus. The authors' procedure is illustrated in Table 1 with the example of an octasaccharide. After nitrous acid deamination and ring contraction to 2,5-anhydromannose, chain-cleaving deamination, commenly used for 2-amino-2-desoxy sugars, allows localization of the N-sulfate and N-acetyl groups in the hexosamine. Labeling of the reducing terminus permits its differentiation from the internal building blocks. Successful elucidation of the whole sequence is made possible by selection of suitable degradation methods and mass spectrometric analysis with verification through calculations which of the coded disaccharide structures are in agreement with the experimental results.

That the detailed structural elucidation of heparin and heparan sulfate is highly topical and is being intensively investigated is demonstrated by two further papers which were published in 1999 and which herald "Sequence analysis…"[13] or "A strategy for rapid sequencing…"[14] in their titles. The chemical (deaminative cleavage) and enzymatic methods are the same as in the *Science* work cited above,[1] but in the first case[13] the analysis of the degradation products is carried out by anion-exchange chromatography of the tritium-labeled oligosaccharides, and in the second example[14] by gel electrophoresis after reductive amination with a fluorophore. Neither study involves mass spectrometry.

All the applications described involve isolated, chemically uniform hexa- and decasaccharides. Isolates of natural oligosaccharides are frequently microheterogeneous and must normally be carefully purified first. Even for Sasisekharan and co-workers, building block analysis by capillary electrophoresis is essential for the restriction of possible sequences. Wherein lies the progress of this strategy? According to the statements of the authors, MALDI-MS does not require renewed isolation and purification of the cleavage products after the different manipulations. Moreover, they see in their approach potential for a fully automated sequencing method, when with better investigated and more specific enzymes the tools for a defined partial degradation of complex polysaccharides are available and thus a more refined and more efficient algorithm can be developed. This is all the more important if the future goal is to analyze larger sequences and mixtures of oligosaccharides in a similar manner and thus reduce the effort of fractionation of biological material, a challenge whose accomplishment is awaited with anticipation.

[1] G. Venkataraman, Z. Shriver, R. Raman, R. Sasisekharan, *Science* **1999**, *286*, 537 – 542.
[2] M. Maccarana, Y. Sakura, A. Tawada, K. Yoshida, U. Lindahl, *J. Biol. Chem.* **1996**, *271*, 17804 – 17810.
[3] H. Egge, J. Peter-Katalinic, *Mass Spectrom. Rev.* **1987**, *6*, 331 – 393.
[4] V. N. Reinhold, B. B. Reinhold, C. E. Costello, *Anal. Chem.* **1995**, *67*, 1772 – 1784.
[5] P. Juhasz, K. Biemann, *Carbohydr. Res.* **1995**, *270*, 131 – 147.
[6] Y. Mechref, A. G. Baker, M. V. Novotny, *Carbohydr. Res.* **1998**, *313*, 145 – 155.
[7] J. Lemoine, F. Chirat, B. Domon, *J. Mass Spectrom.* **1996**, *31*, 908 – 912.
[8] N. Viseux, C. E. Costello, B. Domon, *J. Mass Spectrom.* **1999**, *34*, 364 – 376.
[9] C. W. Sutton, J. A. O'Neill, J. S. Cottrell, *Anal. Biochem.* **1994**, *218*, 34 – 46.

[10] B. Küster, T. J. P. Naven, D. J. Harvey, *J. Mass Spectrom.* **1996**, *31*, 1131 – 1140.
[11] Y. Mechref, M. V. Novotny, *Anal. Chem.* **1998**, *70*, 455 – 463.
[12] B. Finke, B. Stahl, A. Pfenninger, M. Karas, H. Daniel, G. Sawatzki, *Anal. Chem.* **1999**, *71*, 3755 – 3762.
[13] R. R. Vivès, D. A. Pye, M. Salmivirta, J. J. Hopwood, U. Lindahl, J. T. Gallagher, *Biochem. J.* **1999**, *339*, 767 – 773.
[14] J. E. Turnbull, J. J. Hopwood, J. T. Gallagher, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2698 – 2703.

Direct Detection of Hydrogen Bonds in Biopolymers by NMR Spectroscopy

Gerd Gemmecker*

Dedicated to Professor Horst Kessler on the occasion of his 60th birthday

Today every chemist knows about the existence and importance of hydrogen bonds. The concept of hydrogen bonding plays a major role in the understanding of the physicochemical properties of water, inorganic and organic acids, as well as acid – base catalysis. Hydrogen bonds influence the conformational preferences and reactivity of many classes of organic molecules.

In the field of biochemistry, hydrogen bonds are responsible for the selectivity of base pairing in nucleic acids—the basis for the preservation and propagation of genetic information. Furthermore, hydrogen bonds play a crucial role in the three-dimensional folding of proteins: the most prominent structural features of proteins—helical and β-sheet regions—are defined and stabilized by their unique pattern of hydrogen bonds.

Finally, intermolecular hydrogen bonds contribute to the affinity and selectivity of molecular recognition, from simple host – guest systems all the way up to the multi-component complexes of proteins, nucleic acids, etc., which are at the heart of so many biological functions.

However, in the course of structure determinations by X-ray diffraction or NMR spectroscopy, the existence of hydrogen bonds is usually only inferred in a rather indirect way from an appropriate spatial arrangement of possible hydrogen donors and acceptors that is already known to exist in high-resolution molecular structures. Particularly in the field of NMR spectroscopy, several approaches have been proposed to deduce the existence and localization of hydrogen bonds indirectly from parameters such as hydrogen-exchange rates, isotropic and anisotropic chemical shifts, ^{2}H quadrupolar splittings, and ^{1}H/^{2}H isotope shifts.

At the XVIIIth International Conference on Magnetic Resonance in Biological Systems (ICMRBS) in Tokyo in August 1998 Andrew Dingley and Stephan Grzesiek reported the observation of scalar couplings across the hydrogen bonds in the Watson – Crick base pairs of a ^{15}N-labeled 69 nucleotide RNA(their paper appeared in print the same month).[1] In an experiment set up for the detection of scalar couplings between two ^{15}N spins (a type of ^{15}N-COSY spectrum), several fairly intense cross-peaks had been observed (Figure 1) that could only be assigned to correlations between N3 of a uridine and N1 of an adenosine group, and N1 of a guanosine and N3 of a cytosine group, each across the hydrogen bonds within a U-A or G-C Watson – Crick base pair, respectively (Figure 2).

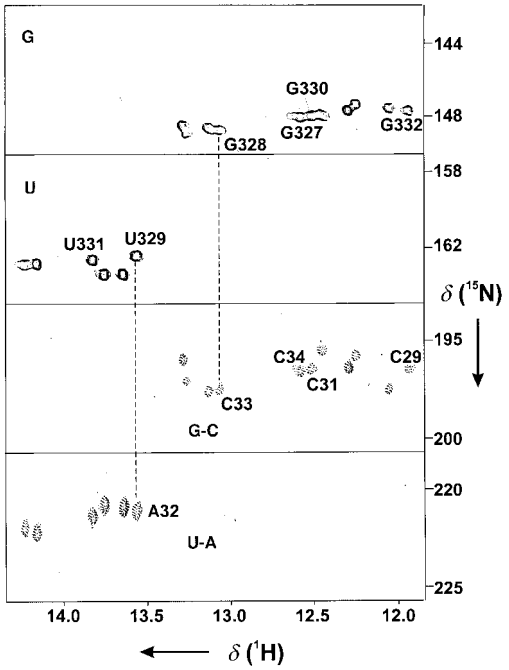

Figure 1. Strips from an HNN-COSY spectrum of a ^{13}C,^{15}N-labeled 69-nucleotide RNA domain.[1] The upper two panels show the COSY "diagonal peaks", that is, the ^{15}N spins (in F1) correlated with their directly bound proton (F2). The cross-peaks in the lower two panels are caused by ^{15}N-^{15}N correlations through a $^{2h}J_{^{15}N,^{15}N}$ coupling across the hydrogen bonds within Watson – Crick base pairs; the correlations G328-N1/C33-N3 and U329-N3/A32-N1 are indicated by dashed lines. (Reproduced with permission from the American Chemical Society 1998.)

[*] Priv.-Doz. Dr. G. Gemmecker
Institut für Organische Chemie und Biochemie
TU München, Lichtenbergstrasse 4
85747 Garching (Germany)
Fax: (+49) 089-289-13210
E-mail: Gerd.Gemmecker@ch.tum.de

  0570-0833/00/3907-1224 $ 17.50+.50/0