University of Maryland Baltimore County

CardioPEG Techniques in Glycobiology– NMR Spectroscopy
Oct 16, 18, 2013
Tour of NMR lab.WBSB 312, Weds Oct 16, 2 PM-3 PM

Instructor: C. Allen Bush and Bruce Johnson – UMBC Chemistry and Biochem. bush@umbc.edu

Bring a laptop for the workshop. You can install the software for NMR data analysis on that.

- **NMR spectroscopy** In NMR spectroscopy, about 1 ml of the liquid sample is placed in a sample tube in a strong magnetic field. Radio frequencies in the 20 Mhz to 800 Mhz range irradiate the sample stimulating spectroscopic transitions between spin states of the nuclei in the magnet. The energy difference between the Zeeman states is proportional to the magnetic field strength, B_0 so the stronger the magnetic field, the higher is the radio frequency and the higher is the sensitivity and spectral resolution. Thus there is strong motivation to use the highest B_0 possible. The radio frequency depends both on the nucleus detected and the field strength and it is generally specified as the frequency at which protons resonate. The NMR probe placed inside the magnet has radio frequency coils tuned to the appropriate frequencies. A superconducting magnet, used due to its high field, has geometry known as a solenoid.[slide 2, 3] In chemical and biochemical applications, a vertical bore magnet is used so the B_0 is vertical rather than horizontal. Electric current flows continuously (no electrical resistance) in the superconducting magnet, which is maintained at 4 K by liquid helium. Once powered up, the magnet draws no additional electric current so behaves like a permanent magnet. The cryogens, N_2 and He, must be replenished at regular intervals.

  Most common applications are to nuclei with spin 1/2, in particular for carbohydrates, ^{1}H and ^{13}C but other nuclei such as ^{31}P and ^{15}N are often used. Modern instruments employ digital RF pulses which produce a response function in the time domain called the free induction decay (FID) rather than a spectrum. [slide 4]. Older instruments used continuous RF whose frequency was scanned to generate the spectrum. The difference between a such a CW and a pulse FT spectrometer can

be likened to that between a flute that makes a musical note at a single frequency (in Hz) and a bell that is struck. As the response of the bell decays away in the time domain its sound can be analyzed into its frequency components. The mathematics is called Fourrier transformation. The FID is recorded digitally by a radio receiver.[slide 4] The spectrum, giving the response in the frequency domain can be recovered with a Fourier transform, a mathematical integration conveniently performed with a computer. Since the experiment requires a computer, and the radio frequencies of both the transmitter and receiver are digital, the whole instrument is computer controlled. The user need not be very aware of or involved in the math required for most experiments.

This method provides the most complete information on carbohydrate structure. NMR is sensitive to stereochemistry so it can give information on anomeric configuration of linkages and is the best general method for it. Simple 1-d ^{1}H NMR spectra can be used to quickly determine the purity and homogeneity of an oligosaccharide preparation and can readily identify known structures from available published reference spectra. Other 2-dimensional NMR methods can be used to provide linkage analysis replacing methylation analysis. Since the method is non-destructive, samples can be recovered for other experiments. The main drawback to NMR is the limited sensitivity which ranges from 100 nM to 1 μM with the more powerful experiments needed for linkage analysis requiring the larger amount.

- **^{1}H NMR spectroscopy** In the early 1980's the introduction of high field NMR spectroscopy with superconducting magnets made possible determination of glycolipid and glycopeptide structures much more completely and reliably than had been previously available, revolutionizing the field of complex carbohydrate structure.

  To detect ^{1}H NMR in oligosaccharides or polysaccharides, amounts from 50 n moles to μmoles are needed. Water soluble glycopeptides, [slide 5] oligosaccharides or almost any derivatives are dissolved in D_2O. Slide 6 shows the spectrum of an oligosaccharide alditol isolated by alkaline borohydride degradation of a blood group A active mucin (O-linked glycans – structures in slide 7.) In such samples, ^{2}H (deuterium) is exchanged for ^{1}H and we observe carbon-bound protons only. Since they exchange with aqueous solvents, hydroxyl and amide protons are

not seen in most experiments.

- **Analysis of spectra features** –

  Frequencies (chemical shift) depend on general features of the chemical environment; anomeric (4.5 to 5.5 ppm), methyl, (1.5 to 2.5 ppm) following usual rules taught in organic chemistry. [slide 8] A major problem is that most of the methine protons are chemically very similar, differing only in stereochemistry. The chemical shifts of the signals are very similar and ^{1}H signal overlap among ring protons is severe so there are few resolved peaks between 3.5 and 4 ppm. Therefore successful carbohydrate NMR spetroscopy was impossible before high field superconducting instruments became common about 1980. The signal resolution problem continues to be serious and most experiments are done on 500-600 MHz instruments. Multidimensional NMR (2-d or 3-d) is essential to improve resolution. The instrument measures frequencies in Hz but spectral data is more commonly discussed in ppm (parts per million) which facilitates comparison of spectra recorded at different magnetic field strengths.

  There are a few isolated and characteristic (structural reporter) signals which can be used to identify known structures even in simple 1-d spectra. Anomeric signals are reasonably well resolved and most generally useful. Some equatorial protons are isolated near 4 ppm.(mannose H2 and galactose H4) [slide 6,8] But chemical shifts are readily measured accurately so even a few lines may be adequate to identify an oligosaccharide or glycopeptide. Methyl groups of 6-deoxy sugars are also useful reporter groups in the upfield region 1.0 - 1.7 ppm. The chemical shifts of complex oligosaccharides are strongly influenced by structure leading to fingerprints characteristic of a specific structure. This provides a quick proof of purity of an oligosaccharide isolated by chromatography since integrated signal amplitudes can be used to count protons in a sample. Likewise, if an NMR spectrum of a certain glycopeptide or oligosaccharide is published, identification of that compound from other sources from a 1-d ^{1}H NMR spectrum is simple. (There are no known examples of two different complex oligosaccharides with the same NMR spectrum.)

- **Assignment of NMR signal lines to specific nuclei**. The fingerprint method works well only if reference spectra of authentic samples

of known structure are available. For a new or unknown structure, additional NMR information must be obtained. Other information about a resonance in addition to the chemical shift comes from the spin coupling constants. (Review the NMR chapter in your elementary organic chemistry text.) These splittings of the spectral lines arise from the influence on a nucleus of a neighboring nucleus which is transmitted through the chemical bonding electrons. (scalar coupling or J-coupling). [slide 9] The values of these coupling constants depend on the relative orientation of the bonds connecting the nuclei. For three-bond coupling between protons, the dihedral angle of the bonds is correlated with the value of the coupling constants by a 'Karplus equation'. [See Taylor and Drickamer, pp. 87-88.] In a pyranoside the coupling constants are either large (7-11 Hz) if both ^{1}H are axial (*trans*) or the coupling constant is small (0-3 Hz) if one of the ^{1}H is equatorial and the ^{1}H are *gauche*. [slide 10] Thus the anomeric configuration of a sugar with an axial H2 can be readily determined from the splitting of the anomeric ^{1}H signal which is generally isolated (a structural reporter). [slide 6, 8] [Taylor and Drickamer, pp 87-88]. There are additional methods for determining anomeric configuration if H2 is equatorial (NOE and one-bond J_{CH}) so NMR is the primary method of choice for determining anomeric configuration.

The most critical, and often most difficult, step in the extraction of valuable structural information about a carbohydrate, or a protein or a polynucleotide requires assignment of the spectrum. Assignment of proton signals to pyranoside ring spin systems is done by sequential correlation of H1 to H2, H2 to H3, H3 to H4 etc. [slide 11] Spin correlation can be detected by decoupling experiments in which a radio frequency source is used to saturate a coupled signal. This causes the splitting of the resonance coupled to it to collapse thus identifying coupled proton pairs of vicinal protons. This is crucial for the assignment of each spectral line to a particular hydrogen in the chemical structure. Since each sugar ring is a spin system or chain of coupled spins, following this chain of connectivity can be used to assign signals to individual monosaccharides. This method of selective decoupling by selective irradiation fails in the crowded region of carbohydrate ^{1}H NMR spectra due to overlapping signals.

- **2-d NMR** In 2-dimensional NMR spectra, the crowded spectrum is

spread out in a second dimension by correlations (cross peaks) between different nucleii at different chemical shifts in much the same way as 2-d chromatography is used to further resolve chromatographic peaks by resolution with a second chromatographic technique. Multidimensional NMR is possible only in pulsed FT NMR spectroscopy. RF pulses are arranged with regularly incremented delays during which the chemical shift evolves creating a second, indirect time dimension. [slide 12, 13] Both the detected dimension (normal NMR FID) and the indirect dimension are Fourier transformed to give a 2-d frequency spectrum. [slide 12]

2-d correlation spectroscopy (COSY) generates cross peaks between two scalar coupled resonances, typically from atoms separated by three or fewer bonds.[slide 12] This has the dual effect of greatly improving the spectral resolution and allowing complete assignment by tracing the chain of proton spins around the ring. [slide 12] Values of the coupling constants (splittings) can be estimated from the cross peak shapes indicating axial or equatorial ^{1}H position and thus identifying the stereochemistry of the pyranoside. [slide 8, 12] Sugars with the gluco configuration [slide 9] have large J coupling between H2-H3, H3-H4 and H4-H5 while those with the galacto configuration with an equatorial H4 have small values of J coupling for H3-H4 and H4-H5. [slide 8] (Taylor and Drickamer, Fig. 6.1).

The nature of the cross peaks in 2-d NMR is determined by the 'pulse sequence', the program of RF pulses sent to the NMR probe which holds the sample.[slide 13] The computer holds the information on the pulse program that tells the box of radios when to turn on and off as well as what radio frequency and power level to use. It also turns on the radio receiver. Thus a different pulse sequence can be made to measure TOCSY (HOHAHA) spectra which provide extended correlation within the sugar ring with cross peaks between H1 and H3 or H1 and H4. [slide 16]

Another pulse sequence generates 2-dimensional Nuclear Overhauser effect (NOESY) connects ^{1}H through space rather than through the chemical bonding electrons. Thus cross peaks depend on conformation in 3-d space. *gauche* vicinal protons give large NOESY cross peaks (but small scalar coupling peaks). For sugars such as β-glucose and β-galactose, the anomeric proton is relatively close in space to H5 gen-

erating a NOESY cross peak that is very useful for assignments of the complete sugar ring. In addition, cross peaks are seen between adjacent sugar residues giving sugar sequence and some information on the linkage.[slide 17] In most, but not all cases, NOE cross peaks are largest between anomeric and aglycone proton directly across the glycosidic linkage. (Taylor and Drickamer, Fig. 7.13). It is best to record NMR spectra in phase-sensitive mode which gives sharper line-shape as well as information on the sign (positive or negative) of the signal [slide 18]. The red color in slide 18 represents a negative signal.

- **^{13}C NMR spectra** ^{13}C is an isotope which is only 1% abundant in nature but it is often used for NMR spectroscopy. The sensitivity is lower because of the low abundance and also the lower frequency which is 1/4 the proton frequency. Carbon spectra have much better chemical shift dispersion and much less overlap than do ^{1}H spectra. [slide 19] ^{13}C spectra have been used for 25 years to characterize bacterial polysaccharides which can generally be isolated in quantity. The sensitivity of directly detected ^{13}C spectra is poor requiring up to 20 mgs. of polysaccharide for an experiment on a high molecular weight complex carbohydrate. Because of the low abundance, there is no ^{13}C-^{13}C coupling in natural abundance samples so there is no rigorous way to assign ^{13}C spectra as there is by spin correlation of ^{1}H. (99.9% abundant) But spectra give an excellent quantitative accounting of carbon atoms and have been of great importance for structure determinations in bacterial polysaccharides, often available in large quantities. This is very rarely the case for glycopeptides or glycoprotein oligosaccharides limiting the utility of directly detected ^{13}C NMR for that case.

  In the past 20 years it has become common for complex carbohydrates to measure ^{1}H-detected ^{13}C correlation (HMQC or HSQC) which is more sensitive than directly detected ^{13}C spectroscopy and the spectra correlate C-H bonded pairs. [slide 20] Thus, the ^{1}H assignment by spin correlation can be used to completely assign the ^{13}C spectra. Resolution of the ^{13}C dimension spreads out overlapping peaks and rigorously connects ^{13}C assignment to the ^{1}H assignment. The improved resolution of ^{13}C - ^{1}H correlation spectroscopy provides an essentially complete solution to the problem of signal overlap in carbohydrates, making it possible to give complete signal assignments for both ^{1}H and ^{13}C for determination of complex structures of up to about 10 residues.

[slide 21]

Anomeric proton NMR chemical shifts are generally found between 4.5 ppm and 5.5 ppm where few other resonances occur. The ^{13}C chemical shifts are generally between 90 and 110 ppm a region where essentially no other signals appear so that region in the C-H correlation spectrum can be used to count uniquely anomeric signals and hence sugar residues. Those signals also provide a good starting point for assignment of the other signals of each monosaccharide ring by correlation. Each C-H cross peak in the anomeric region represents a single sugar residue (pyranoside or furanoside) in the oligosaccharide, glycopeptide or glycolipid. Therefore the peaks can be assigned a unique letter (A-E in this case) representing a residue in this polysaccharide which has five sugars and one ribitol (F) in the repeating subunit.

Slide 21 is a 'multiplicity-edited' C-H HSQC spectrum in which the cross peaks of methylene groups (a 3-spin system with two protons and one ^{13}C) are negative, i.e. below the plane of the figure. They appear in red in slide 21, a feature that is especially useful in identifying the C6 of pyranosides and the methylenes of the ribitol (residue F). Methine groups (C-H) are phased positive and represented as black.

- Long-range coupling correlation (HMBC) – A slight modification of the HMQC pulse sequence allows for the detection of 2-bond and 3-bond C-H coupling correlations at somewhat reduced sensitivity. [slide 22] Cross peaks are seen between ^{1}H and ^{13}C within the sugar ring and also across the glycosidic linkage. HMBC can be used to resolve some ambiguities in the assignments due to overlaps in HMQC spectra. With full assignment of ^{1}H and ^{13}C spectra, long-range C-H correlation (HMBC) indicates linkage positions thus avoiding the need for Methylation analysis. [Abeygunawardana and Bush, *Advan. Biophys. Chem.* **3**, 199-249 (1993)]

- **Processing of the NMR Data** The NMR spectrometer records data in the time domain–that is the Free Induction Decay (FID) [slide 4] which must be Fourier transformed to obtain the spectrum. This can be done on the spectrometer computer using the manufacturer's software which is fine for 'quick and dirty' inspection of the data but for a more careful inspection of the hard-won spectral information, it is preferable to transfer the data to a general purpose computer. This is especially

true for complex 2-dimensional heteronuclear data that may contain a wealth of valuable structural information.

Advantages are that one can use much better and more flexible software than that provided by the spectrometer vendors, who make good instruments but generally poor software. In addition, processing and analysis can be done on your own time and in your own space so the spectrometer, which costs lots more than an ordinary computer, can be used for other experiments. One can easily transfer the data over a network to:

- The UMBC NMR data laboratory [slide 27]
- Your laboratory at JHU or anywhere
- Your own laptop.

For the intial processing of data, we recommend 'nmrPipe' a program originally written at NIH and now distributed by Frank Delaglio. For advanced data analysis, we recommend nmrview distributed by One Moon Scientific. These programs are generally available for use by academic institutions.

- **NMR spectroscopy and conformation.** ^{1}H spectra assignment is required for 3-dimensional structure determination by the methods described for primary structure determination. The solution NMR method gives the solution conformation which may be a more or less rigid structure, but if it is flexible, the solution structure will be an average over several different conformers. Heteronuclear $^{3}J_{CH}$ coupling correlated with glycosidic dihedral angles following rules similar to those for J coupling of protons; trans conformations with dihedral anges near 180 °give relatively large coupling constants (7 - 8 Hz) while gauche conformations give smaller values. [slide 28] Measurement of $^{3}J_{CH}$ is difficult for natural abundance ^{13}C samples but is possible. $^{3}J_{CC}$ coupling accross the glycosidic linkage is correlated with the same dihedral angles but requires ^{13}C isotope enrichment for practical detection since neighboring ^{13}C atoms occur with a probability of 1 in 10,000. Biosynthetic isotope enrichment can be done in bacterial polysaccharides but for glycoproteins or glycolipids expression in eukaryotic systems is required. At present, isotope enrichement in eukaryotic cells is not very practical.

- **NOE** Much experimental information on oligosaccharide conformation in solution comes from NOE data. [slide 17,29] Nuclear Overhauser effect measures relaxation of a ^{1}H nucleus by dipolar interaction with neighboring nucleii. This effect depends on distance between ^{1}H. Internuclear distances derived from the data can be combined with the molecular modeling method to construct molecular models of oligosaccharides and glycopeptides. This method has been very effective in determining protein structue in solution. Likewise, dihedral angle values deduced from coupling constants can be used as constraints in model building. [C. A. Bush, M. Martin-Pastor and A. Imberty, (1999) "Structure and Conformation of Complex Carbohydrates of Glycoproteins, Glycolipids and Bacterial Polysaccharides" Annu. Rev. Biophys. and Struct. Biol. **28**, 269-293.]

- **Examples** Blood group epitopes (A, B, H and Lewis type) are relatively rigid tri- and tetrasaccharides due to crowded arrangement of the sugar linkages. Evidence comes from molecular modeling, NMR data and from x-ray crystallography.[slide 41] In contrast, the α-(2→3)-sialic acid linkage to β-galactose in the sialyl Lewis^X epitope (selectin receptor) is flexible. [slide 30] Mannose linkages in N-glycopeptides are more flexible.