Unravelling Structural Information from Complex Mixtures Utilizing Correlation Spectroscopy Applied to HSQC Spectra

ARTICLE *in* ANALYTICAL CHEMISTRY · JULY 2013

Impact Factor: 5.64 · DOI: 10.1021/ac4014379 · Source: PubMed

9 AUTHORS, INCLUDING:

Timothy Robert Rudd
NIBSC
50 PUBLICATIONS **561** CITATIONS
SEE PROFILE

Davide Gaudesi
Università Vita-Salute San Raffaele
13 PUBLICATIONS **125** CITATIONS
SEE PROFILE

Giangiacomo Torri
Istituto di Ricerche Chimiche e Biochimiche G....
264 PUBLICATIONS **6,768** CITATIONS
SEE PROFILE

Marco Guerrini
Istituto di Ricerche Chimiche e Biochimiche G....
105 PUBLICATIONS **2,568** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

Unravelling Structural Information from Complex Mixtures Utilizing Correlation Spectroscopy Applied to HSQC Spectra

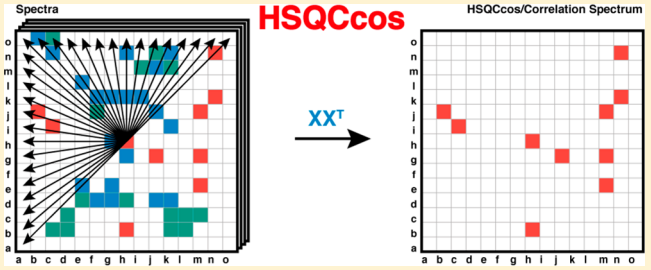
Timothy R. Rudd,*[,†] Eleonora Macchi,[†] Laura Muzi,[†] Monica Ferro,[†] Davide Gaudesi,[†] Giangiacomo Torri,[†] Benito Casu,[†] Marco Guerrini,[†] and Edwin A. Yates[‡]

[†]Istituto di Ricerche Chimiche e Biochimiche "G. Ronzoni", Via Giuseppe Colombo 81, 20133 Milano, Italia

[‡]Department of Structural and Chemical Biology, University of Liverpool, P.O. Box 147, Liverpool, L69 3BX U.K.

Supporting Information

ABSTRACT: The first use of *statistical* correlation spectroscopy to extract chemical information from 2D-HSQC spectra, termed HSQC correlation spectroscopy (**HSQCcos**), is reported. **HSQCcos** is illustrated using heparin, a heterogeneous polysaccharide, whose diverse composition causes signals in HSQC spectra to disperse. **HSQCcos** has been used to probe the chain modifications that cause this effect and reveals hitherto unreported structural details. An interesting finding was that the signal for position 2 of trisulfated glucosamine [N-, 3-O-, and 6-O-sulfated] (**A***) is bifurcated, owing to the presence of **A*** residues in both the "normal" antithrombin binding site and also at the nonreducing end of the molecule, which is reported in intact heparin for the first time. The method was also applied to investigating the environment around other rare sequences/disaccharides, suggesting that the disaccharide; 2-O-sulfated iduronic acid linked to 6-O-sulfated N-glucosamine, which contains a free amine at position 2, is adjacent to the heparin linkage region. **HSQCcos** can extract chemically related signals from information-rich spectra obtained from complex mixtures such as heparin.



Nuclear magnetic resonance (NMR) spectroscopy is a widely used and versatile technique for the analysis of heterogeneous polymers, providing information about chemical content ($^{13}C/^{1}H$, HSQC NMR) and short-range connectivity (NOESY, TOCSY NMR). Additional information, that is not obvious by visual inspection of the data, can be extracted from complex spectral data sets by using *statistical* correlation spectroscopy, which was proposed as a general method by Noda et al.[1] (generalized two-dimensional correlation spectroscopy). This has subsequently been applied widely to one-dimensional NMR spectra.[2−4] The aim of these techniques is to link indirect changes (including those that are possibly not physically linked) in spectral data sets and to use these correlations to extract information about the multicomponent mixtures that were measured, which may include, for example, contaminants in pharmaceuticals and biomarkers in biofluids, but as far as the authors are aware, this analytical approach has not been applied previously to two-dimensional NMR spectra. A group of techniques termed covariance spectroscopy has been used in multidimensional NMR.[5,6] The goal of these techniques is to emulate more complex, experimentally time-consuming, pseudospectra from supposedly simpler and more easily recorded spectra. In those techniques, connections are formed between nuclear spins in molecular systems through the determination of the matrix square root of the inner or outer product of the 2D-spectra in question, (indirect covariance NMR[6]). This forms a symmetrical pseudospectrum and, in the example cited, a pseudo-^{13}C TOCSY spectrum is produced from a HSQC-TOCSY spectrum. Asymmetric spectra have been formed using unsymmetrical indirect covariance NMR, for example, the formation of a GHSQC−COSY spectrum via a GHSQC spectrum and a COSY spectrum.[5]

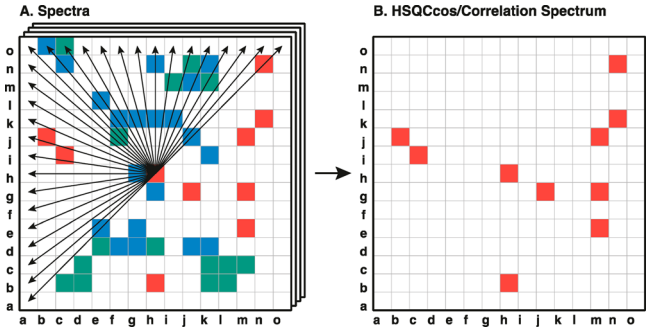
Of the more routine NMR experiments, HSQC can provide ample information about a chemical system without suffering as much from the problem of overlapping signals that besets ^{1}H NMR spectra of multicomponent mixtures, whether those mixtures actually consist of separate chemical entities or comprise heterogeneous compounds. Here, we use *statistical* correlation spectroscopy to explore HSQC spectra (**HSQCcos**), producing correlation spectra, which can be used to elucidate information concerning the composition of mixtures of molecules whose sequences are heterogeneous, correlating a single component or signal, with every other component or signal, in a series of two-dimensional spectra. The technique works by taking a single element from a series of HSQC spectra, or any other two-dimensional data set, and correlating that with every other element, making it possible to see which signals are associated with each other. Unlike indirect covariance analysis, which is used to establish correlations within spectra by using at least one spectrum that contains such information, for example, a TOCSY spectrum, **HSQCcos** determines feature connectivity/association by using the

Received: May 13, 2013
Accepted: July 10, 2013
Published: July 10, 2013

   dx.doi.org/10.1021/ac4014379 | *Anal. Chem.* 2013, 85, 7487−7493

variation within the series of spectra (Scheme 1), which is analogous to the mechanism behind generalized two-dimensional correlation spectroscopy.

Scheme 1. Schematic of the Process Behind Correlation Spectroscopy Applied to Two-Dimensional Data[a]

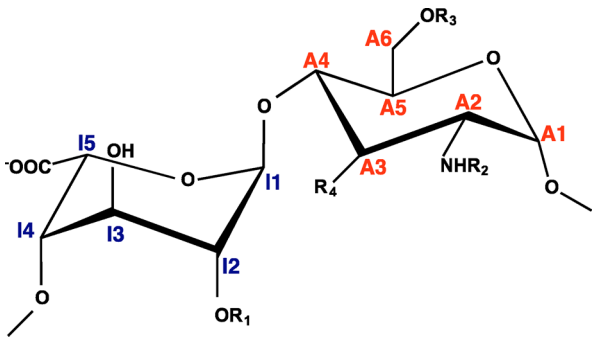


[a](A) Single element or signal from a series of two-dimensional spectra (element *h,h* in this example) is correlated with every other element of the spectra. (B) The result is a correlation/HSQCcos spectrum, which reports all the signals that correlate highly with the signal that was initially interrogated.

The examples reported here investigate the medically important polysaccharide heparin,[7] which is, along with insulin, the most used pharmaceutical product by mass. Recently, it was subject to serious contamination, which led to scores of deaths and hundreds of seriously ill patients, mainly in the U.S.A., following adulteration with another chemically modified polysaccharide.[8,9] There is a pressing need, therefore, for improved methods of analysis, capable of probing the structure of this material, the details of which remain only partially characterized. The methods presented are, however, equally applicable to any 2-dimensional data set of heterogeneous preparations or samples, whether they be pharmaceutical, industrial, biotechnological or bioinformatic in origin.

Heparin is a heterogeneous linear anionic polymer, composed of a repeating disaccharide containing a uronic acid bound via a 1 → 4 linkage to glucosamine (**A**) (Scheme 2). The uronic acid (**U**) can be present as β-D-glucuronic acid (**G**) or, more commonly, its C-5 epimer α-L-iduronic acid (**I**) and

Scheme 2. Predominant Repeating Disaccharide Structure of Heparin[a]



[a]-4) L-IdoA α(1-4) D-GlcN α(1-], where R_1 = H or SO_3^-, R_2 = H/$COCH_3$ or SO_3^-, R_3 = H or SO_3^-, and R_4 = H or SO_3^-. The α L-IdoA can be replaced by its C-5 epimer, β D-GlcA. The carbon atom numberings for the hexuronic acid residue are reported in blue and for the glucosamine residue in red.

the latter can be *O*-sulfated at position 2 ($\mathbf{I}_{2S}$). While the glucosamine (**A**) can be O-sulfated at positions 6 ($\mathbf{A}^{6S}$) and 3 ($_{3S}\mathbf{A}$), at position 2 it can be N-acetylated ($\mathbf{A}_{NAc}$), N-sulfated ($\mathbf{A}_{NS}$) or present as a free amine ($\mathbf{A}_{NH3+}$). **A*** refers to glucosamine, which is also N-sulfated and O-sulfated at position 6 and 3. The biosynthesis of heparin starts from a nonsulfated tetrasaccharide, **G**-β(1−3)-**Gal**-β(1−3)-**Gal**-β(1−4)-**Xyl** (where **Xyl** refers to xylose and **Gal** to galactose), which is attached to a serine in the precursor proteoglycan, this sequence being termed the "linkage-region". The biosynthesis of heparin is not template driven, resulting in a highly heterogeneous polysaccharide. Porcine mucosal heparin (PMH) typically has ∼90% N-sulfation, ∼75% 2-O-sulfation, and ∼80% 6-O-sulfation.[7] The heterogeneity of heparin produces complicated ^{1}H NMR spectra, with many overlapping features and this problem in one-dimensional spectra can be overcome to some extent by the use of multidimensional NMR experiments. Consequently, in the case of heparin, the leap from using ^{1}H to HSQC NMR spectroscopy can provide a wealth of additional information (the HSQC spectrum of a sample of heparin can be found in Supporting Information, where regions of interest are indicated on the spectrum). HSQC spectra have been used to determine the composition of heparin[10] and low molecular weight heparin[11]—heparin that is depolymerized to negate the possible adverse effects of full molecular weight heparin, such as heparin-induced-thrombocytopenia. The innate properties of this heterogeneous polymer cause further complications, because features within HSQC spectra are often composed of multiple signals, distinct, but closely arrayed. The ability to assign these features, without requiring chain depolymerization, would provide sequence information for the polymer being investigated, since they are caused by sequence effects in the neighboring chain environment, and it would then be possible, in principle, to determine the mono- or disaccharides residues that are associated with each other in the heterogeneous polymer. Here, **HSQCcos** of 24 pharmaceutical porcine heparin spectra has been used to extract additional sequence information from the spectra. The use of multiple spectra provides variation, in a way akin to a perturbation, and it is this variation that enables spectral features to be correlated together.

EXPERIMENTAL SECTION

Material. Pharmaceutical heparin is variable in composition between manufacturers and between batches from the same manufacturer. This variation has been explored and characterized previously[2,12,13] and in the present work, 24 heparin samples, originating from five different European manufacturers, that were considered to fall comfortably within the group considered as representing heparin, while also exhibiting variation between them typical of a population of heparin samples, were chosen for this study.

Nomenclature. To further clarify the abbreviations laid out in the introduction, **I** stands for iduronate, **A** for aminosugar (glucosamine), and **nr** indicates that the residue is at the nonreducing end of the molecule. The sub- and superscripts denote the position of sulfation (S) or acetylation (Ac), respectively. Throughout the text, **A***N* or **I***N* refer to position *N* (either C atom or H atom depending on context) of the amino sugar (glucosamine) or iduronate residue, respectively. For example, $\mathbf{I}_{2S}$-$\mathbf{A}^{6S}_{NAc}$ corresponds to the disaccharide 2-O-sulfated iduronic acid linked to 6-O-sulfated N-acetyl glucosamine, while **A2*** signifies position 2 of glucosamine, which is

dx.doi.org/10.1021/ac4014379 | Anal. Chem. 2013, 85, 7487−7493

N-sulfated and O-sulfated at positions 6 and 3. Spectral signals that divagate are numbered to differentiate them, the signal for **A2*** is bifurcated and the daughter peaks are labeled **A2***-1 and **A2***-2.

Digestion of A-G-A*-I-A Pentasaccharide to Form A-G + A*nr-I-A. Enzymatic hydrolysis was performed by incubating 1 mg of **A-G-A*-I-A** pentasaccharide with varying amounts of recombinant human heparanase (5−50 μg) [kindly provided by Professor Israel Vlodavsky] at 37 °C (20 mM ammonium acetate, 2 mM calcium acetate, 1 mM β-mercaptoethanol, pH 5.8), attaining a final volume of 1.25 mL. After 24 h, the incubation mixture was treated with formic acid (0.025%) to disrupt any possible oligosaccharide−protein complexes.

Two Dimensional ^{1}H−^{13}C Heteronuclear Single Quantum Correlation (HSQC) NMR Spectroscopy. The heparin samples were prepared by dissolving 20 mg of each in 0.6 mL of 0.15 mM TSP (trimethyl-silyl-3-propionic acid) solution in deuterium oxide. Spectra were recorded on a Bruker AVIII-600 instrument operating at 600.13 MHz (Bruker, Karlsruhe, Germany). ^{1}H−^{13}C HSQC experiments were acquired with 16 scans for 320 increments in the F1 dimension. The matrix size 1k × 512 was zero filled 2k × 1k by application of a squared cosine function prior to Fourier transformation. The samples were held at a temperature of 298 K during data acquisition. Line broadening of 1.0 and 0.3 Hz, respectively, for F2 and F1, were applied before Fourier transformation and all spectral data sets were processed using TOPSPIN 3.1 (Bruker, Karlsruhe, Germany).

HSQC Correlation Spectroscopy. All analyses were performed on a MacBook Pro (Apple), 2.66 GHz Intel Core i7, 8 Gb memory, running Mac OS X, version 10.8.1. The NMR spectra were processed using Topspin 3.1 (Bruker, Karlsruhe, Germany). The spectra were then imported into R (R, version 2.15.0 (2012−03−30))[14] using the rNMR library (rNMR, version 1.1.7 (2011−08−03)).[15] Using the internal functions of rNMR (*parseAcqus*, *parseProcs*, *bruker2D*, and *ucsf2D*) allowed the 24 HSQC spectra being investigated to be batch imported quickly and easily. When the HSQC matrices were imported into R, they were converted into vectors and placed in one data matrix (one column per HSQC spectrum). This matrix was then normalized for spectral area, mean centered, $X_{ij} = x_{ij} - x_{\text{averagei}}$, and Pareto scaled, $X_{ij} = (x_{ij} - x_{\text{averagei}})\sqrt{x_{\text{sdi}}}^{-1}$. To perform **HSQCcos** the element of interest (HSQC peak) in the HSQC matrix was correlated with every other element. In the mechanics of the analysis this is the equivalent row of the HSQC vector data matrix against every other row. This was performed using the cor.test function (Pearson's coefficient method) in R. Correlations were only considered if they had a p-value greater than 99.9% as used previously in STOCSY.[4] The correlations were converted into a matrix and the square of that matrix was plotted. Figures were produced using the Lattice package.[16] In the resultant correlation spectra, bold annotations are related to the features assigned to the disaccharide/monosaccharide under investigation, and the italic annotations are related to other features of interest.

RESULTS AND DISCUSSION

Heparin, in this case, porcine intestinal mucosal heparin, which is used widely as an anticoagulant in medicine, is a heterogeneous polymer and, despite its medical importance, is only partially characterized. This natural variation, which is reflected in the divergence of its HSQC spectra, serves in a way analogous to a physical perturbation in conventional 2D correlation analysis, allowing **HSQCcos** to correlate features together within the spectra. Four examples are included in the manuscript to illustrate the **HSQCcos** analysis approach. They are the analyses of the signals: **I1** of **I$_{2S}$-A$_{NH2}$** (101.3/5.25 ppm), the comparison of **I1** of **I$_{2OH}$-A^{6OH}** (104.94/5.01 ppm) and **I$_{2OH}$-A^{6S}** (104.7/4.94 ppm), analyses of signals attributed to **A***, and finally the comparison of **G1-A$_{NS}$** (104.7/4.61 ppm) and **G1-A$_{NAc}$** (105.2/4.51 ppm). The examples were chosen because they represent rarer sequences within the heparin chain, illustrating the ability of HSQCos to reveal new structural information, even with signals of relatively low intensity. This analysis assumed that the correlation spectrum generated reflects the immediate environment surrounding the nucleus that is being examined. It should be also recognized that the correlated features extracted by the analysis do not represent defined sequences but correspond to disaccharides/monosaccharides that may be adjacent on either side of the position in question. If not cited specifically, the majority of the spectral assignments were from,[17−19] the assignment for the iduronic[19] and glucuronic[17] acid containing polysaccharides come from the measurement of homogeneous chemical-modified polysaccharides.

Anomeric Signal (I1) of I$_{2S}$-A$_{NH2}$ (101.3/5.25 ppm). An example of the ability of correlation spectroscopy applied to HSQC spectra to extract signals from related structures is shown in Figure 1. The correlation spectrum shown is the result of the analysis of the feature assigned to **I1$_{2S}$-A$_{NH2}$**

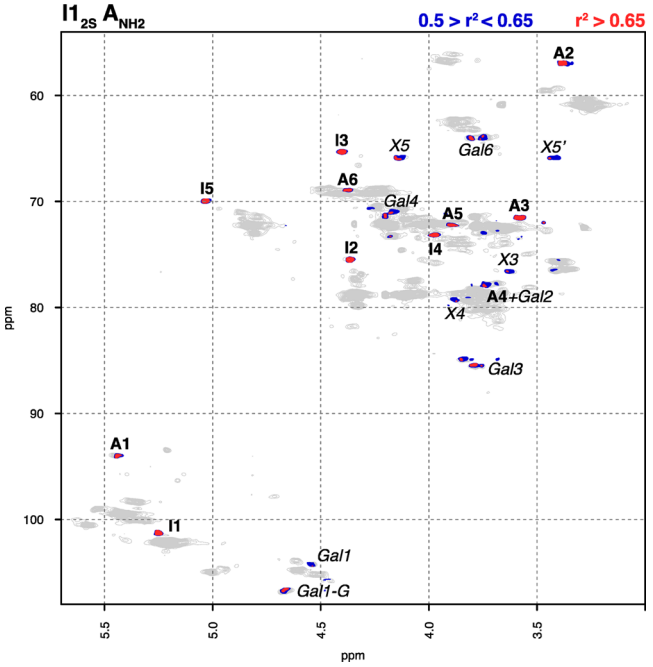


Figure 1. HSQCcos analysis of **I1$_{2S}$-A$_{NH2}$** (101.3/5.25 ppm). By interrogating a single signal, **I1$_{2S}$-A$_{NH2}$**, **HSQCcos** analysis is able to extract the other nuclei associated with that disaccharide (**bold** assignments), in this case confirming that the disaccharide **I$_{2S}$-A$_{NH2}$** is O-sulfated at **A6**. Additionally, the presence of signals arising from the heparin linkage-region (*italic* assignments) indicates that the disaccharide **I1$_{2S}$-A$^{6S}_{NH2}$** is correlated with, and presumably located near the linkage-region of the heparin chain. The underlying spectrum (gray) is a representative porcine intestinal mucosal heparin HSQC spectrum, which is present in the set of spectra that were analyzed. It will appear in all subsequent correlation spectrum figures.

dx.doi.org/10.1021/ac4014379 | *Anal. Chem.* 2013, 85, 7487−7493

(101.3/5.25 ppm). Evaluation of this single peak extracts signals which are due to the disaccharide $\mathbf{I}_{2S}$-$\mathbf{A}^{6S}_{NH2}$[19] (Figure 1 bold assignment) and the heparin linkage-region.[18] Before this analysis it was not known whether the disaccharide $\mathbf{I}_{2S}$-$\mathbf{A}_{NH2}$ was sulfated at position 6 or not. This analysis indicated that it is 6-O-sulfated. The other signals found are due specifically to the heparin linkage-region; the proportion of the heparin chain that links it to its protein component, positions-3, -4, -5, and -5′ of xylose, positions-1, -2, -3, and -4 of galactose (**Gal-Gal**, **Gal-G**) (Figure 1 *italic* assignments).[18] This correlation pattern suggests that $\mathbf{I}_{2S}$-$\mathbf{A}^{6S}_{NH2}$ is adjacent to the reducing end of the heparin samples tested, **-I_{2S}-A_{NH2}-G-Gal-Gal-Xyl-Ser**, (where **Ser** is serine and **Xyl** is xylose), which was unexpected; the linkage region is usually assumed to flank unsulfated regions.[20,21]

Anomeric Signal (I1) of I_{2OH}-A^{6OH} (104.7/4.94 ppm) and I_{2OH}-A^{6S} (104.94/5.01 ppm). Porcine intestinal mucosal heparin is principally composed of $\mathbf{I}_{2S}$, ∼75%. More rarely, **I** can be in the 2-O-nonsulfated form, $\mathbf{I}_{2OH}$, to the extent of ∼10%. Proton 1 of **I** for both $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$ and $\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}$ are clear from interference by other signals, (Figure 2), making them a suitable subject for **HSQCcos** analyses.

It is most likely that the disaccharide $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$ is N-sulfated at position 2 of N-glucosamine ($\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}_{NS}$), because the signal due to **I**-1 for the disaccharide in question correlated with that of $\mathbf{A2}_{NS}$ (Figure 2A). The disaccharide $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}_{NS}$ correlated with a limited number of disaccharides, primarily with $\mathbf{I}_{2S}$-$\mathbf{A}^{6OH}_{NS}$ and weakly with **G**-$\mathbf{A}_{NAc}$. The small correlation with $\mathbf{A2}_{NAc}$ arose from $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$ correlating with **G**-$\mathbf{A}_{NAc}$ (**G**1). The signals extracted matched closely the assignments previously published for $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$ and $\mathbf{I}_{2S}$-$\mathbf{A}^{6OH}_{NS}$.[19]

Unlike the previous disaccharide, $\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}$ is N-acetylated at position 2 of glucosamine ($\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}_{NAc}$), within the heparin samples tested, as $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6S}$ correlated strongly with $\mathbf{A2}_{NAc}$ (Figure 2B). There were no correlations with $\mathbf{I}_{2S}$ (**I2** for example) suggesting the regions containing $\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}_{NAc}$ are exclusively 2-O-nonsulfated, unlike for $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$. The correlation pattern for $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6S}_{NAc}$ was clear, the signals with stronger correlations match the previously published assignment for $\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}_{NAc}$ (Figure 2B bold assignments). This disaccharide is present in the antithrombin (AT)-binding sequence **-I-A-G-A*-I-A** which, in porcine heparin, contains the nonreducing glucosamine almost exclusively N-acetyl-6-O-sulfated. Correlations with **A2***, **A4***, and **G-A*** confirm this observation. Other strong correlations with the disaccharide **G**-$\mathbf{A}_{NS}$ indicate that the disaccharide $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}_{NAc}$ is also present in similar sequences, which do not contain **A***: -$\mathbf{I}_{2S}$-$\mathbf{A}^{6S}_{NS}$-$\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}_{NAc}$-$\mathbf{G}_{2OH}$-$\mathbf{A}^{6S}_{NS}$-[17,22] (Figure 2B italic assignments).

HSQCcos Analysis of A*. Pharmaceutically one of the most important components of heparin is considered to be the trisulfate glucosamine ($_{3S}\mathbf{A}^{6S}_{NS}$ (**A***)) residue of the high affinity AT-binding domain. Within the HSQC spectrum of heparin, there are 4 clear signals that are due to **A*** directly (**A1***, **A2*** (split into two signals) and **A4***(split into two signals)) or **A*** containing disaccharides (**G1-A***). **HSQCcos** analysis of **G1-A*** reveals signals arising from **A*** primarily (**A1***, **A2*** **A3***, and **A4*** (both signals)). Interestingly, **G**4nr and **G**3nr signals arising from glucuronic acid located at the nonreducing end of the polysaccharide, also correlate with **G1-A***, indicating that this disaccharide might be present at the nonreducing end of the chain. Interrogation of **A1*** provided similar information to the analysis of **G1-A***, but **A1*** also correlated with peaks assigned to both **G1**-$\mathbf{A}_{NS}$ and **G1-A***.

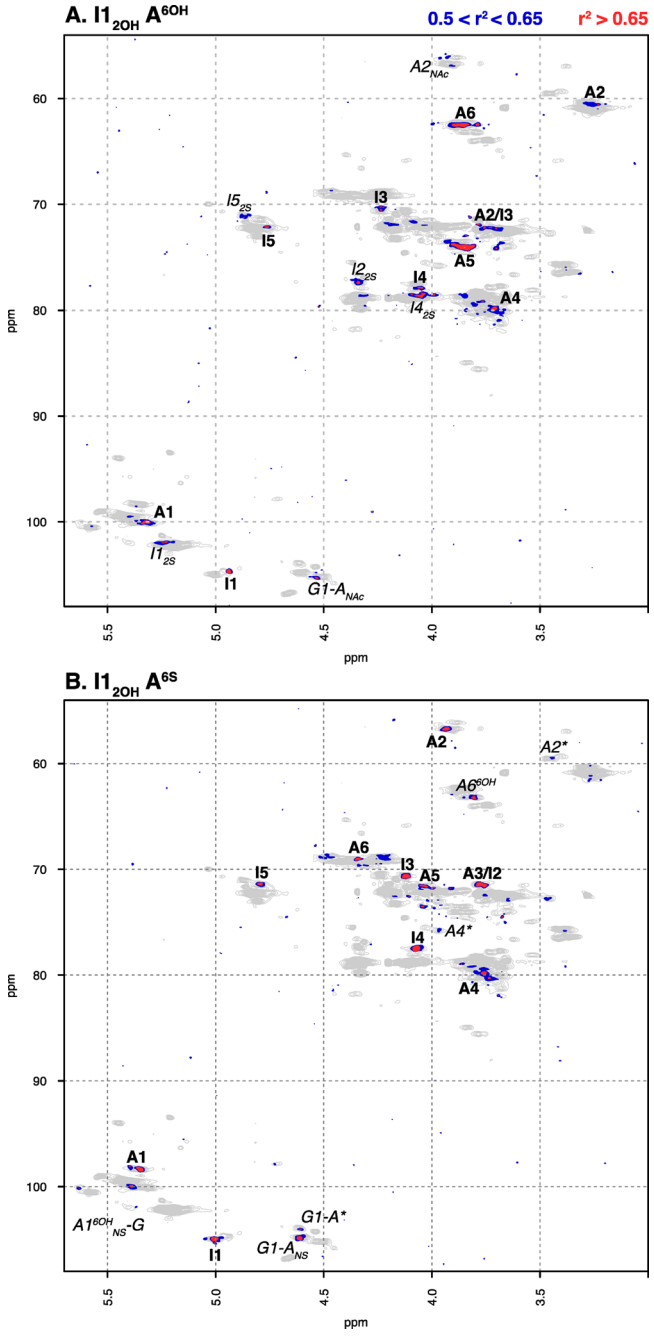


Figure 2. HSQCcos analysis of $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6OH}$ (104.7/4.94 ppm) and $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6S}$ (104.94/5.01 ppm). The assignments in bold pertain to the disaccharide being probed, $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6OH}$ and $\mathbf{I1}_{2OH}$-$\mathbf{A}^{6S}$, while the italic assignments are other features of interest. These analyses indicate that the disaccharide $\mathbf{I}_{2OH}$-$\mathbf{A}^{6OH}$ (**A.**) is N-sulfated at **A2**, and that it is associated with the disaccharide **G**-$\mathbf{A}_{NAc}$ and with regions of heparin containing $\mathbf{I}_{2S}$, whereas $\mathbf{I}_{2OH}$-$\mathbf{A}^{6S}$ is N-acetylated at **A2** and coincides with the disaccharides **G**-$\mathbf{A}_{NS}$ and **G-A***.

There were also correlations with disaccharides containing $\mathbf{I}_{2OH}$ and $\mathbf{I}_{2S}$-$\mathbf{A}^{6S}$.

The two signals arising from **A2*** (the star (*) denotes N-, 3-, 6-trisulfated glucosamine) were scrutinized (**A2***-1, 3.40/59.3 ppm and **A2***-2 3.45/59.5 ppm (Figure 3A and B)), the bifurcation of the signal is due to **A*** being present in two different environments, the main region of the heparin chain and toward the nonreducing end of the chain. The **HSQCcos** analysis of **A2***-1 (Figure 3A) clearly extracts signals for **A4***nr,

dx.doi.org/10.1021/ac4014379 | *Anal. Chem.* 2013, 85, 7487−7493

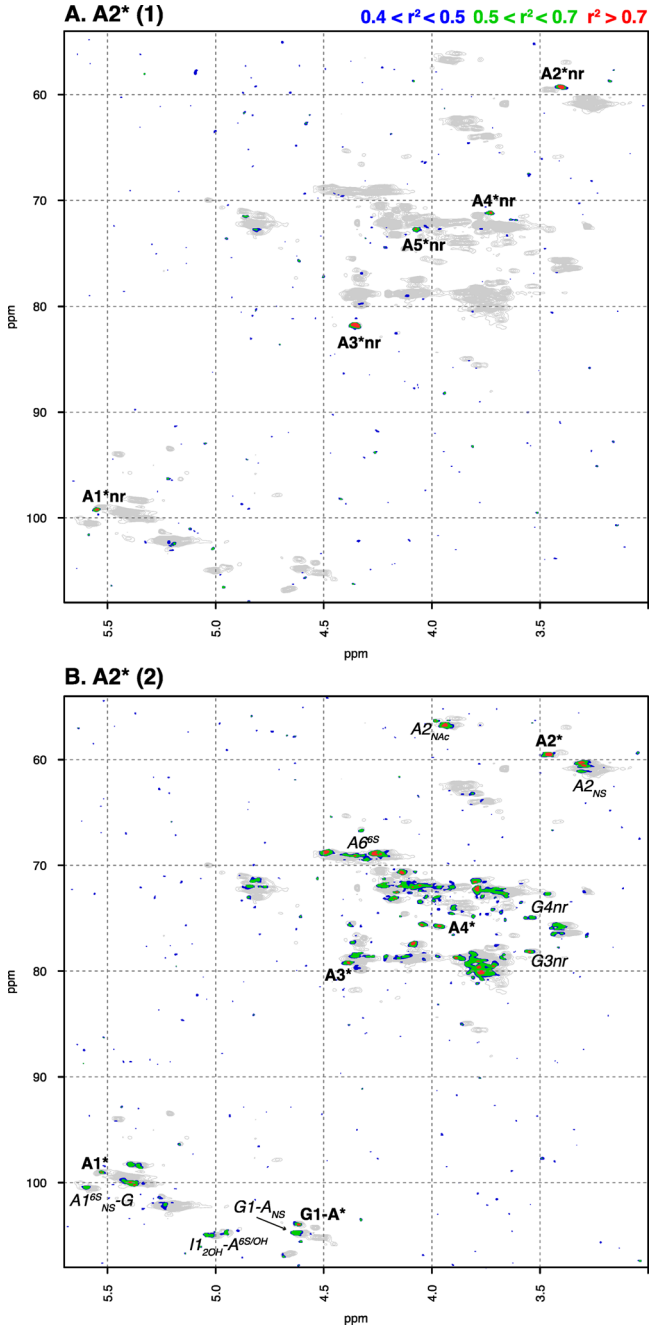


Figure 3. HSQCcos analysis of **A2***, **A**. **A2***-1 (59.3/3.40 ppm) and **B**. **A2***-2 (59.5/3.45 ppm. Within the correlation spectra (**A** and **B**) the assignments in bold pertain to the disaccharide being probed, while the italic assignments are other features of interest. **A2***-1 correlates with signals for **A*** located at the nonreducing end of the molecule, while, **A2***-2 corresponds to **A*** in the body of the heparin chain.

A3*nr, and **A1***nr, and these signals closely match those of a trisaccharide with **A*** located at the nonreducing end (Figure 4).

The trisaccharide was formed by depolymerising fondaparinux (**A-G-A*-I-A**) with an endoglucuronidase, heparanase, producing a mixture of the saccharides **A-G** and **A*-I-A**, the chemical shifts of which can be found in Supporting Information. The other **A2*** signal (**A2***-2 3.45/59.5 ppm) correlated with the general sequences associated with **A***, as found by the analyses of the signals **G1-A*** and **A1*** (Figure

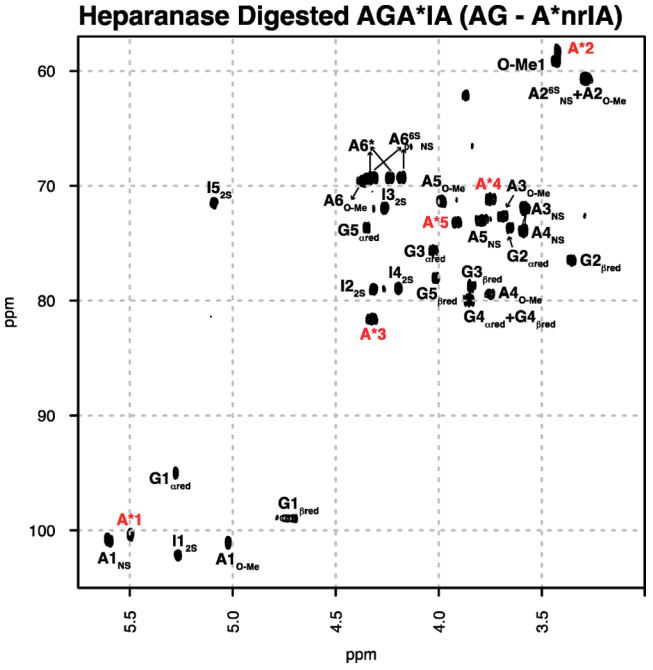


Figure 4. H^1−C^{13} HSQC spectra of heparanase digested pentasaccharide **A-G-A*-I-A**, itself a pharmaceutical agent, fondaparinux. The heparanase digestion of **A-G-A*-I-A** forms the fragments **A-G** and **A*nr-I-A**, allowing the signals of **A*nr** to be assigned (Figure 3**A**). In the spectra of **A-G** and **A*nr-I-A**, the red assignments are for the monosaccharide A*nr.

3**B**). Both $I1_{2OH}$-A^{6S}_{NAc} and $I1_{2OH}$-A^{6OH}_{NS} correlated with **A2***-2, although only **I1** was found for the disaccharide $I1_{2OH}$-A^{6OH}_{NS}, whereas the majority of signals arising from $I1_{2OH}$-A^{6S}_{NAc} were observed. **A2***-2 correlated to both **G1-A*** and **G1-**A_{NS}, and also extracted signals arising from **G2** (**G2** 1 74.71/3.40 ppm and **G2** 2 76.50/3.41 ppm), **G3**nr and **G4**nr, and interestingly, there were correlations with $I_{2S/OH}$-A_{NH2}. Thus, additional structural information has been gained directly from the heterogeneous mixture without recourse to additional fractionation or purification.

Glucuronic Acid Containing Sequences within Porcine Heparin: G-A_{NS} (104.7/4.61 ppm) and G-A_{NAc} (105.2/4.51 ppm). While porcine intestinal mucosal heparin is mainly, ∼85%, composed of disaccharides containing iduronic acid, the remaining disaccharides contain glucuronic acid. The most common of these disaccharides are **G-**A_{NS} and **G-**A_{NAc}, ∼7% for each, and the least common is the trisulfated glucosamine containing disaccharide, **G-A***, ∼3%.

The N-sulfate containing disaccharide (Figure 5**A**) correlated weekly with **G-**A_{NAc} and vice versa, **G-**A_{NS} also correlated with $A^{6S/6OH}$-**G**, indicating that **G-**A_{NS} is present in sulfaminoheparosan-like sequences.[17] The correlation with N-acetyl glucosamine is specific for the lower lobe of the $A2_{NAc}$ signal. The disaccharide **G-**A_{NS} is associated with **I**, correlating with I_{2OH}-A^{6S} and A_{NS}-I_{2OH}, which suggests that **G-**A_{NS} occurs in sequences with stretches of primarily 2-O-nonsulfated iduronic acid. The presence of A_{NS} attached to **G** and **I** could explain the two separate correlations with $A2_{NS}$, as well as their being **G-**A^{6OH}_{NS} and/or **G-**A^{6S}_{NS}.

Unlike **G-**A_{NS}, it is clear that **G-**A_{NAc} is almost exclusively de-6-O-sulfated, the correlation was very strong with $A6^{6OH}$ (Figure 5**B**). As stated previously, **G-**A_{NAc} and **G-**A_{NS} correlated with each other, but in the case of **G-**A_{NAc} the features in $A2_{NAc}$

dx.doi.org/10.1021/ac4014379 | Anal. Chem. 2013, 85, 7487−7493

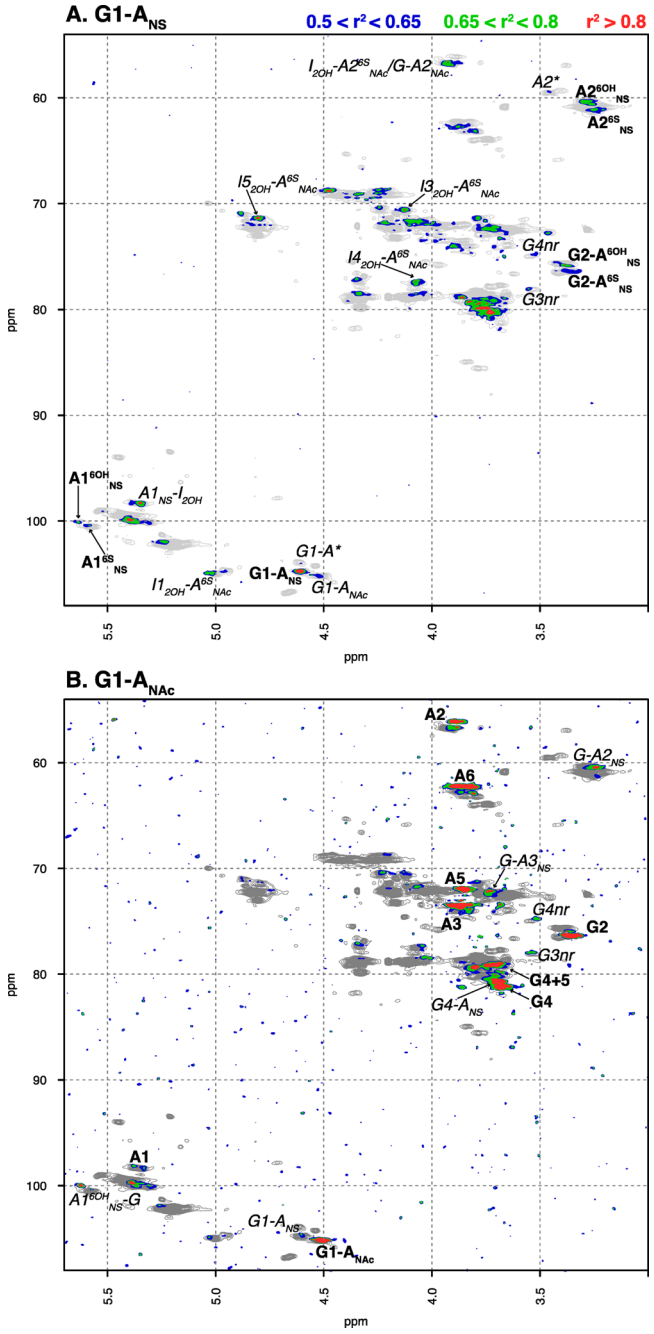


Figure 5. HSQCcos analysis of (A) **G1-A**$_{NS}$ (104.7/4.61 ppm) and (B) **G1-A**$_{NAc}$ (105.2/4.51 ppm). The assignments in bold pertain to the disaccharide being probed, while the italic assignments are other features of interest. **G1-A**$_{NS}$ correlated to **I**$_{2OH}$-**A**$^{6S}_{NAc}$, **G-A**$_{NAc}$ and weakly to disaccharides containing **A***, while **G1-A**$_{NAc}$ correlated to fewer disaccharides, principally **G-A**$_{NS}$.

and **A2**$_{NS}$ that are highlighted are different to those that correlate with **G-A**$_{NS}$. For example, the upper lobe of **A2**$_{NAc}$ correlated predominantly with **G-A**$_{NAc}$. Although **G-A**$_{NAc}$ correlations indicate that this disaccharide is almost exclusively present in sequences with lower sulfation, its correlations with **G-A**$_{NS}$ and the upper lobe of the **A**$_{NS}$-**G** signal suggest that **G-A**$_{NAc}$ is also present with repeating units of **G-A**$_{NS}$; Unlike **G-A**$_{NS}$, **G-A**$_{NAc}$ correlates very weakly with **I**$_{2OH}$ and **I**$_{2S}$ signals, confirming its presence preferentially in heparan sulfate like sequences. A common correlation of both **G-A**$_{NAc}$ and **G-A**$_{NS}$ is s with nonreducing **G**3 and **G**4 at 3.51/75.3 ppm and 3.52/72.0 ppm, respectively.

CONCLUSIONS

Correlation spectroscopy analysis of HSQC spectra provides the means for extracting signals within cluttered, dense data sets, illustrated here with 2-dimensional HSQC NMR spectra of the complex pharmaceutical, heparin. The process can discriminate the sequence effects that cause signals to diverge for a given peak. Heparin was chosen as an example because it is composed of heterogeneous polymeric chains with considerable overlap in 1-dimensional spectra, yet its signals can be resolved sufficiently in 2-dimensions to allow a correlation-based analysis to be attempted. Obviously, other materials could also be studied in a similar manner, the fundamental limit being reached only when signals can no longer be resolved in 2-dimensions. **HSQCcos** was used here to interrogate the environment around the less common disaccharides within heparin, providing sequence information that would normally require depolymerization of the polymer.

HSQCcos analysis indicated that the signal arising from **I1**$_{2S}$-**A**$_{NH2}$ originated from the disaccharide **I**$_{2S}$-**A**$^{6S}_{NH2}$, it was not known previously whether this disaccharide was 6-O-sulfated or not in heparin until this analysis. Furthermore, **HSQCcos** identified that this disaccharide is strongly associated with the reducing end of the polysaccharide, specifically the heparin linkage-region. This was unexpected since it is usually assumed that the linkage region is adjacent to residues without sulfation and epimerization, although the details of the mechanisms by which heparin biosynthesis are controlled are far from complete.

Moreover, the approach was able to determine that the disaccharides **I**$_{2OH}$-**A**^{6OH} and **I**$_{2S}$-**A**^{6S} have preferential groups at position 2 of glucosamine: N-sulfate for the former and N-acetyl for the latter. These two disaccharides appear in different sequences within heparin, **I**$_{2OH}$-**A**$^{6OH}_{NS}$ is located within regions containing **I**$_{2S}$-**A**$^{6OH}_{NS}$ and small amounts of **G-A**$_{NAc}$, while **I**$_{2OH}$-**A**$^{6S}_{NS}$ is associated with **G-A**$_{NS}$ and **G-A***. Similar information was extracted for the disaccharides **G-A**$_{NS}$ and **G-A**$_{NAc}$. The former disaccharide is present in the heparin samples tested in a mixed environment, in sequences containing **G-A***, **G-A**$_{NAc}$ and **I**$_{2OH}$-**A**$^{6S}_{NAc}$, while **G-A**$_{NAc}$ is on the whole present in very uniform regions containing itself and **G-A**$_{NS}$. There are minor correlations with other signals but, these are very weak compared to the correlations with itself and **G-A**$_{NS}$.

The presence of **A*** at the nonreducing end of the chain is also interesting, although A*nr has already been detected after periodate cleavage of heparin,[23] its presence in unfractionated heparin has never been described. The periodate treatment, sometimes used during the purification processes, might be the reason for the presence of minor amounts of A*nr in the heparin samples. Another possibility, is that the latter residue was generated by the action of natural endoglucuronidases, that is, heparanase.

In the example cited here, heparin structure was investigated, but **HSQCcos** could be applied widelyto many systems containing intrinsic heterogeneity and could include other heterogeneous preparations such as drugs based on mixtures of macromolecules as well as in the quality control of a range of industrial pharmaceutical or biotechnological products, the limiting requirement being that signals can be resolved in 2-dimensions. Correlation analysis of 2D-NMR spectra is not limited to HSQC spectra, which would report chemical content

dx.doi.org/10.1021/ac4014379 | Anal. Chem. 2013, 85, 7487−7493

of a compound or compounds; it could be equally applied to TOCSY or NOESY spectra for example, thereby extracting additional structural information about the sample.

■ ASSOCIATED CONTENT

Supporting Information

Additional material as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: timothy.rudd@diamond.ac.uk.

Present Addresses

Timothy R. Rudd: Beamline 23 (Circular Dichroism), Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire, OX11 0DE U.K.

Davide Gaudesi: Dulbecco Telethon Institute, Biomolecular NMR Laboratory, c/o Ospedale S. Raffaele, via Olgettina 58, 20132 Milano, Italia

Monica Ferro: Dipartimento di Chimica, Materiali ed Ingegneria Chimica "G. Natta", Politecnico di Milano, via Mancinelli 7, 20131 Milano, Italia

Author Contributions

M.G. and E.A.Y. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from Finlambardia SPA "Fondo per la promozione di Accordi Istituzionali". The authors would like to thank Dr Mark Tully for reading the manuscript and providing feedback.

■ REFERENCES

(1) Noda, I.; Dowrey, A. E.; Marcott, C.; Story, G. M.; Ozaki, Y. *Appl. Spectrosc.* **2000**, *54*, 236A−248A.

(2) Rudd, T. R.; Gaudesi, D.; Lima, M. A.; Skidmore, M. A.; Mulloy, B.; Torri, G.; Nader, H. B.; Guerrini, M.; Yates, E. A. *Analyst* **2011**, *136*, 1390−8.

(3) Rodrigues, J. A.; Barros, A. S.; Carvalho, B.; Brandao, T.; Gil, A. M. *Anal. Chim. Acta* **2011**, *702*, 178−87.

(4) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282−9.

(5) Martin, G. E.; Hilton, B. D.; Irish, P. A.; Blinov, K. A.; Williams, A. J. *J. Nat. Prod.* **2007**, *70*, 1393−6.

(6) Zhang, F.; Bruschweiler, R. *J. Am. Chem. Soc.* **2004**, *126*, 13180−1.

(7) Rabenstein, D. L. *Nat. Prod. Rep.* **2002**, *19*, 312−31.

(8) Guerrini, M.; Beccati, D.; Shriver, Z.; Naggi, A.; Viswanathan, K.; Bisio, A.; Capila, I.; Lansing, J. C.; Guglieri, S.; Fraser, B.; Al-Hakim, A.; Gunay, N. S.; Zhang, Z.; Robinson, L.; Buhse, L.; Nasr, M.; Woodcock, J.; Langer, R.; Venkataraman, G.; Linhardt, R. J.; Casu, B.; Torri, G.; Sasisekharan, R. *Nat. Biotechnol.* **2008**, *26*, 669−75.

(9) Kishimoto, T. K.; Viswanathan, K.; Ganguly, T.; Elankumaran, S.; Smith, S.; Pelzer, K.; Lansing, J. C.; Sriranganathan, N.; Zhao, G.; Galcheva-Gargova, Z.; Al-Hakim, A.; Bailey, G. S.; Fraser, B.; Roy, S.; Rogers-Cotrone, T.; Buhse, L.; Whary, M.; Fox, J.; Nasr, M.; Dal Pan, G. J.; Shriver, Z.; Langer, R. S.; Venkataraman, G.; Austen, K. F.; Woodcock, J.; Sasisekharan, R. *N. Engl. J. Med.* **2008**, *358*, 2457−67.

(10) Guerrini, M.; Naggi, A.; Guglieri, S.; Santarsiero, R.; Torri, G. *Anal. Biochem.* **2005**, *337*, 35−47.

(11) Guerrini, M.; Guglieri, S.; Naggi, A.; Sasisekharan, R.; Torri, G. *Semin. Thromb. Hemost.* **2007**, *33*, 478−87.

(12) Rudd, T. R.; Gaudesi, D.; Skidmore, M. A.; Ferro, M.; Guerrini, M.; Mulloy, B.; Torri, G.; Yates, E. A. *Analyst* **2011**, *136*, 1380−9.

(13) Rudd, T. R.; Macchi, E.; Gardini, C.; Muzi, L.; Guerrini, M.; Yates, E. A.; Torri, G. *Anal. Chem.* **2012**, *84*, 6841−7.

(14) R Development Core Team. *R*, version 2.13.1 (2011-07-08); R Foundation for Statistical Computing: Vienna, Austria, 2008.

(15) Lewis, I. A.; Schommer, S. C.; Markley, J. L. *Magn. Reson. Chem.* **2009**, *47* (Suppl 1), S123−6.

(16) Sarkar, D. *Lattice: Multivariate Data Visualization with R*; Springer: New York, 2008.

(17) Casu, B.; Grazioli, G.; Razi, N.; Guerrini, M.; Naggi, A.; Torri, G.; Oreste, P.; Tursi, F.; Zoppetti, G.; Lindahl, U. *Carbohydr. Res.* **1994**, *263*, 271−84.

(18) Iacomini, M.; Casu, B.; Guerrini, M.; Naggi, A.; Pirola, A.; Torri, G. *Anal. Biochem.* **1999**, *274*, 50−8.

(19) Yates, E. A.; Santini, F.; Guerrini, M.; Naggi, A.; Torri, G.; Casu, B. *Carbohydr. Res.* **1996**, *294*, 15−27.

(20) Lindahl, U.; Cifonelli, J. A.; Lindahl, B.; Roden, L. *J. Biol. Chem.* **1965**, *240*, 2817−20.

(21) Lindahl, U. *Biochim. Biophys. Acta* **1966**, *130*, 368−82.

(22) Yamada, S.; Yamane, Y.; Tsuda, H.; Yoshida, K.; Sugahara, K. *J. Biol. Chem.* **1998**, *273*, 1863−71.

(23) Islam, T.; Butler, M.; Sikkander, S. A.; Toida, T.; Linhardt, R. J. *Carbohydr. Res.* **2002**, *337*, 2239−43.