Determination of structure of glycans and glycan interactions with proteins by NMR
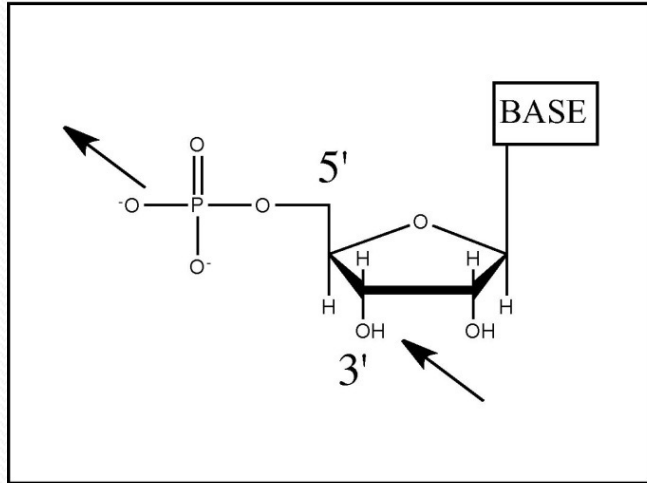
Antonio Molinaro

molinaro@unina.it
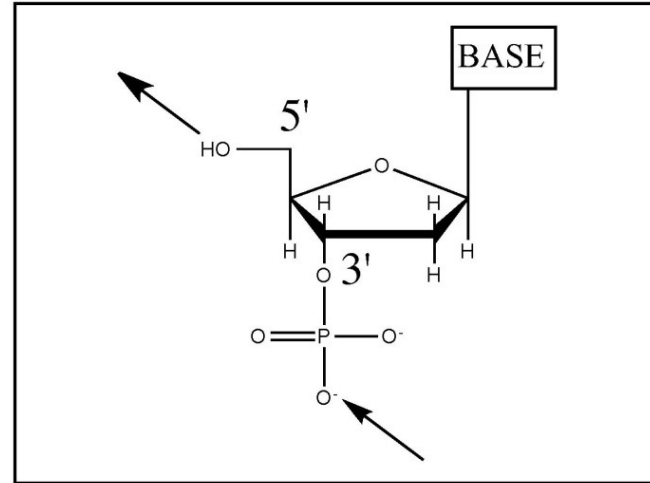
Department of Chemical Sciences, University of Napoli "Federico II"
Department of Chemistry, School of Science, Osaka University

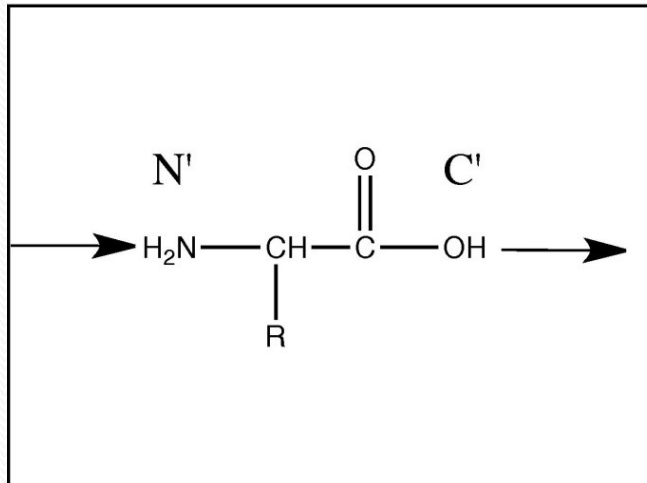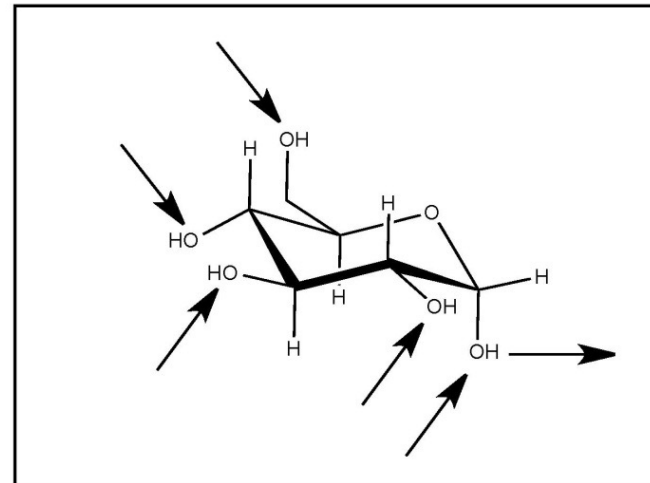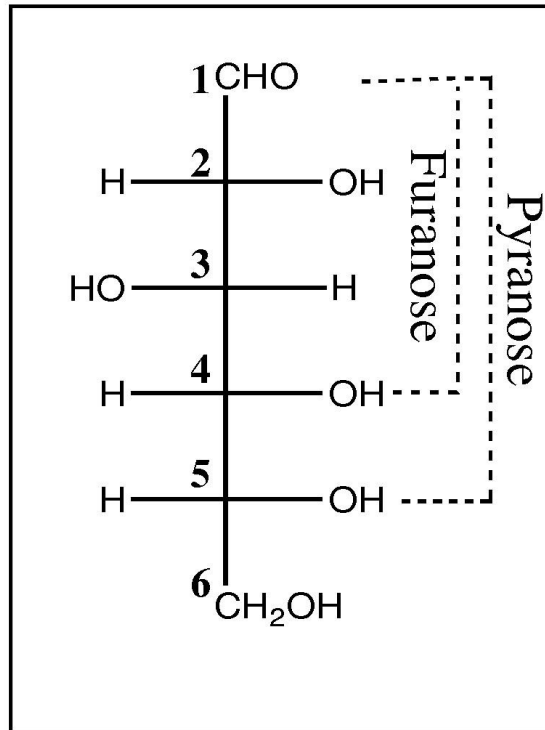# Why glycans are puzzling, why are they not obvious?

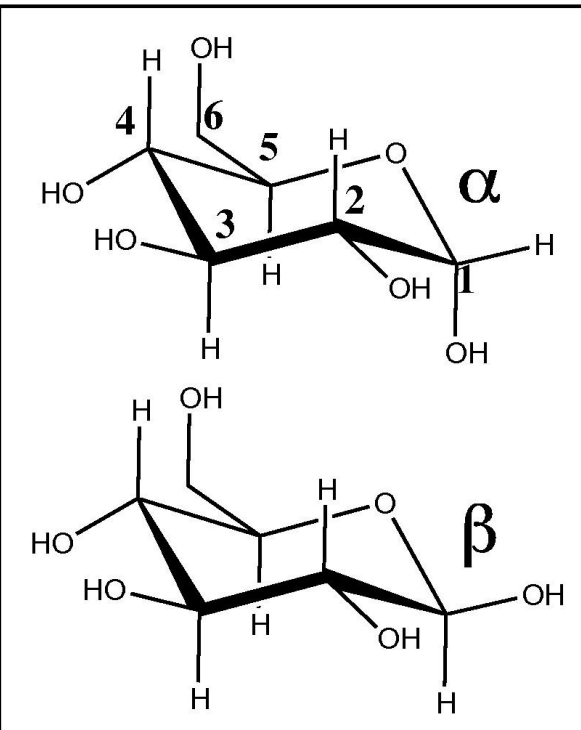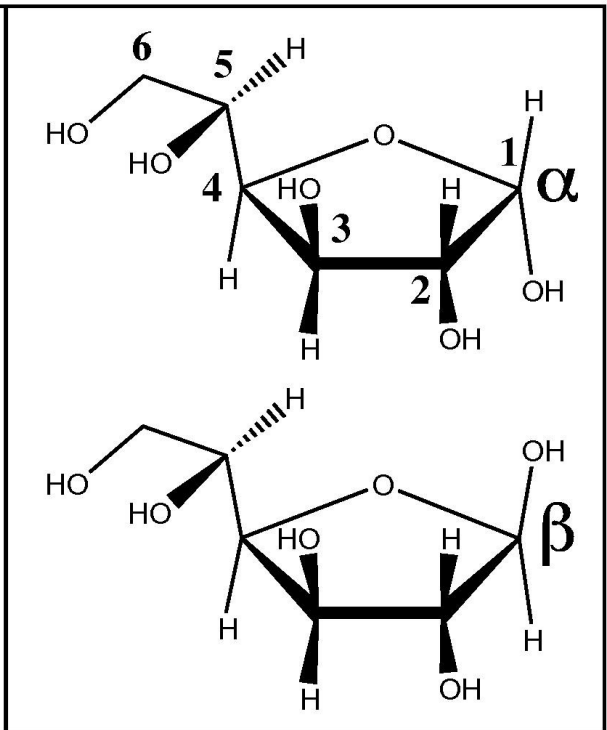**A** Open Form   **B** Pyranose (5→1)   **C** Furanose (4→1)

Open, pyranose and furanose forms of an aldose, showing the many equivalent OH groups. The OH groups can be modified by $-PO_4^{-3}$, $-SO_4^{-2}$, $-NH_2$, $-CO-CH_3$, $-NH-CO-CH_3$, $-H$.

Why glycans are puzzling, why are they not obvious?

Structure determination of a glycan chain: wet organic chemistry for every step

❖ Quali-quantitative analysis (**OC,** GC-MS, NMR)

❖ Absolute configuration (**OC,** GC-MS, NMR)

❖ Size of the ring (**OC,** GC-MS, NMR)

❖ Anomeric configuration (**OC,** NMR)

❖ Linkage analysis (**OC,** GC-MS, NMR)

❖ Monosaccharides sequence (**OC,** MALDI-MS, 2D NMR)

❖ Determination of non-carbohydrate appendages (**OC,** GC-MS, MALDI-MS, 2D NMR)

KEEP CALM IT'S ONLY ORGANIC CHEMISTRY

Typical Applications of NMR:

- **Structural (chemical) elucidation**

  Natural product chemistry

  Organic chemistry: Analytical tool of choice for synthetic chemists.

- **Study of dynamic processes**

  Reaction kinetics.

  Study of equilibrium (chemical or structural).

- **Structural (three-dimensional) studies**

  Proteins

  DNA/RNA

  Polysaccharides

  Drug design

- **Structure Activity Relationships (SAR) by NMR**
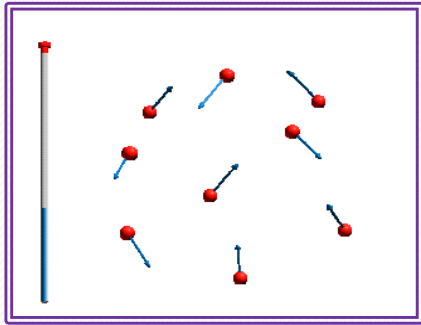
- **Medicine - Magnetic Resonance Imaging (MRI)**

NMR active nuclei

Nucleus (odd atomic number): ^1H, ^{13}C, ^{15}N, ^{19}F, ^{31}P

How does NMR work?

A spinning charge creates a magnetic moment, so these nuclei can be thought of as tiny magnets.

In presence of a magnetic field Magnetic moments precess and orient with or against the field

B_0 (magnetic field)

μ
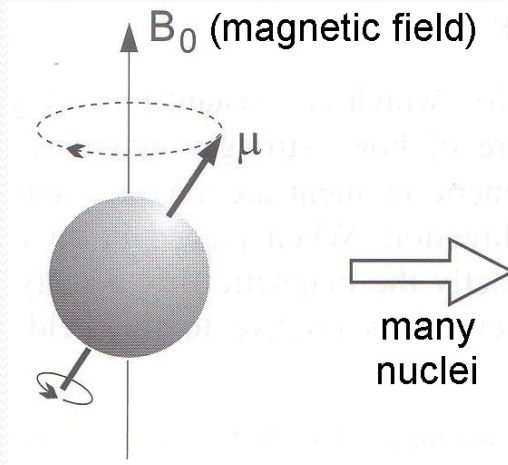
many nuclei

Line up with magnetic field

M_0

z

x

y
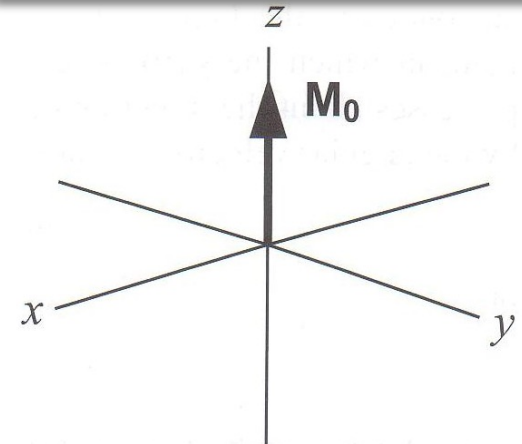
tip using RF pulse

Flip over
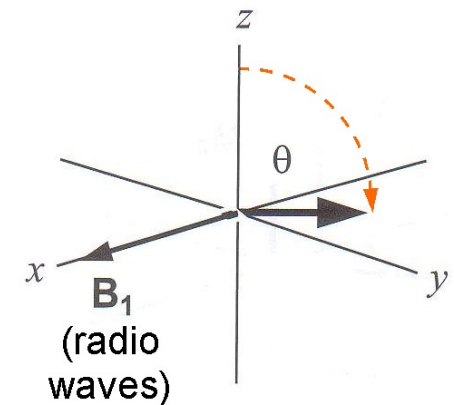
nuclei precess in xy plane

~few seconds

z

θ

x

B_1 (radio waves)

y
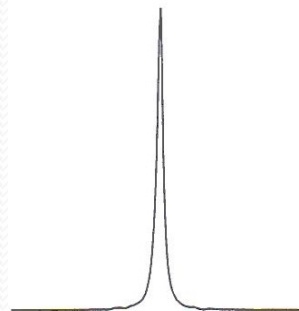
radio signal from nuclei

NMR Signal

Free Induction Decay (FID)
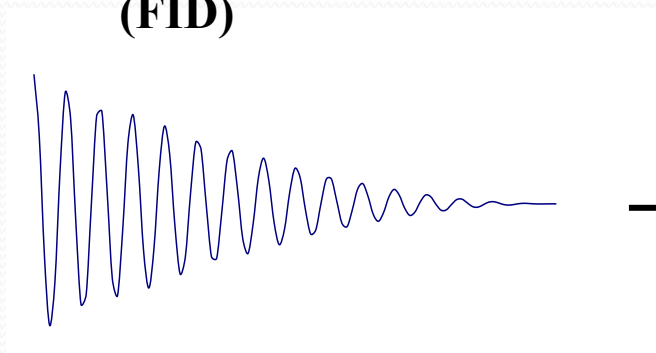
Time

FT $\rightarrow$

NMR Signal
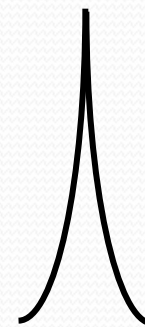
Frequency (Hz or MHz)

$$\nu_1 = \gamma\, B_0/2\pi.$$

$$\nu_1 = \gamma \, B_0/2\pi.$$

Chemical shift (ppm)

¹³C NMR spectrum

Signals

Ref
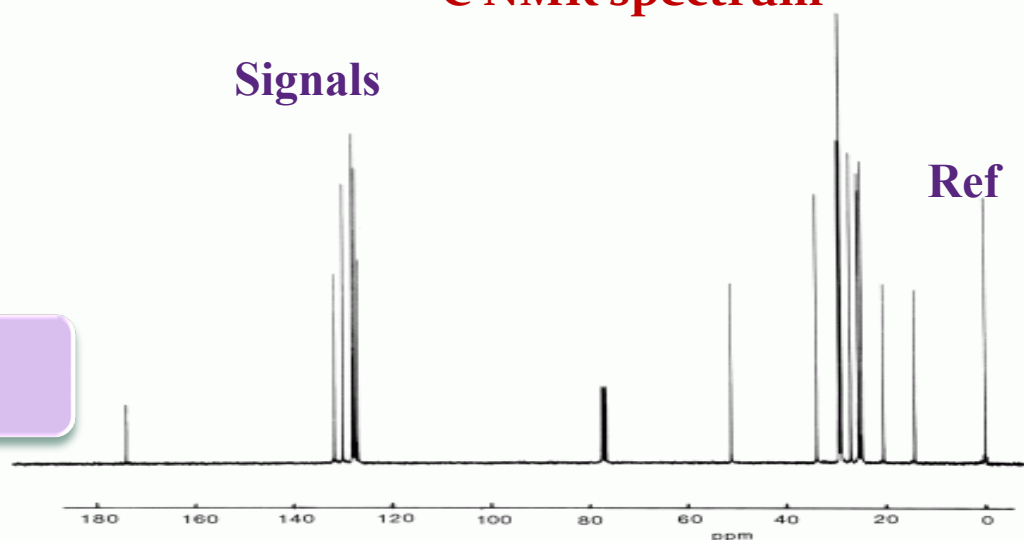
$$\delta \, (ppm) = \frac{\nu_{PEAK} - \nu_{REF} \; (Hz)}{Freq \; of \; the \; nuclei \; (MHz)} = ppm$$

- Depending on the *chemical environment* we have variations on the magnetic field that the nuclei feel, even for the same type of nuclei. It affects the local magnetic field.

¹H chemical shifts
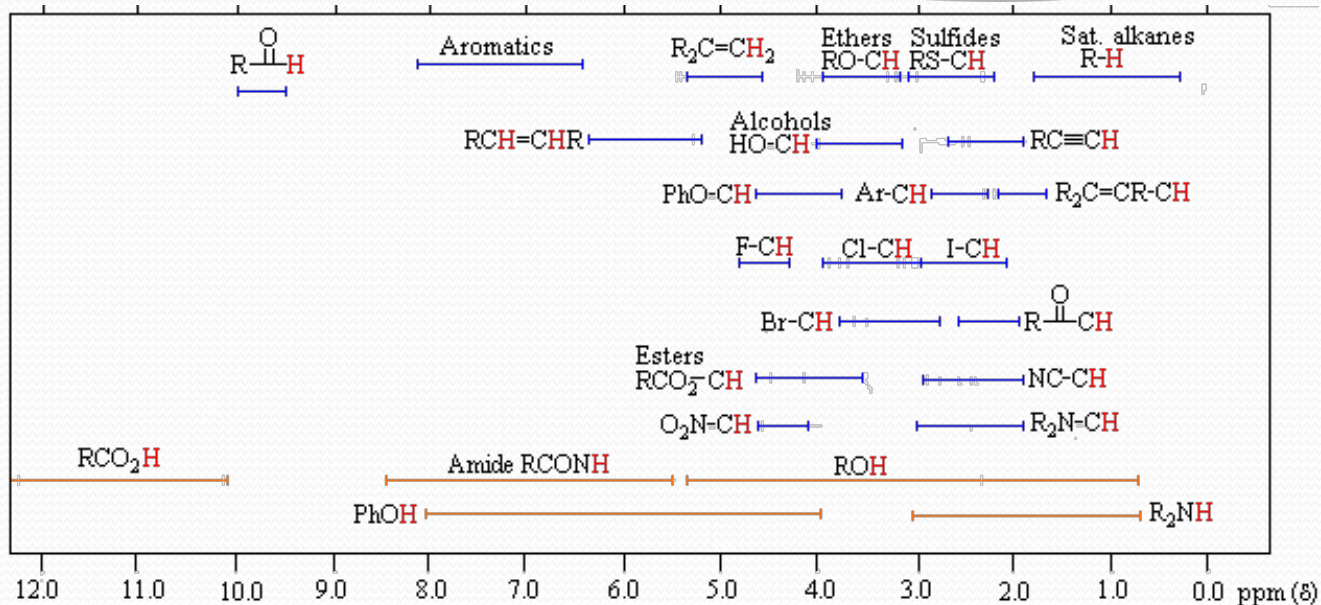
¹³C chemical shifts
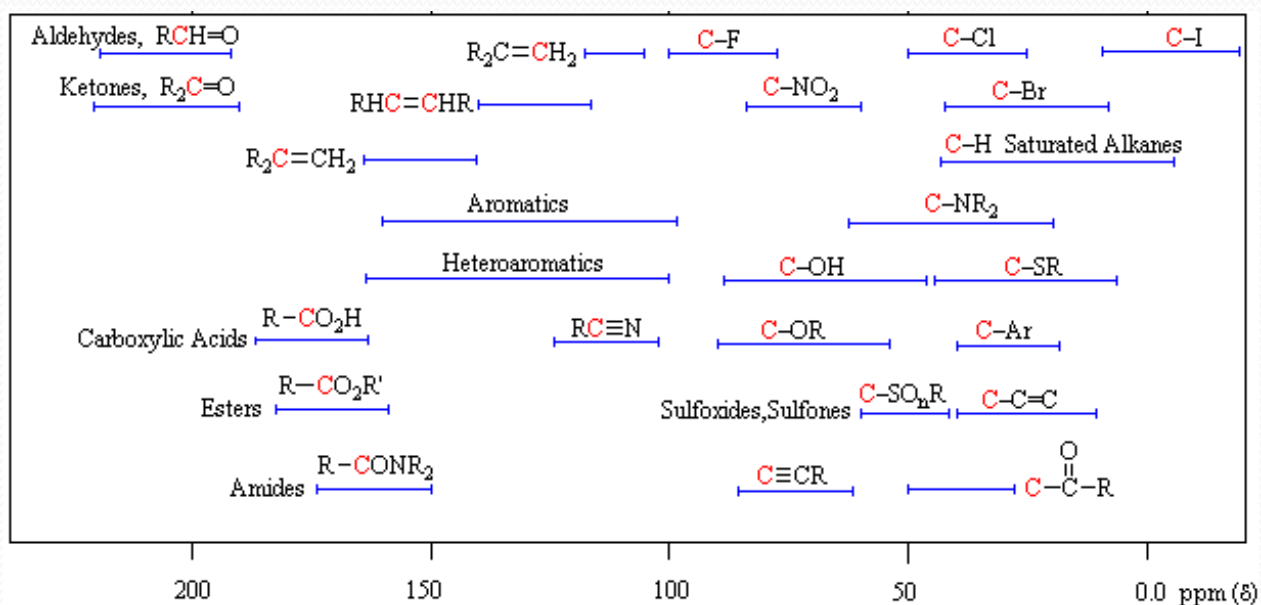
^1H - ^1H Coupling

Signals do not appear as single lines, sometimes they appear as multiple lines. This is due to ^1H - ^1H coupling (also called spin-spin splitting **or J-coupling**).

For this line, H_B is lined up **with** the magnetic field (adds to the overall magnetic field, so the line comes at higher frequency)

For this line, H_B is lined up **against** the magnetic field (subtracts from the overall magnetic field, so the line comes at lower frequency)

H_A

H_B

$-\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-$ H_A H_B
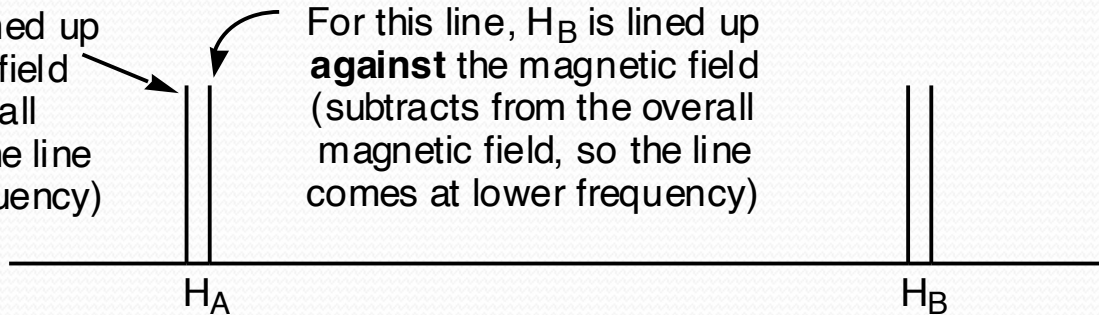
H_A is split into two lines because it feels the magnetic field of H_B.

H_B is split into two lines because it feels the magnetic field of H_A.
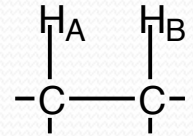
A and B are scalarly coupled nuclei

Doublet: 1 Adjacent Proton
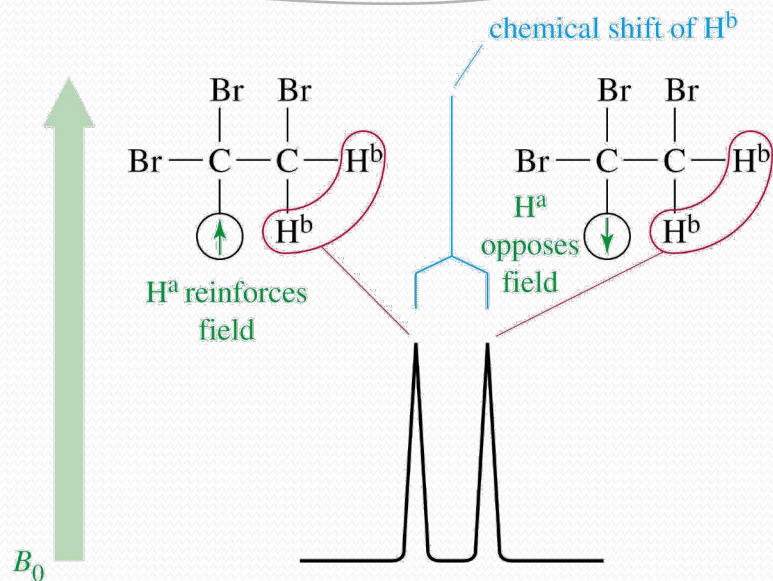
Triplet: 2 Adjacent Protons
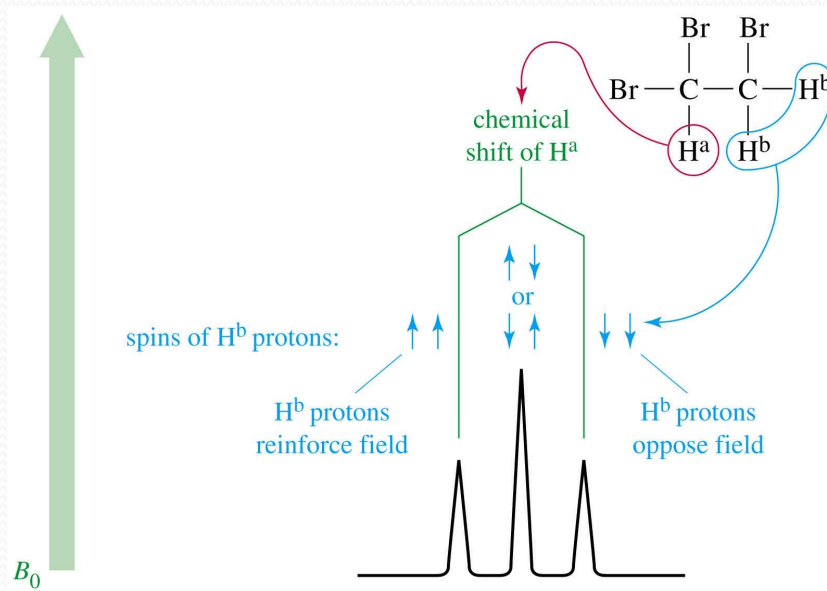
The *N* + 1 Rule

If a signal is split by *N* equivalent protons, it is split into *N* + 1 peaks.
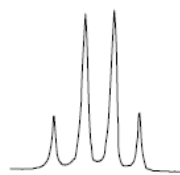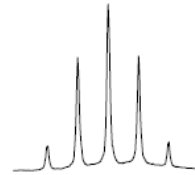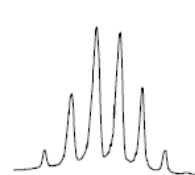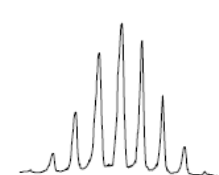
n	2^n	Multiplet Intensities
0	1	1
1	2	1 1
2	4	1 2 1
3	8	1 3 3 1
4	16	Pentet 1 4 6 4 1
5	32	Sextet 1 5 10 10 5 1
6	64	Septet 1 6 15 20 15 6 1
7	128	Octet 1 7 21 35 35 21 7 1
8	256	Nonet 1 8 28 56 70 56 28 8 1

^{1}H NMR

☞ Number of signals (number of non-equivalent H)

☞ Chemical shift (differences in chemical environment)

☞ Splitting or Coupling (number of neighboring H)

☞ Integration (relative number of H at each signal)

- Chemical shift data - tells us what kinds of protons we have.
- Integrals - tells us the ratio of each kind of proton in our sample.
- ^1H - ^1H coupling - tells us about protons that are near other protons.

- **Number of signals**

- **Chemical shift**

- **Splitting or coupling**

- **Integration**

$CH_3CH_2CH_2Br$

triplet, so n=2
deshielded

triplet, so n=2
shielded

6 lines,
so n=5

3 types of H
middle methylene appears as 6 lines due
to 5 neighbours, 3 + 2

Integrals   2   2   3

11   10   9   8   7   6   5   4   3   2   1   0
ppm
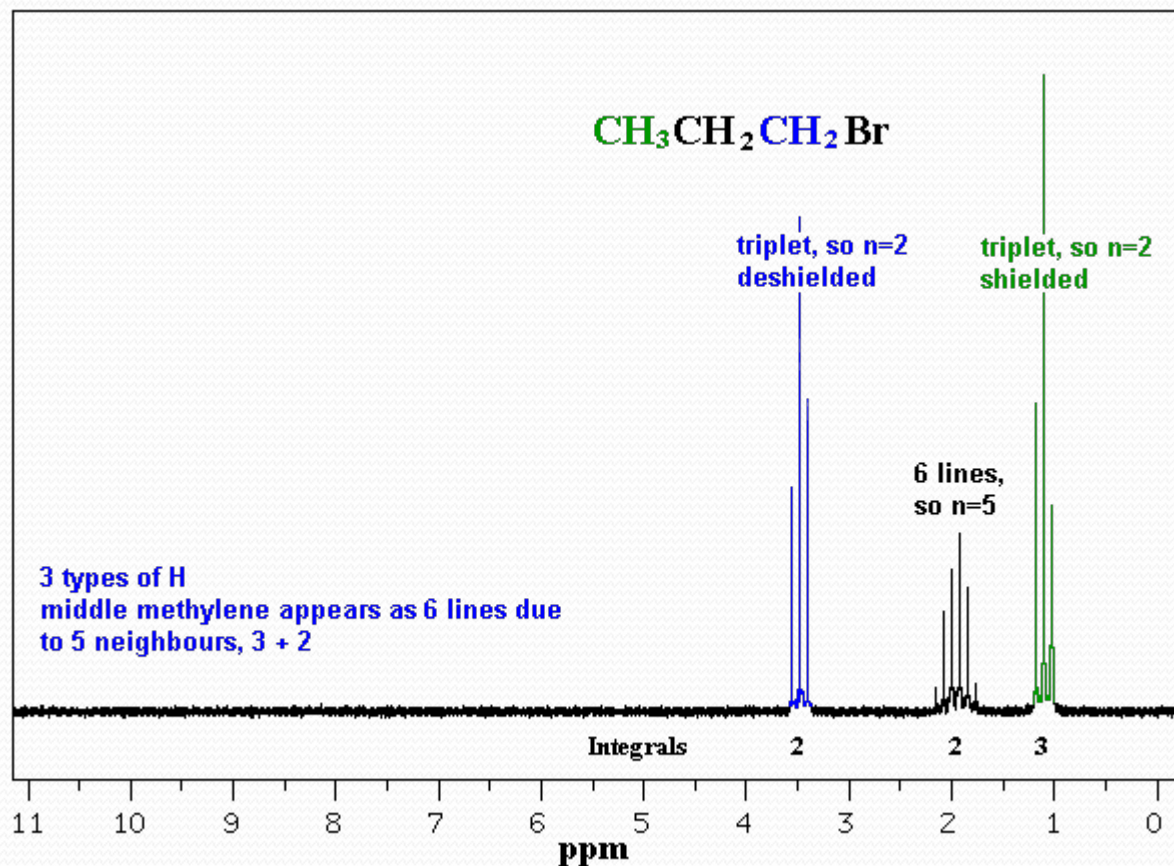
^{13}C NMR

☞ Chemical shift is normally 0 to 220 ppm

☞ Similar factors affect the chemical shifts in ^{13}C as seen for H NMR

☞ Long relaxation times (excited state to ground state) mean no integration

☞ Number of peaks indicates the number of types of C
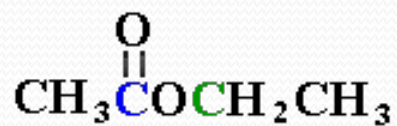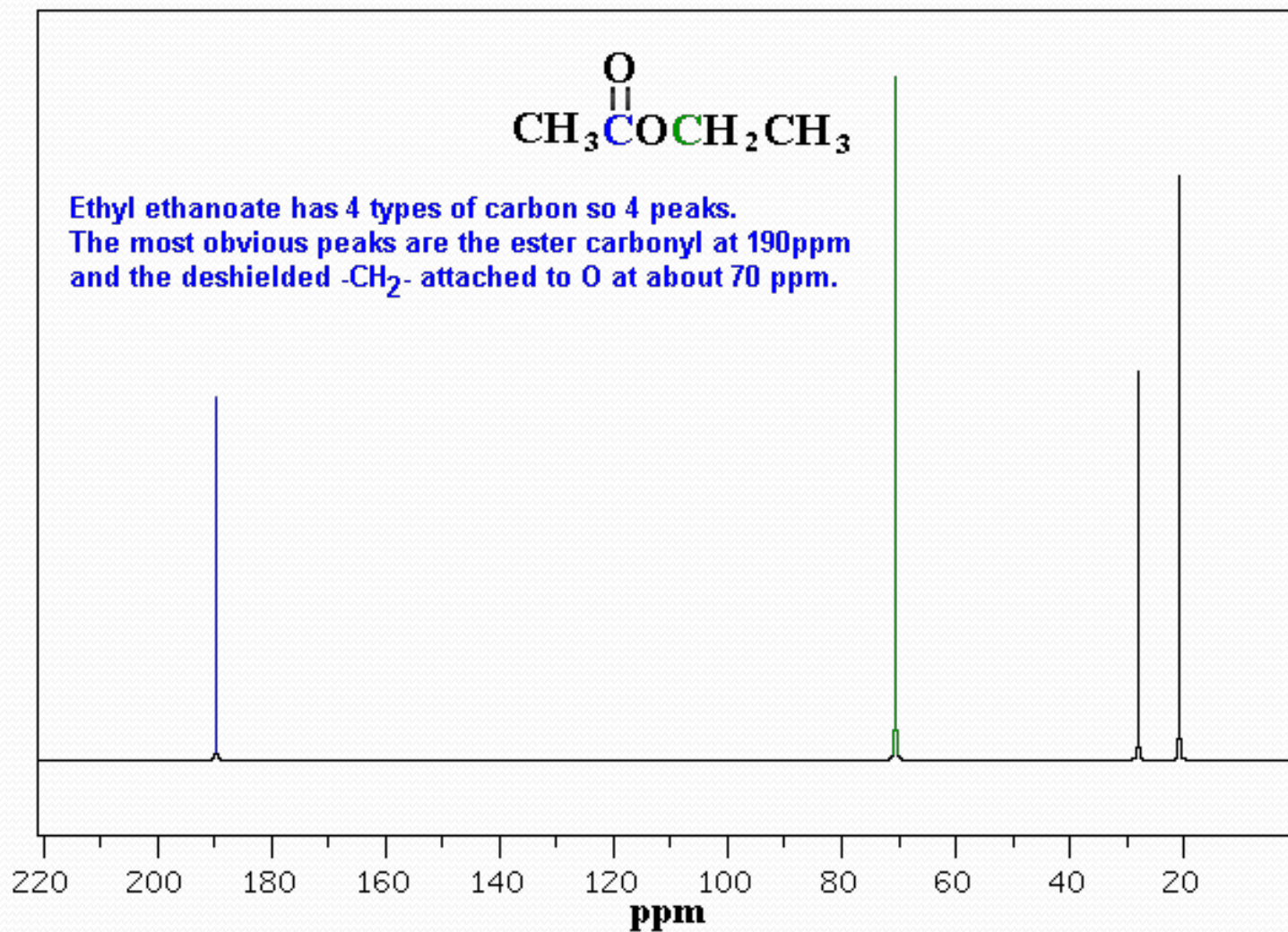
Types of NMR experiments

● 1D Experiments:

- ^1H
- ^{13}C
- ^{31}P

● 2D Experiments:

- COSY (Correlation Spectroscopy)

- TOCSY (Total Correlation Spectroscopy)

- NOESY (Nuclear Overhauser effect spectroscopy)

$^1H - \,^1H$

- HSQC (Heteronuclear Single Quantum Coherence)

- HMBC (Heteronuclear Multiple Bond Coherence)

$^1H - \,^{13}C$

Why glycans are puzzling, why are they not obvious?
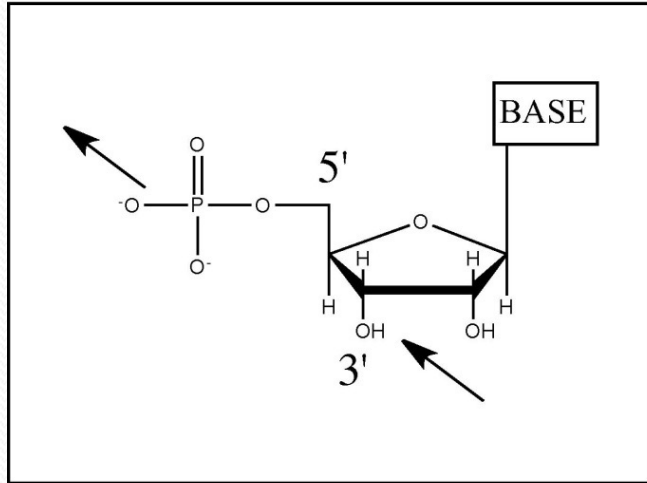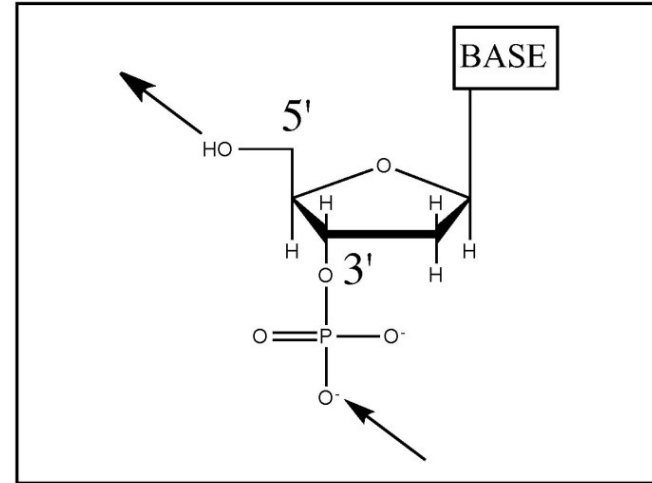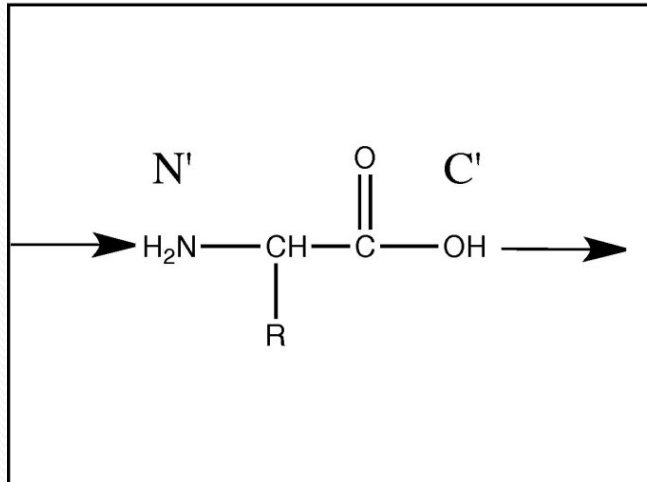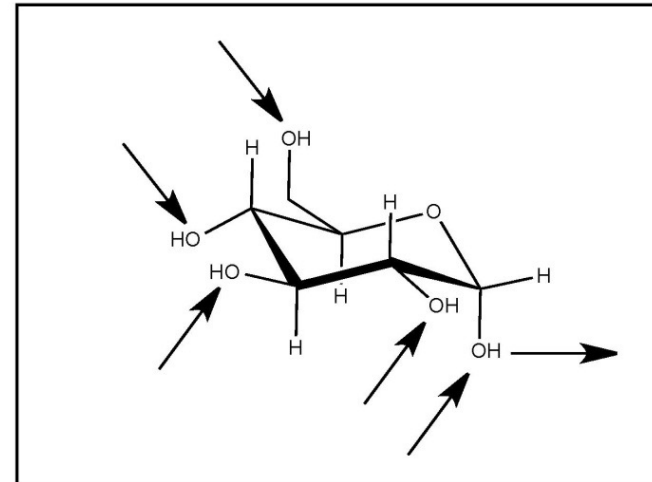
Typical Applications of NMR:

- **Structural (chemical) elucidation**

  Natural product chemistry
  Organic chemistry: Analytical tool of choice for synthetic chemists.

- **Study of dynamic processes**

  Reaction kinetics.
  Study of equilibrium (chemical or structural).

- **Structural (three-dimensional) studies**

  Proteins
  DNA/RNA
  Polysaccharides
  Drug design

- **Structure Activity Relationships (SAR) by NMR**

- **Medicine - Magnetic Resonance Imaging (MRI)**

Structure Determination

- **Each observable NMR resonance needs to be assigned or associated with the atom .**

  - Molecular Formula
  - Functional groups
  - Carbon Connectivity
- Position of substitution on the carbon framework
  - Stereochemical properties.

NMR of carbohydrates:

- Sugar assignment (COSY, TOCSY, HSQC, HMBC)

- Anomeric configuration ($^3J_{H,H}$ coupling constants, NOE)

- Ring size (^{13}C chemical shift, HMBC)

- Non carbohydrate substituents (Homo- and Heteronuclear NMR)

- Sugar sequence: Linkage analysis and glycosylation pattern (HMBC, NOE)

^1H and ^{13}C typical regions of carbohydrates:

The ^1H NMR Spectra can be <u>roughly</u> divided into the following regions:

Anomeric and Acylated Protons : 5.5-4.5 ppm.
Ring Protons : 4.5-3 ppm
Acetyl Groups, Methylene Protons: 3-2 ppm
Methyl Groups: 0.8-2.0 ppm

The ^{13}C NMR Spectra can be <u>roughly</u> divided into the following regions :

Anomeric Carbons Resonate Between 90-105 ppm
Ring Carbons Between 60-85 ppm
Nitrogen Bearing Carbons (In Amino Sugar) 50-60 ppm
Acetyl Groups 20-25ppm
Methylene Protons: 25-35 ppm
Methyl Groups: 15-20ppm

Sucrose

$J = A - B\cos\phi + C\cos2\phi$

β

α
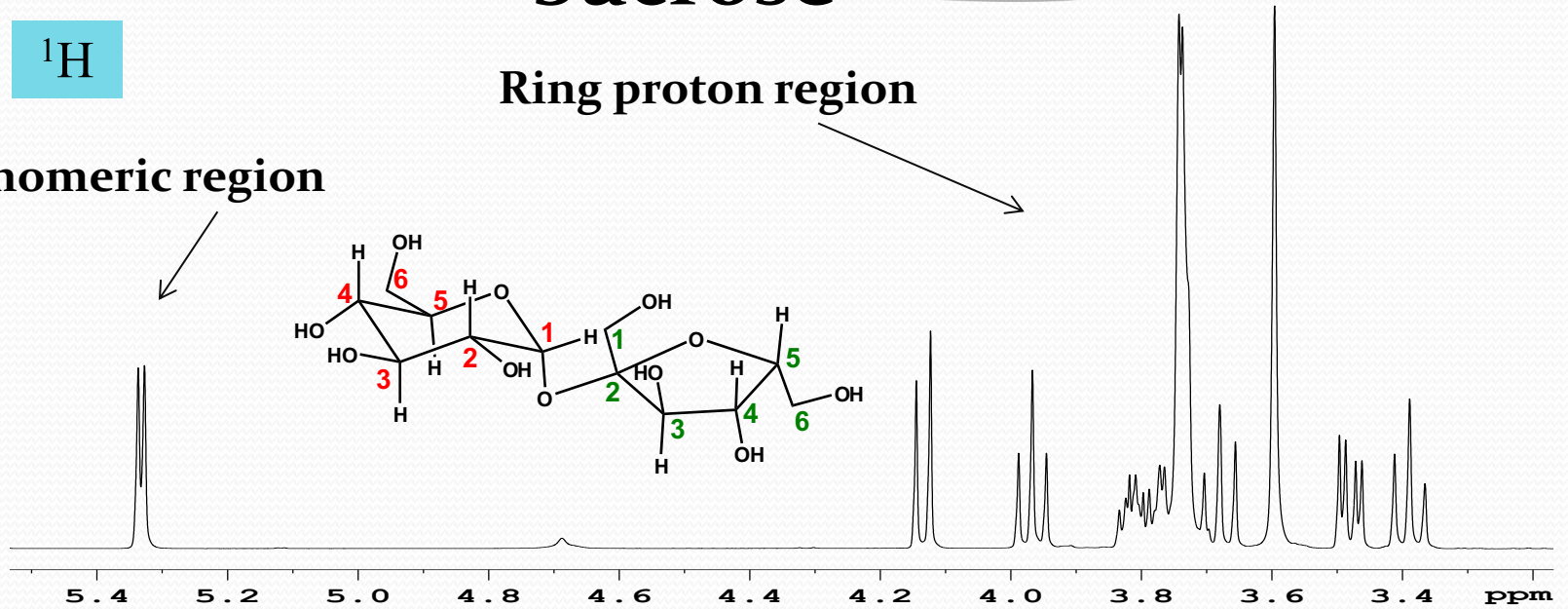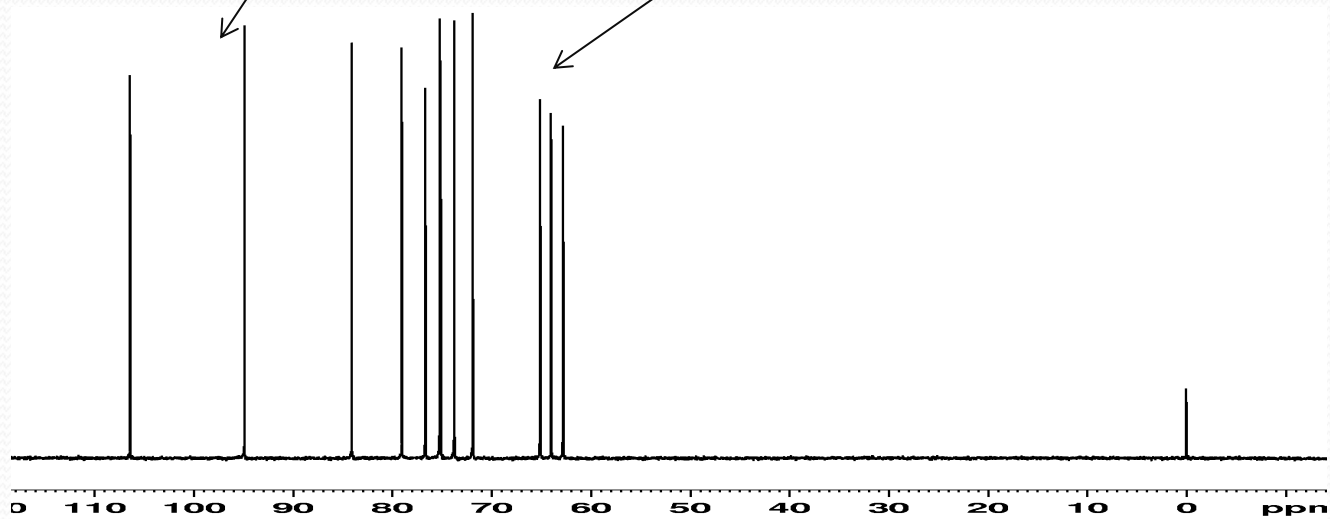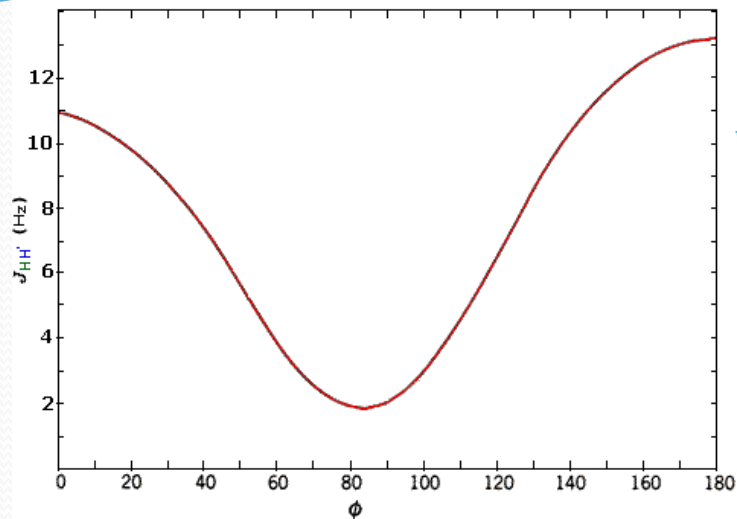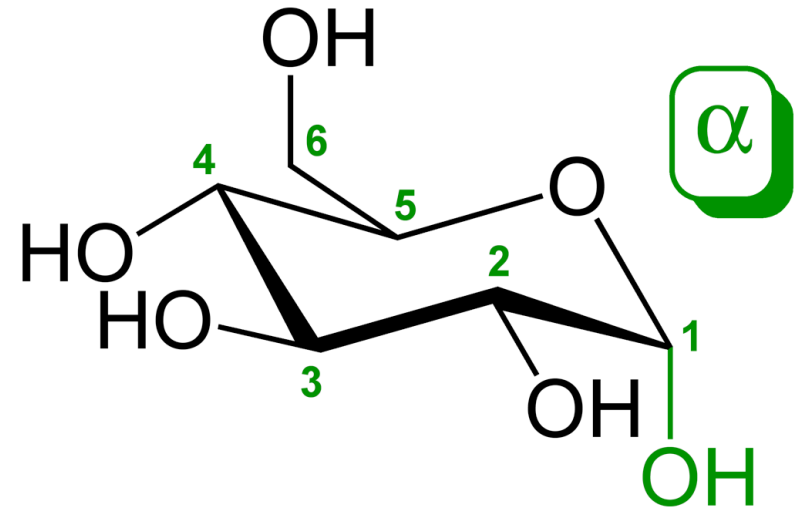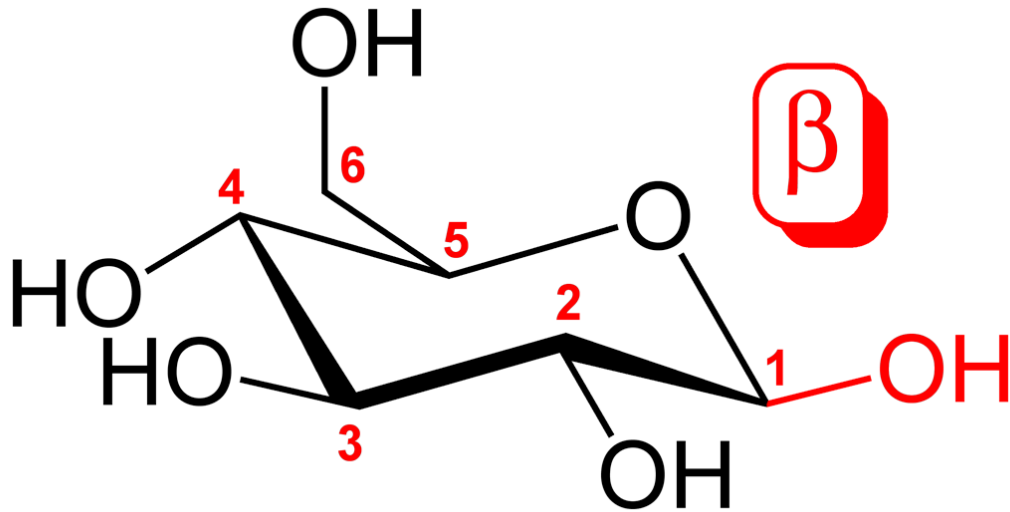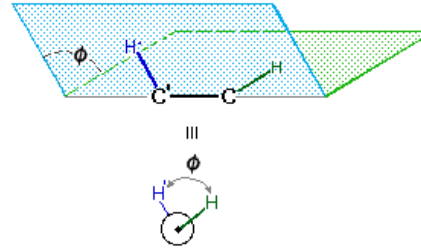
❖**Anomeric configuration**

^1H NMR spectrum contains information on the configuration of glycosidic linkages
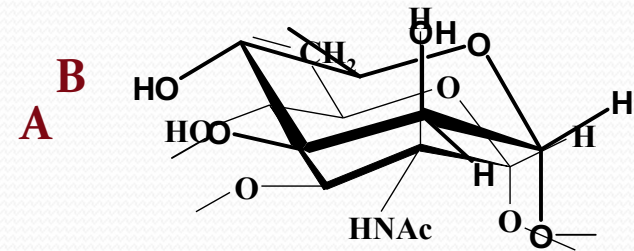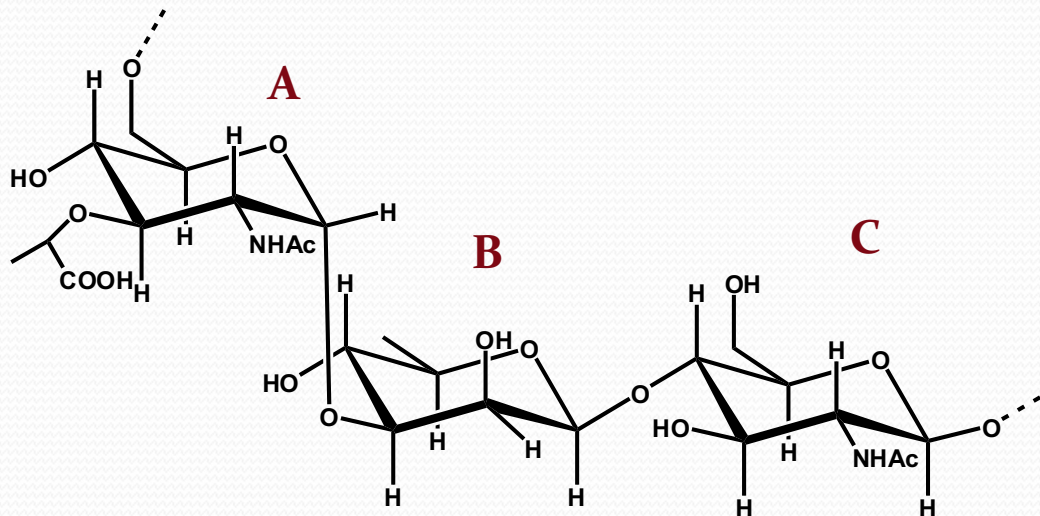
β-linkage: $J_{1,2} < 2$ Hz
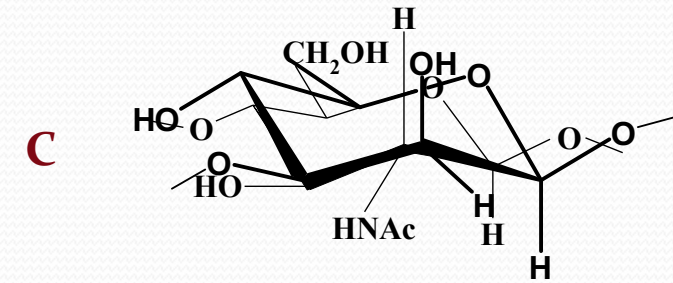
β-linkage: $J_{1,2} > 6$ Hz
α-linkage: $J_{1,2} < 2$ Hz
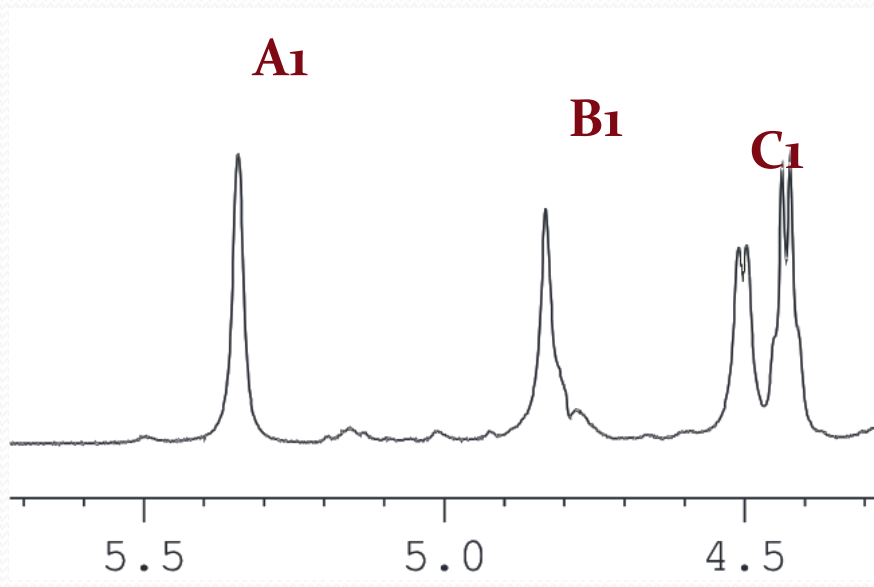
α-linkage: $^1J_{C_1,H_1}$ 170-175 Hz
β-linkage: $^1J_{C_1,H_1}$ 160-165 Hz

Sucrose
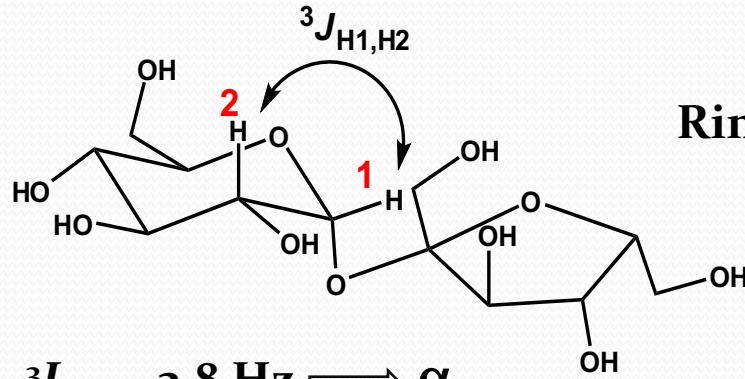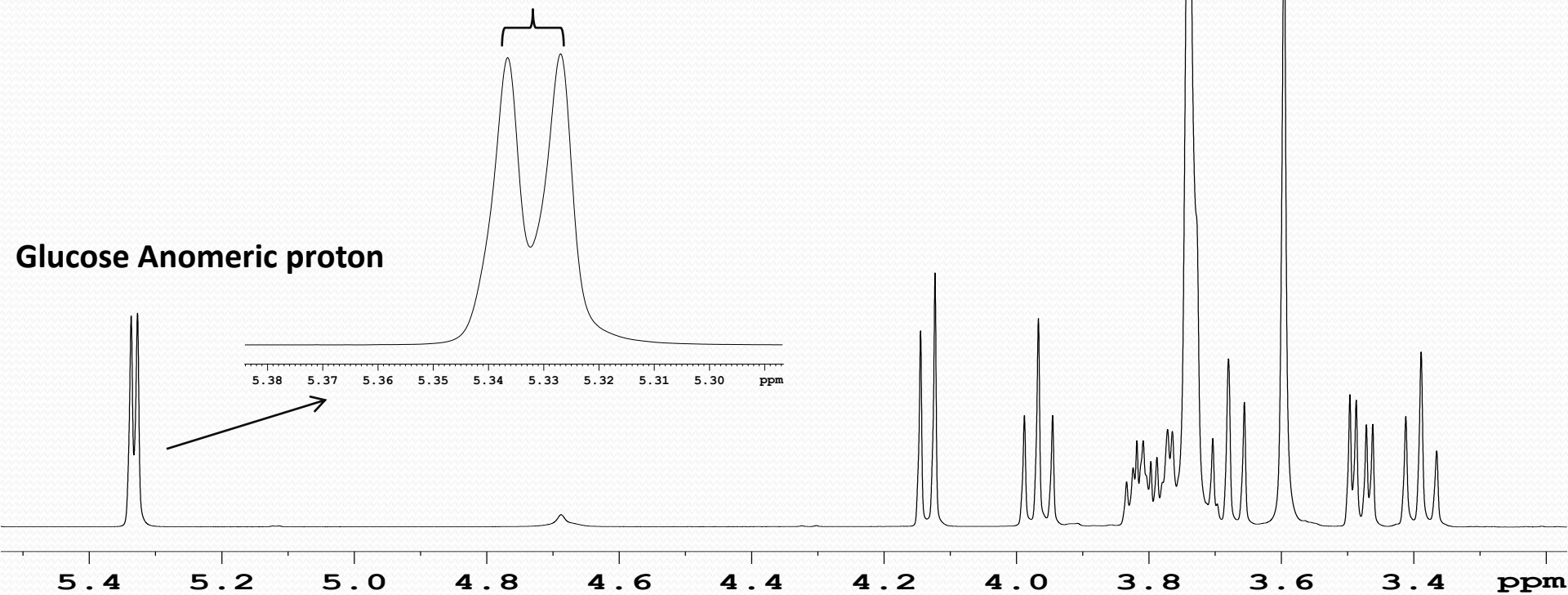
Anatomy of a 2D NMR Experiment
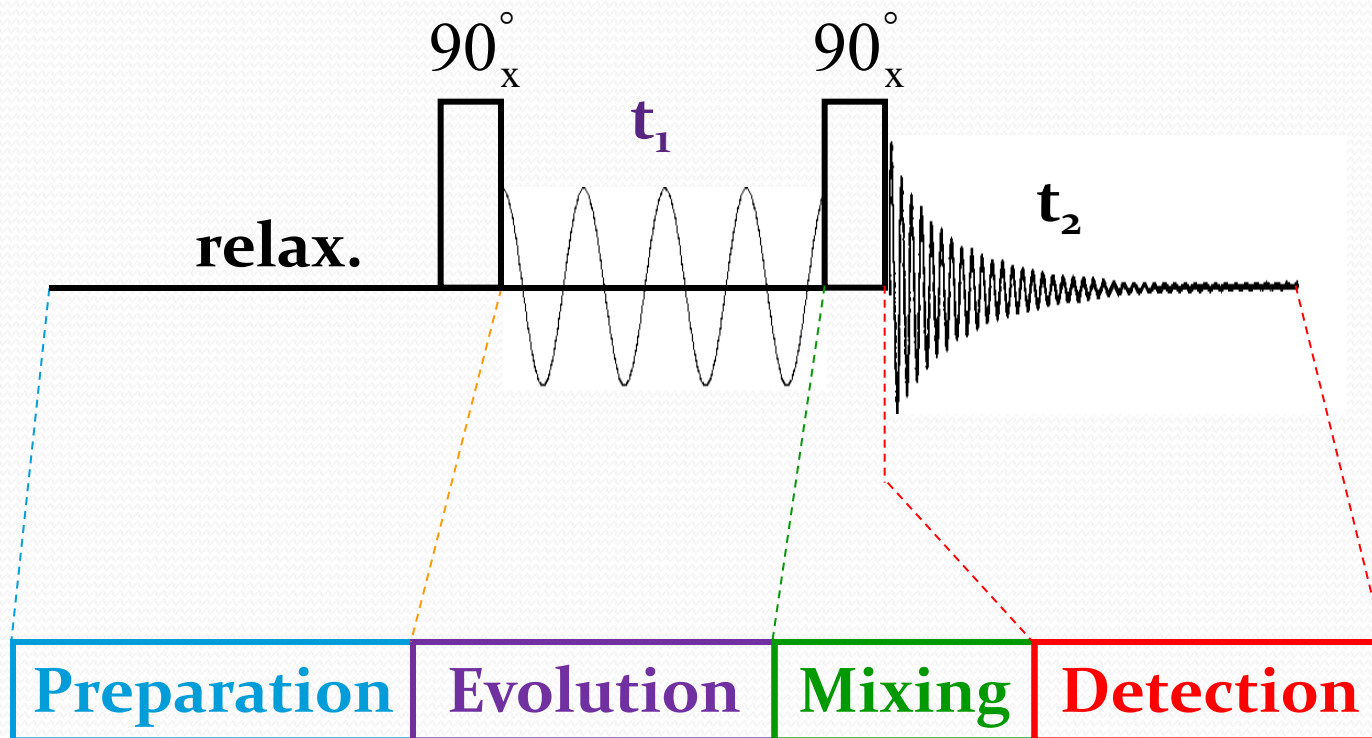
2D NMR - The Interferogram

Interferogram

t_1

f_2
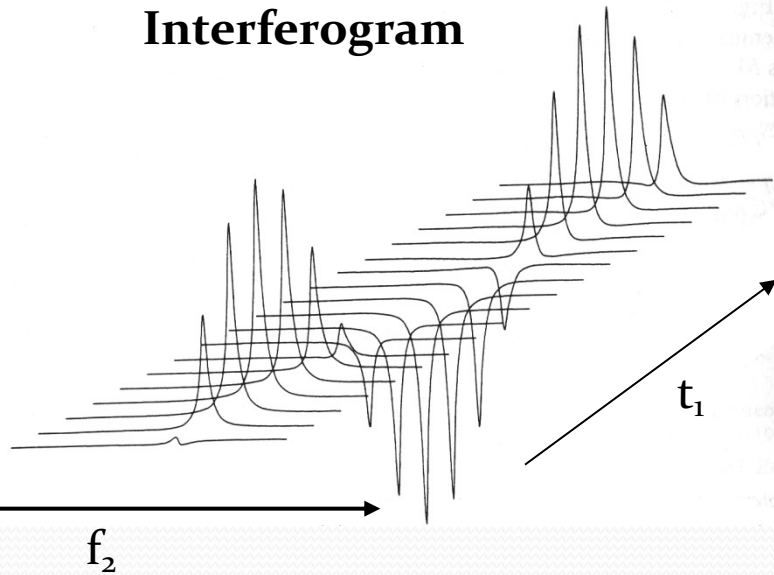
A 2D data set can be thought of as a series of 1D . Each 1D file is different from the next by a change in t_1.

Fourier transformation of each 1D in the t2 domain creates an interferogram.

2D plot of data

f_1

f_2

Contour plot.

f_1

f_2
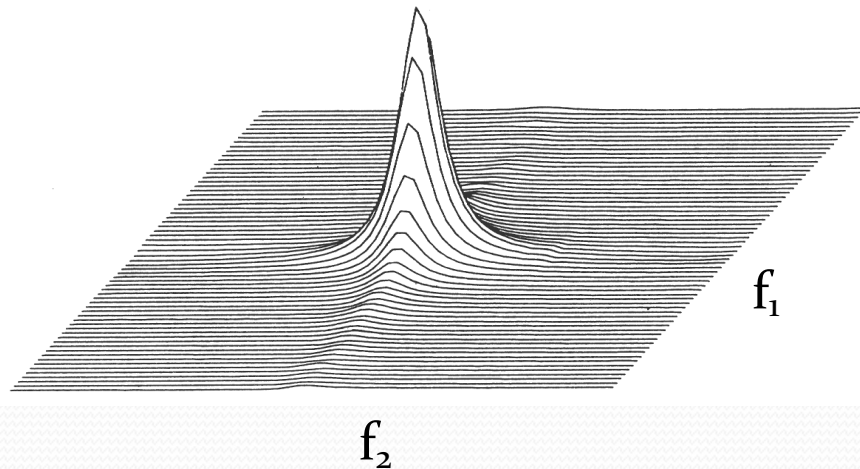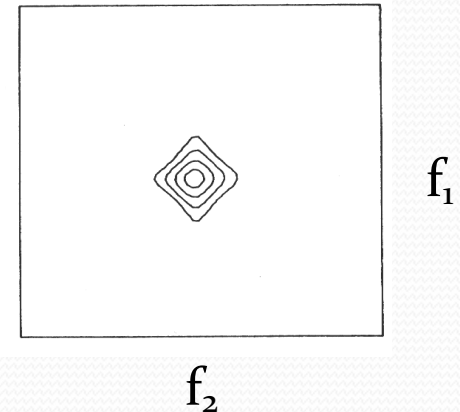
Two Dimensional NMR

✓A 2D data set can be thought of as a series of 1D experiments collected with different timing.

✓Fourier transformation of each 1D in the t2 domain creates an interferogram.

✓The t1 domain is then Fourier transformed resulting in a 2D file with the frequency in each dimension.

✓This 2D file will provide a map of all spin-to-spin correlations

✓Each 2D experiment can provide either through bond (COSY type) or through space (NOESY type) correlation

COrrelation SpectroscopY (COSY)

In a 2D COSY spectrum, **cross-peaks** will exist where there is spin-spin coupling between nuclei.
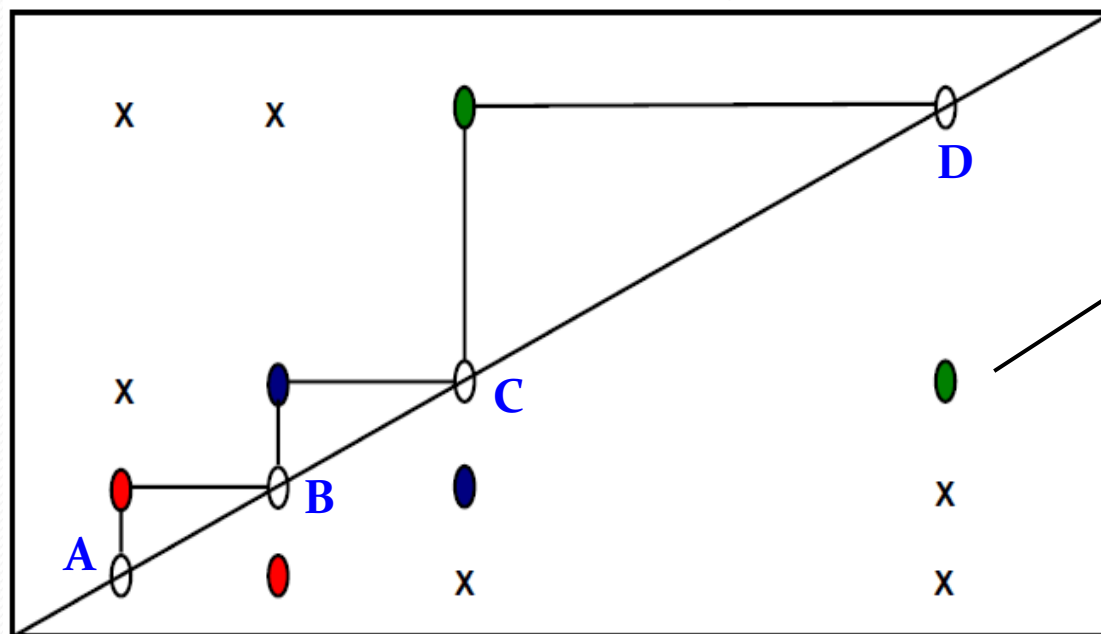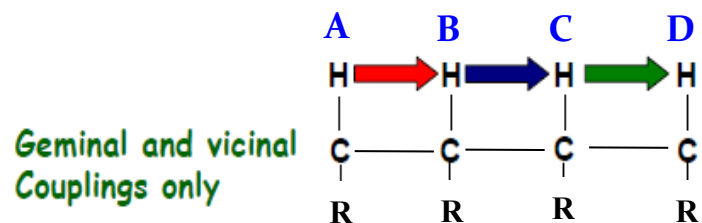
Used to identify spins which are coupled to each other.
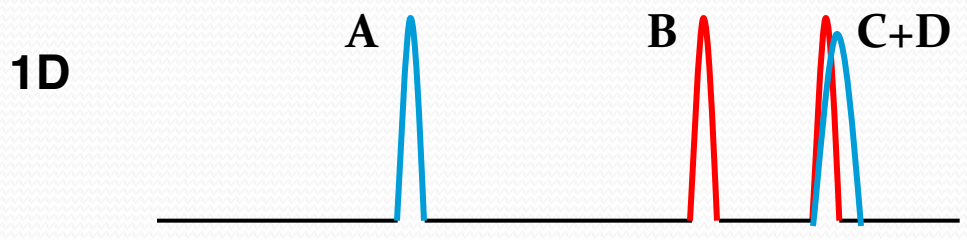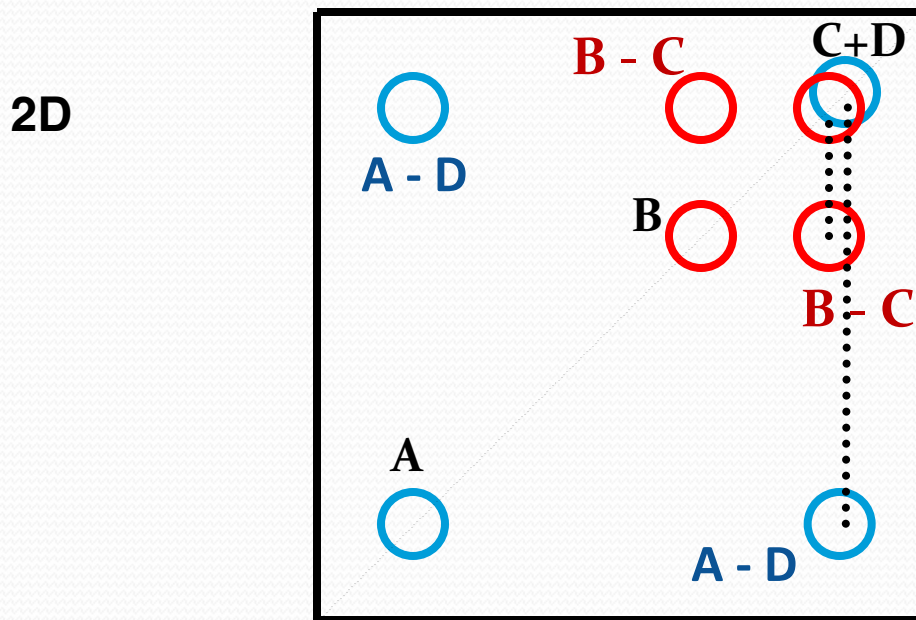
2D Experiments – COSY
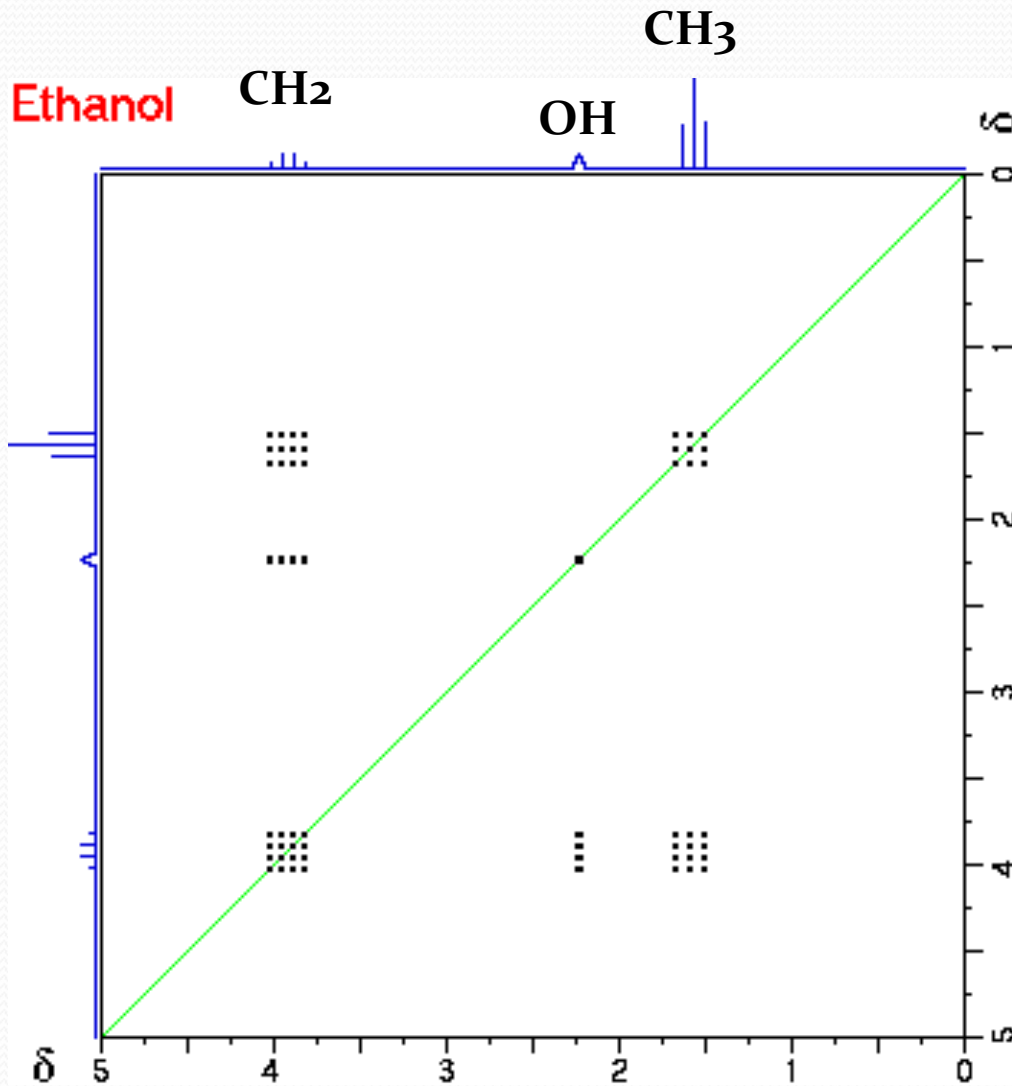
Cross peaks due to geminal and vicinal coupling

CH_3-CH_2-OH

Sugar assignment

COSY of sucrose
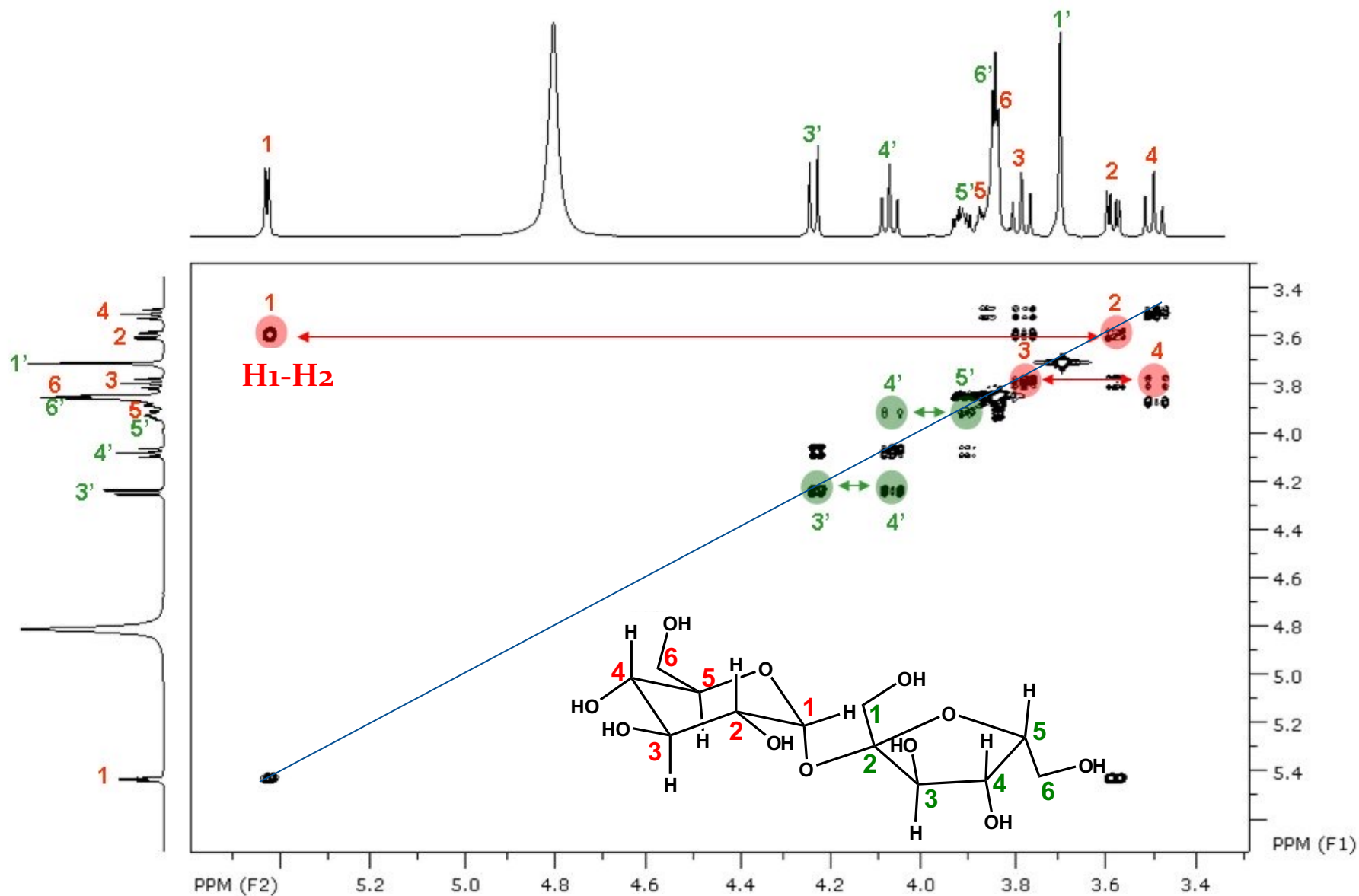
TOtal Correlation SpectroscopY (TOCSY) experiment

TOCSY Experiment

In general, the TOCSY mixing time determines the number of bonds over which signal can be Transferred, assuming that none of the coupling Constants = 0
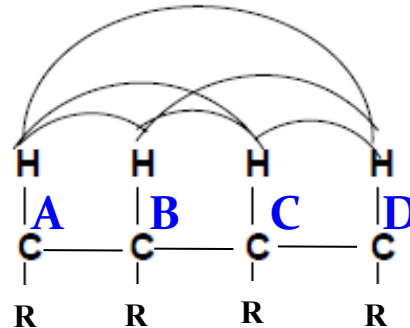
•Cross peaks generated between all members of a coupled spin network
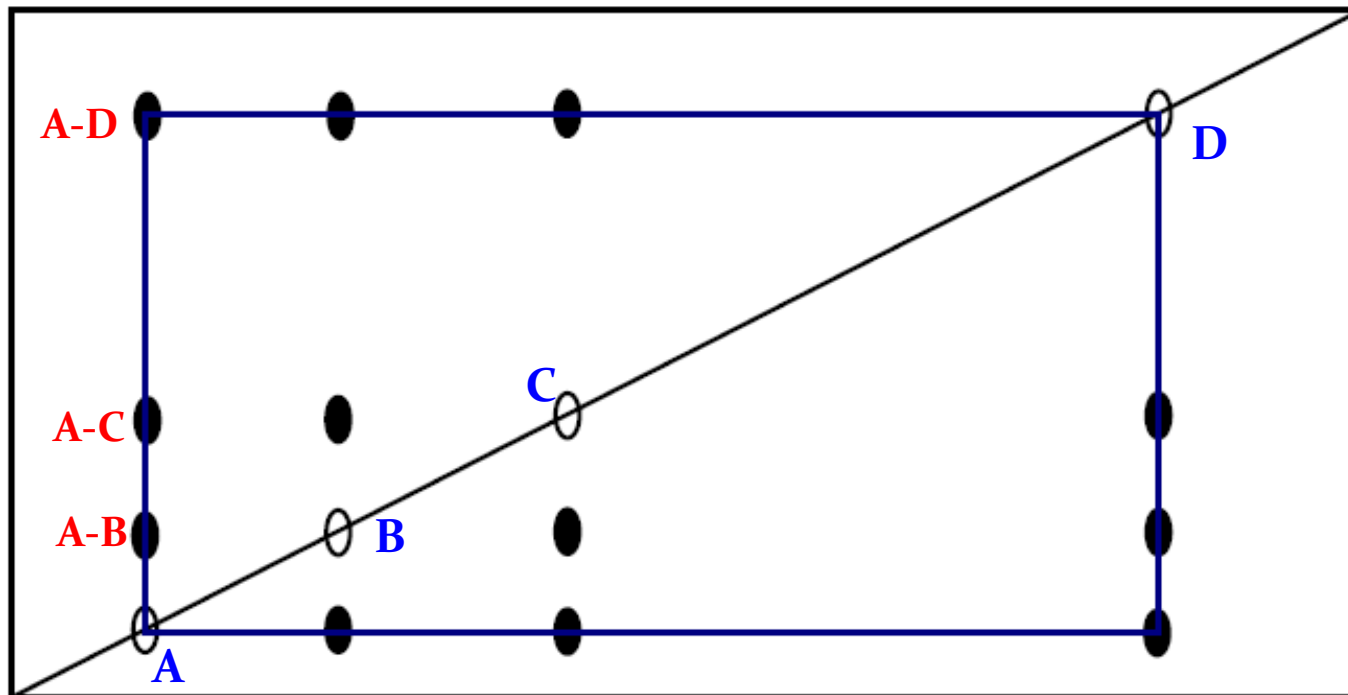
TOCSY

•**Cross peaks are generated between all members of a coupled spin network**
• **Coherence transfer period occurs during a multi-pulse spin-lock period;**
•**Length of spin-lock and J-coupling constants determine how far the spin coupling network will be probed**

COSY

TOCSY

Spin-Lock Pulse (~14ms)
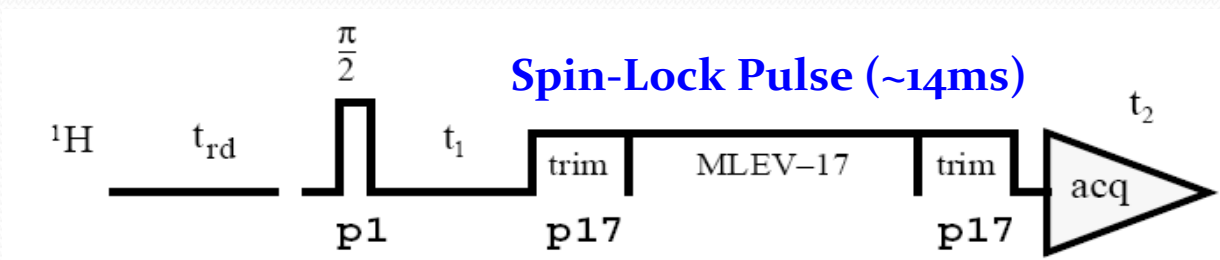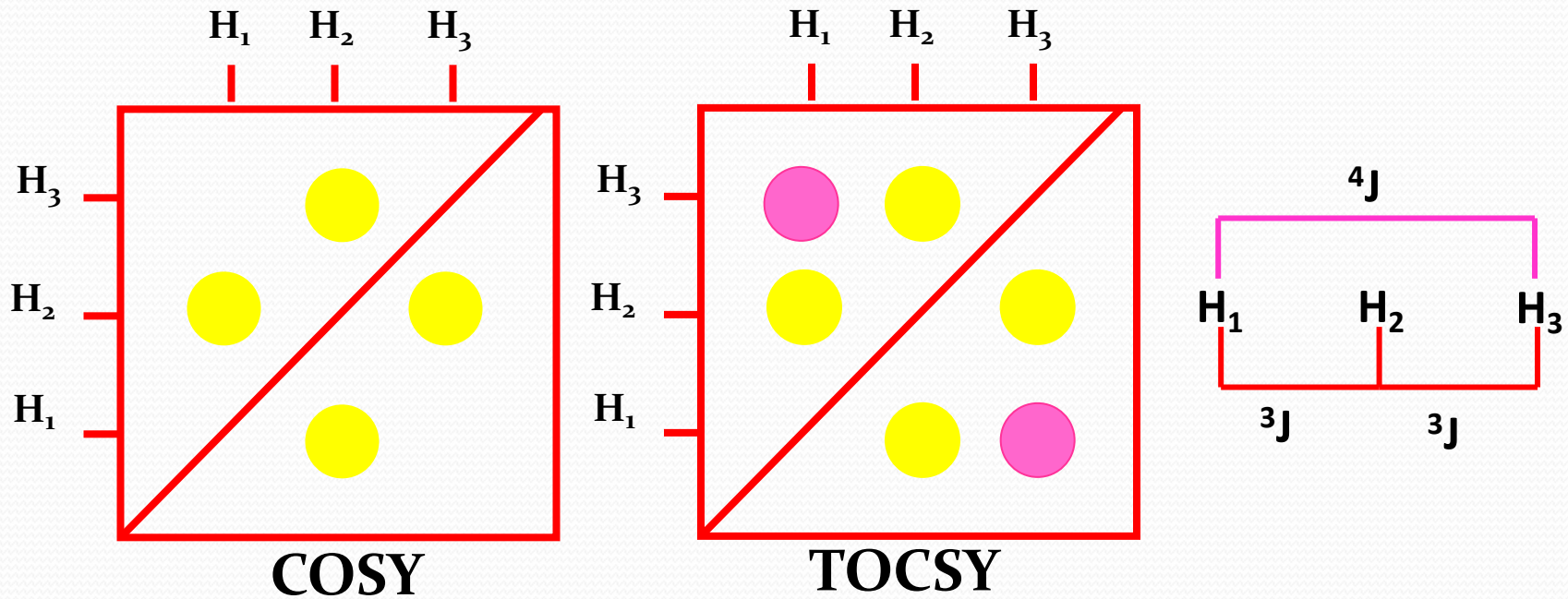
COSY and TOCSY – Sugar assignment
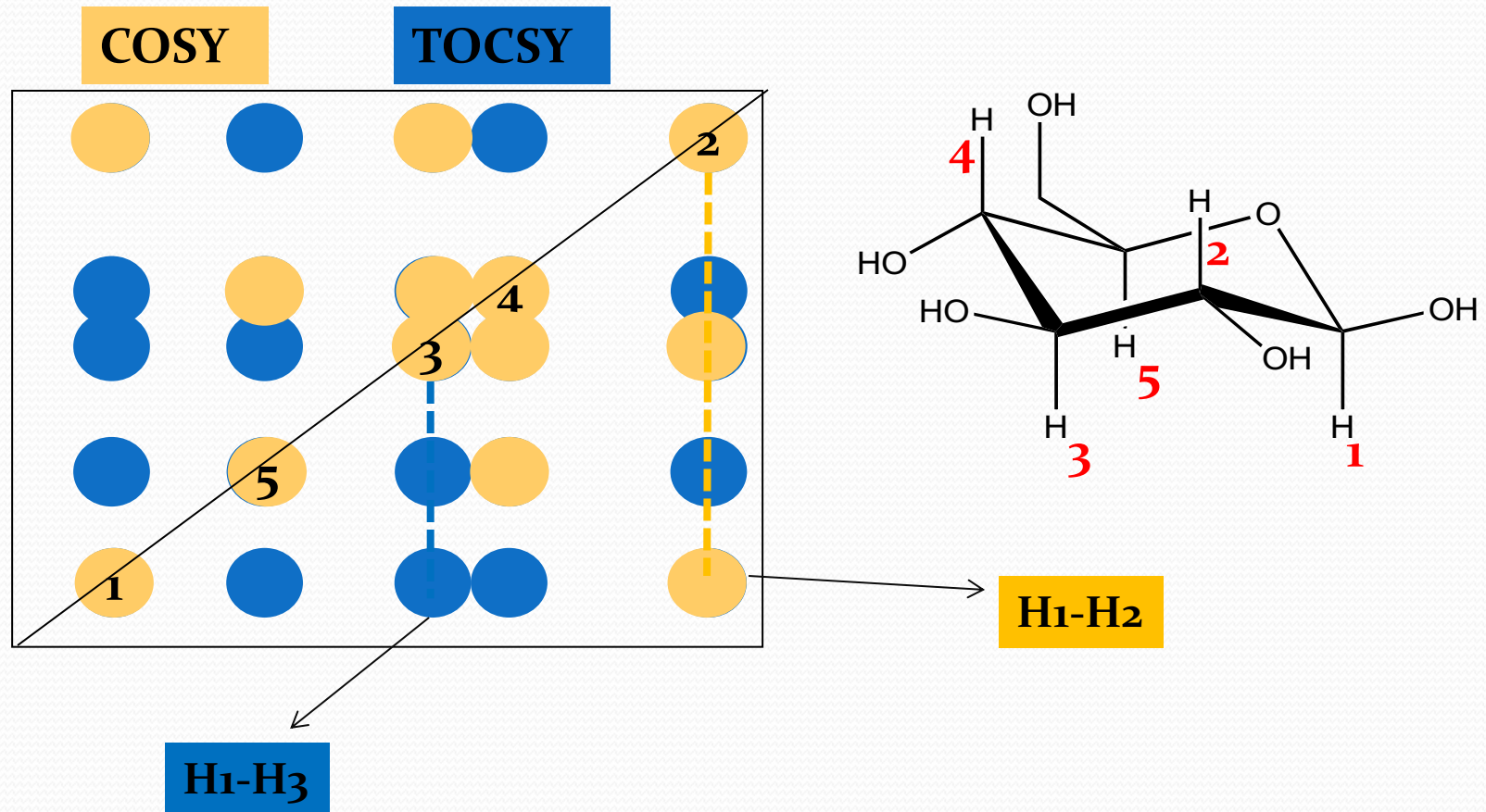
In Glucose, H1 and H2 protons are scalarly coupled, H1 and H3 are not.

In **COSY** spectra $\longrightarrow$ H1 andH2 correlation observed ;

In **TOCSY** spectra $\longrightarrow$ H1 and H3 observed

TOCSY of sucrose
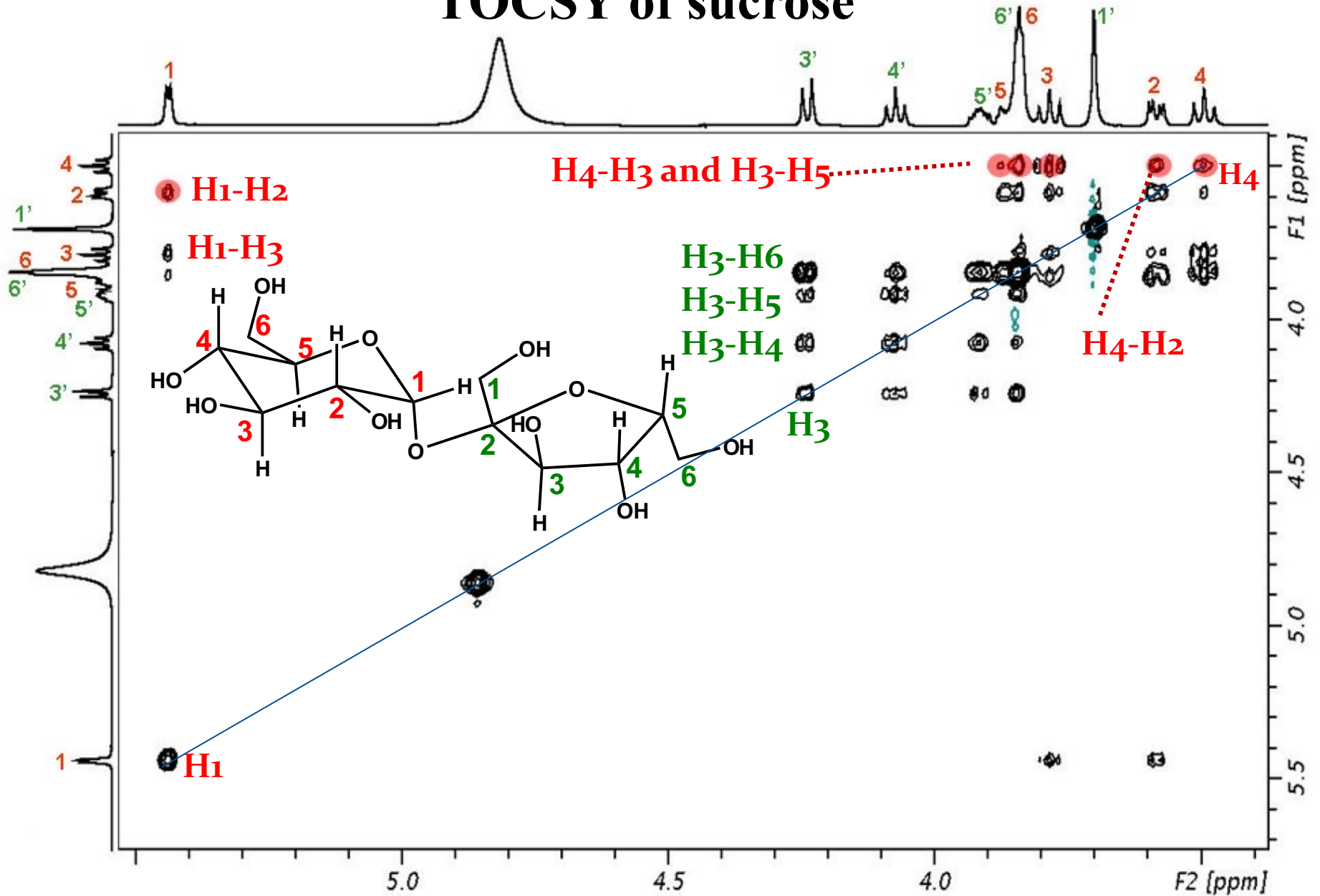
HSQC: Heteronuclear Single-Quantum Correlation

Sucrose HSQC
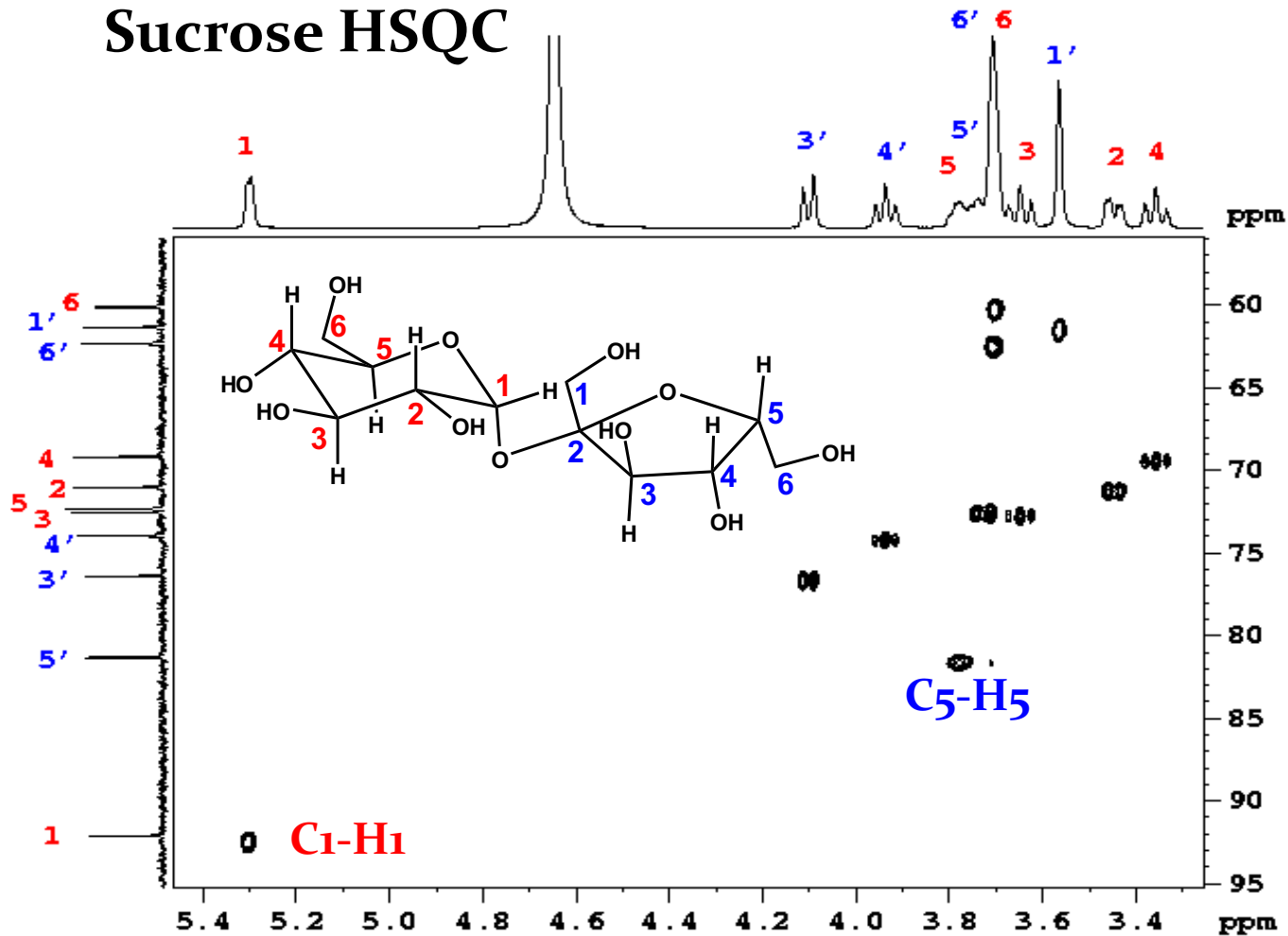
The spectrum contains a peak for each unique proton attached to the heteronucleus being considered.

The **2D HSQC experiment** permits to obtain a 2D heteronuclear chemical shift correlation map between directly-bonded ^1H and X-heteronuclei (an atomic nucleus other than a proton), often ^{13}C or ^{15}N.

HSQC: Heteronuclear Single-Quantum Correlation

¹H–¹⁵N HSQC spectrum of a fragment of the protein NleG3-2.
Each peak in the spectrum represents a bonded N-H pair, with its two coordinates corresponding to the chemical shifts of each of the H and N atoms. Some of the peaks are labeled with the amino acid residue that gives that signal
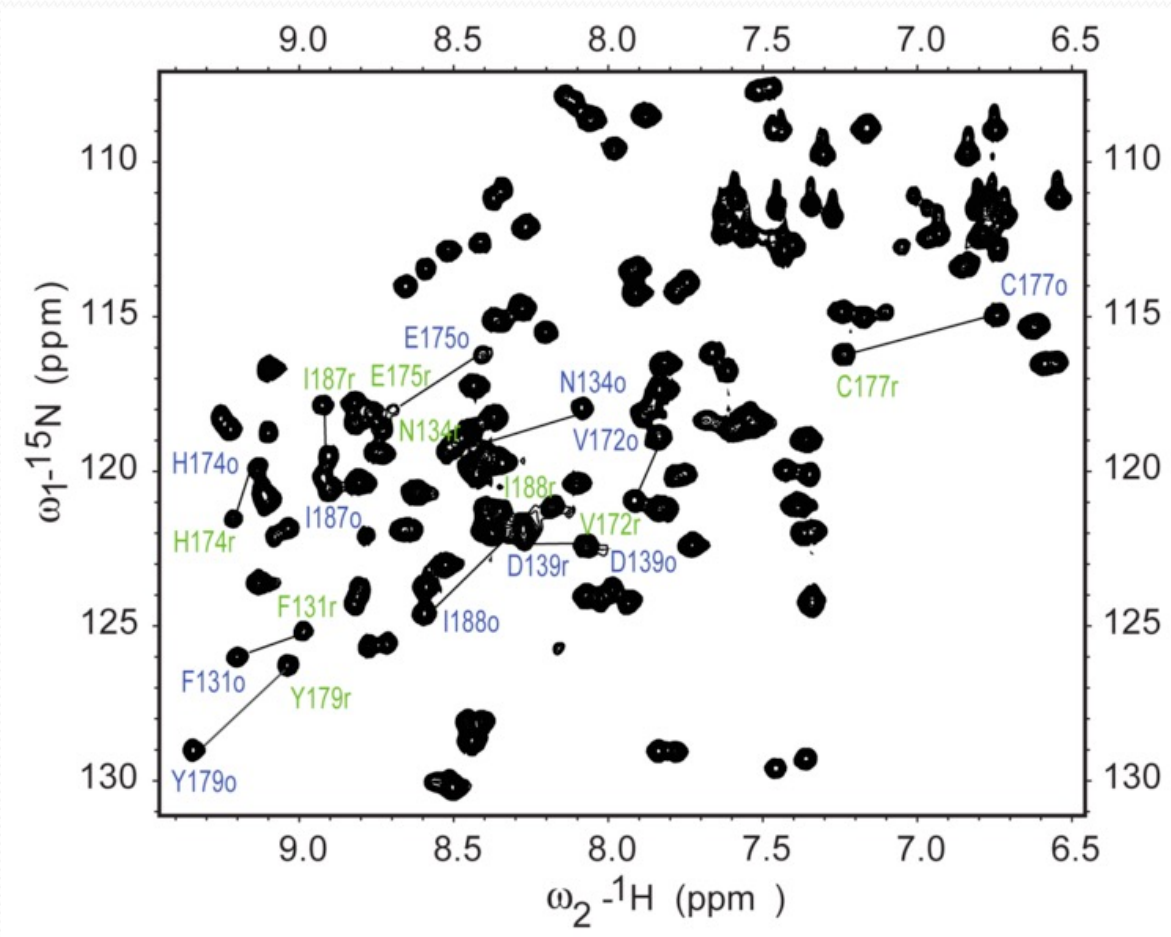
HMBC (Heteronuclear Multiple Bond Correlation)

2D HMBC experiment correlates chemical shifts of two types of nuclei **separated from each other with two or more chemical bonds.**
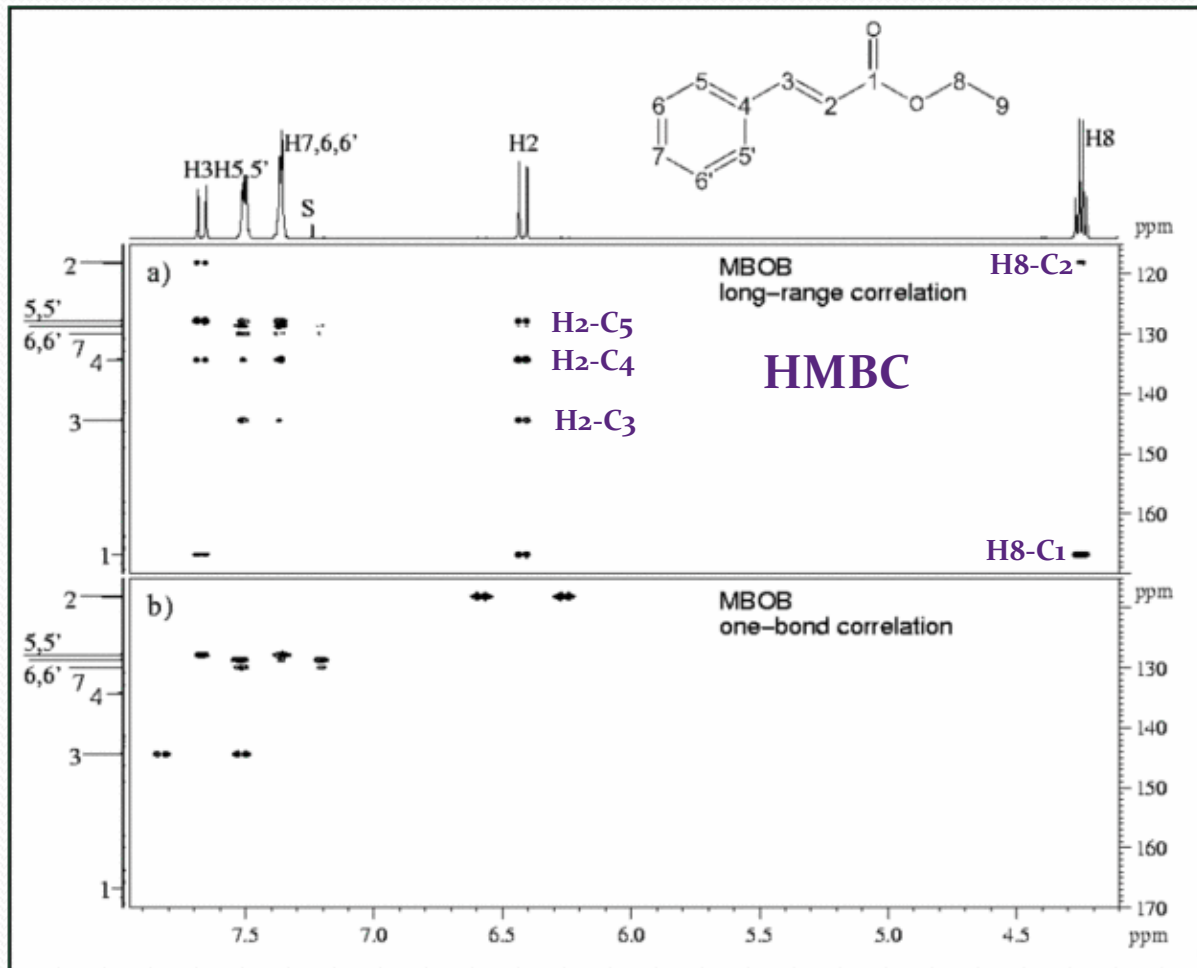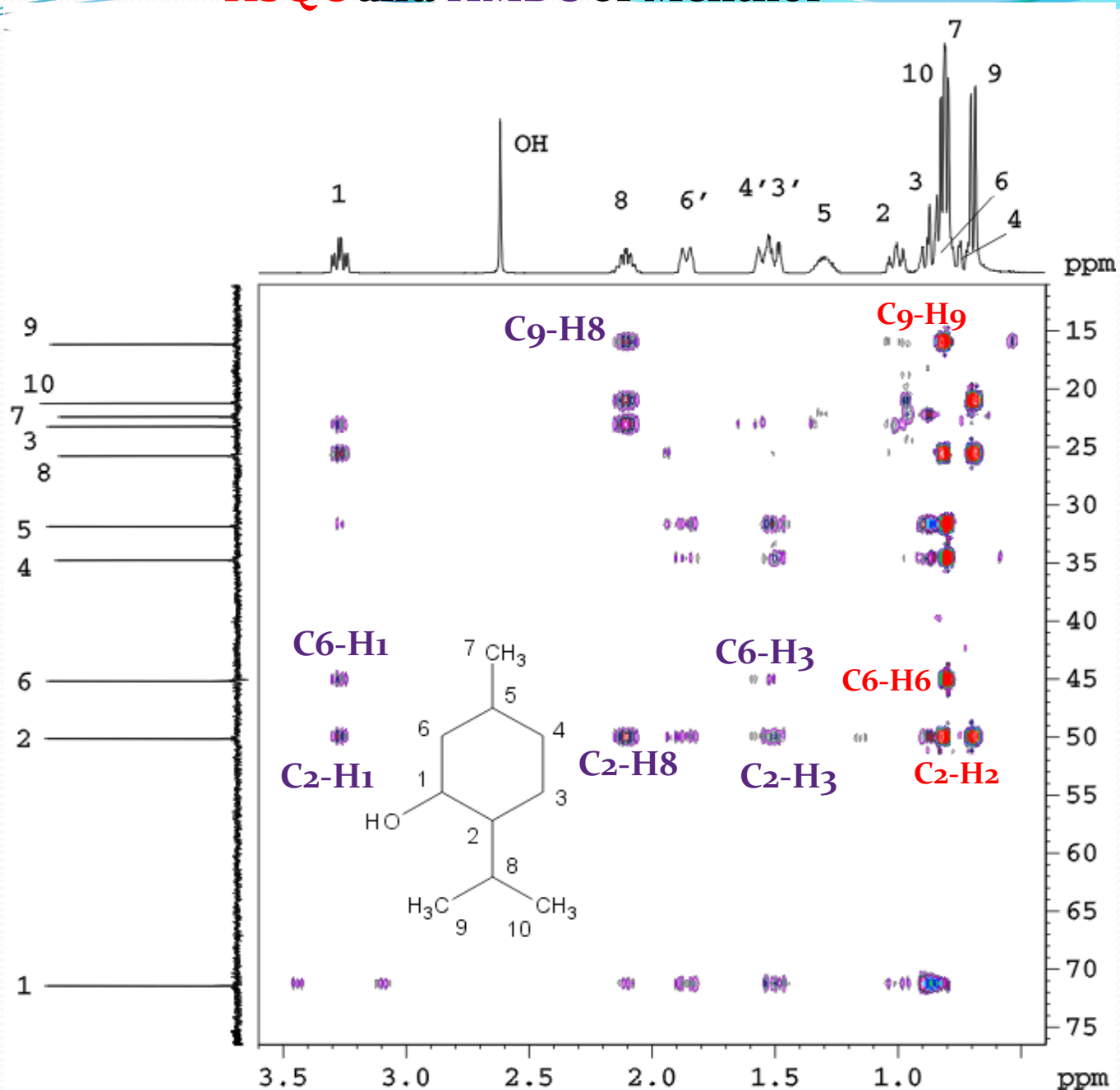
Long range *inter*-residual correlations in the HMBC spectrum
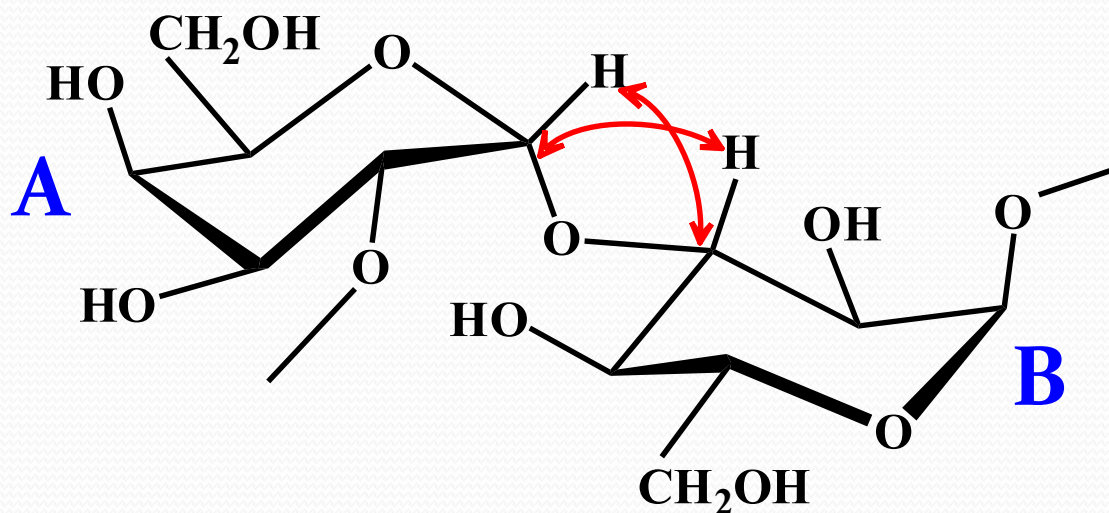
A1 – B3 long range correlations
(H1 A – C3 B and C1 A – H3 B)

A1- B3 linkage / A and B are 1-3 linked

HMBC of Sucrose
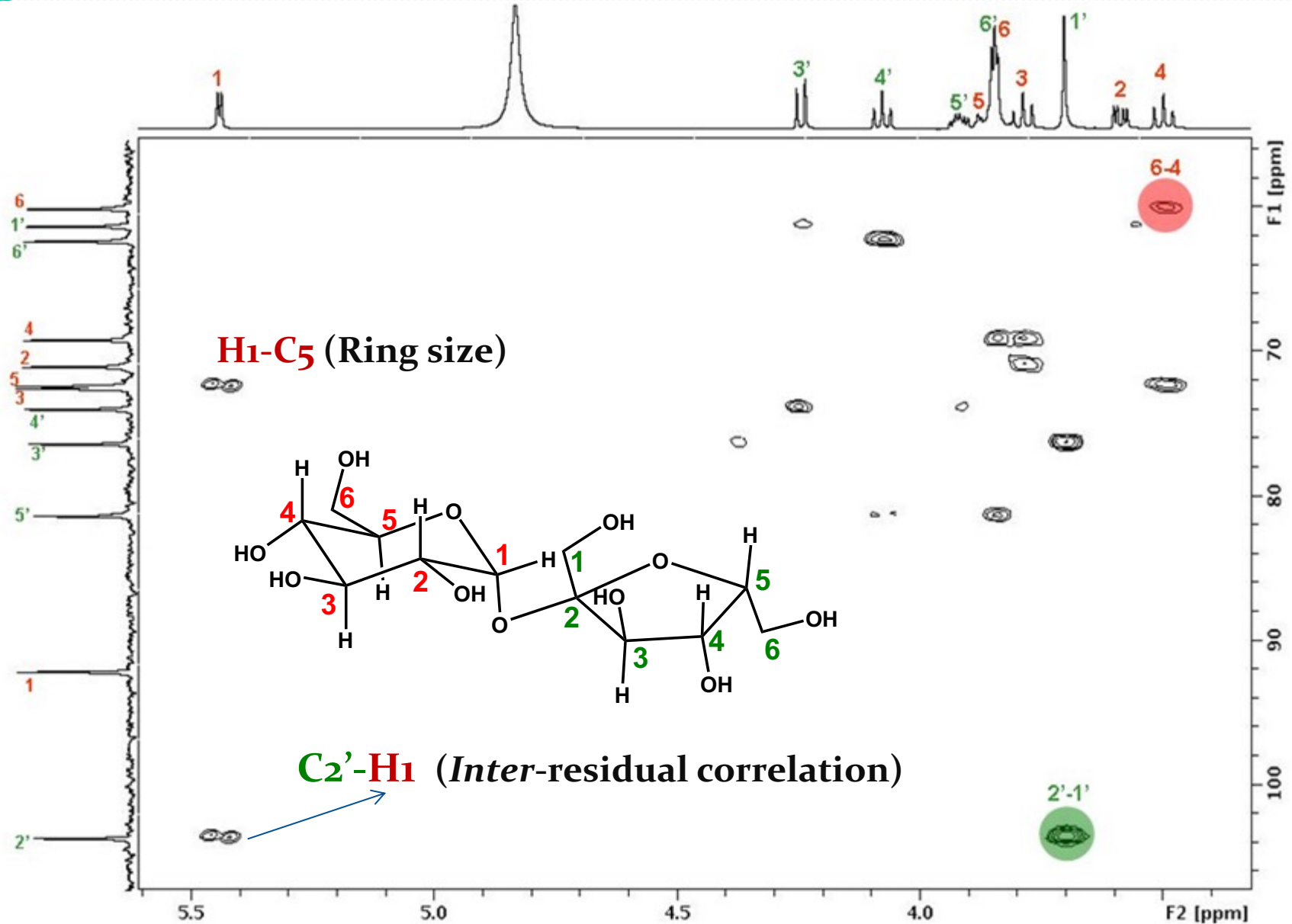
Nuclear Overhauser Effect (NOE) Spectroscopy

The 2D spectrum will have chemical shifts in f1 and f2.

The cross peaks are for nuclei that are coupled dipolarly.
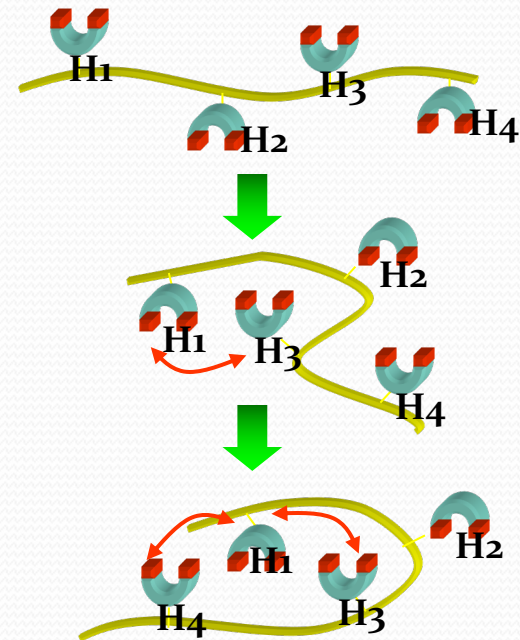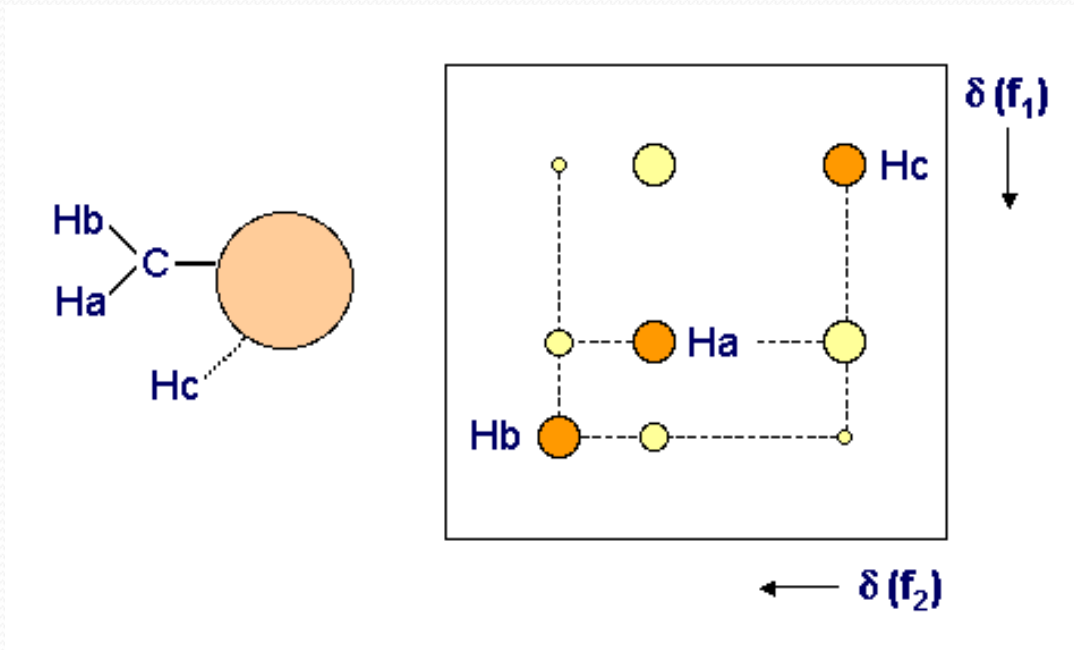
NOE contact: <u>C is close in space to spin A</u>

The NOE effect is the method for the elucidation of 3D structural features and stereochemistry

cosy   tocsy   noesy
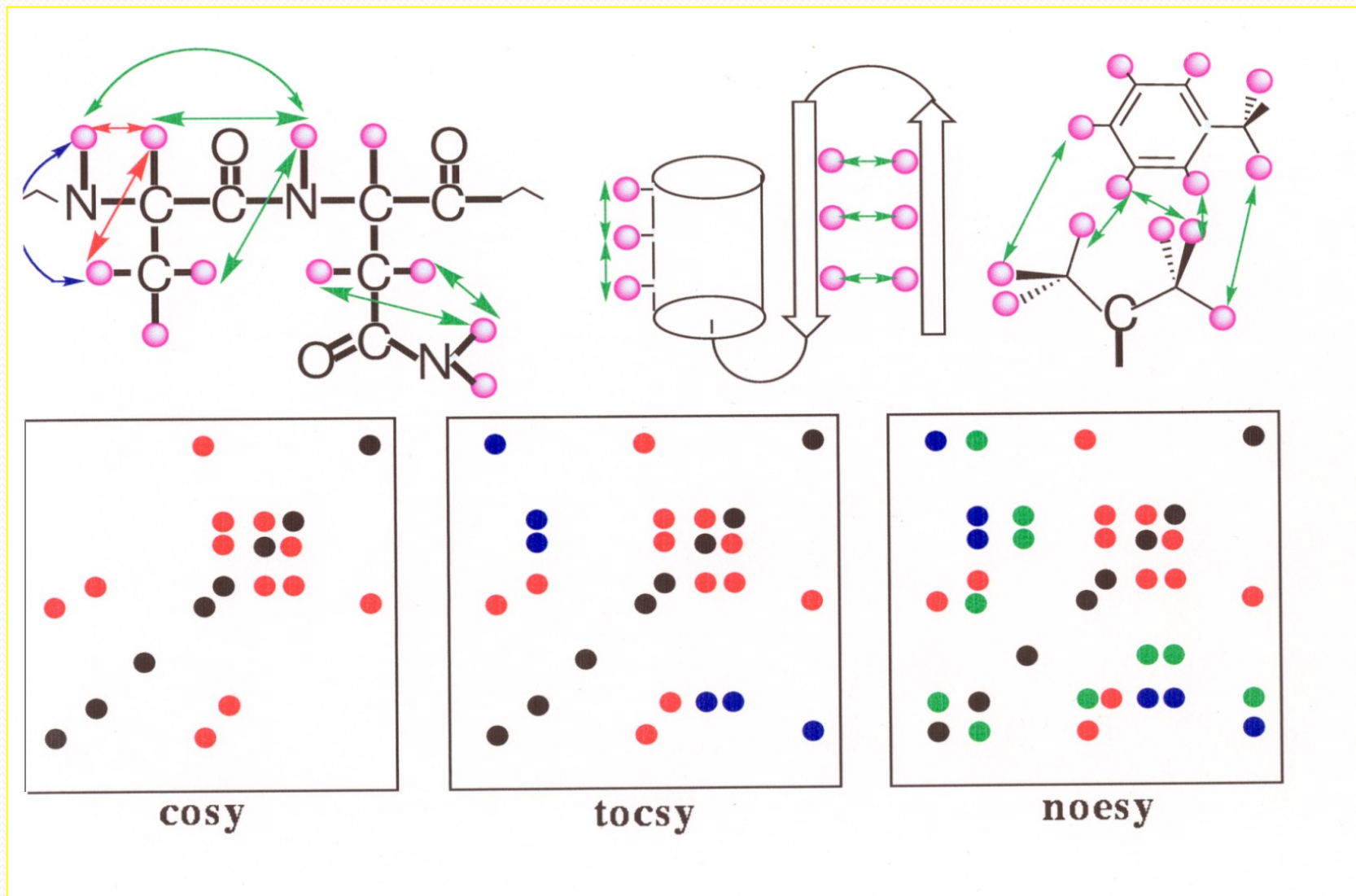
NOE and Distances
Isolated spin pair aproximation (ISPA)

NOE H1'-H3'   **NOE H1'-H4**

NOE build up curves

$$\frac{\sigma_{H1'-H4}}{\sigma_{H1'-H3'}} = \frac{r^{-6}_{H1'-H4}}{r^{-6}_{H1'-H3'}}$$

$\eta_{ab} \propto r_{ab}^{-6}$

$$r_{ac} = r_{ab} * (\eta_{ab} / \eta_{ac})^{-1/6}$$

$\eta_{ac} \propto r_{ac}^{-6}$

NOE

$\propto r_{IS}^{-6}$

τ_{mix}
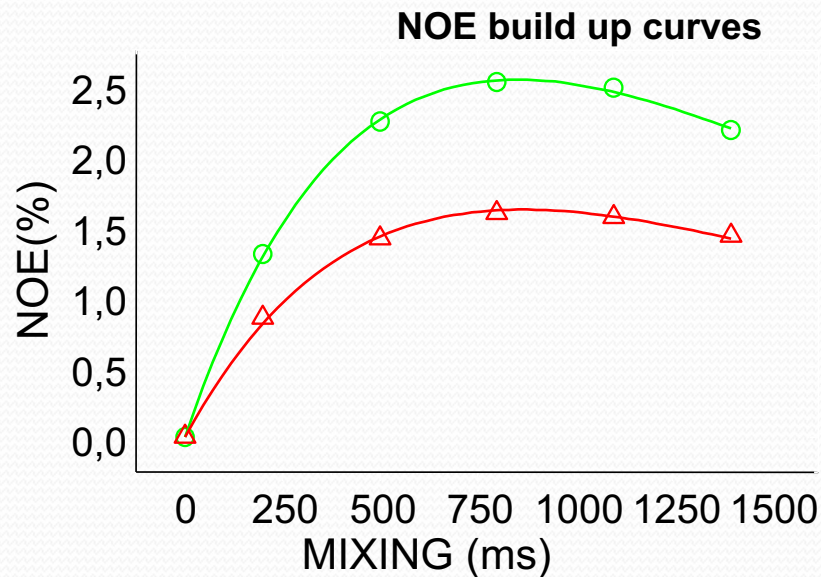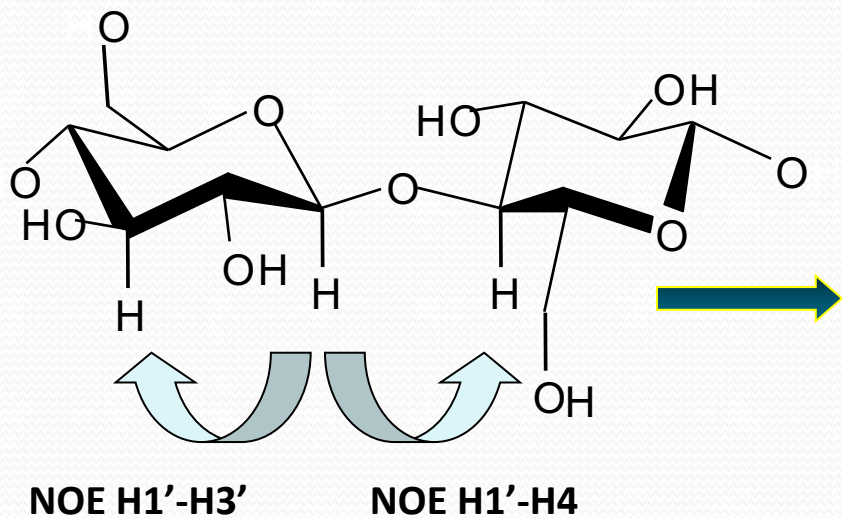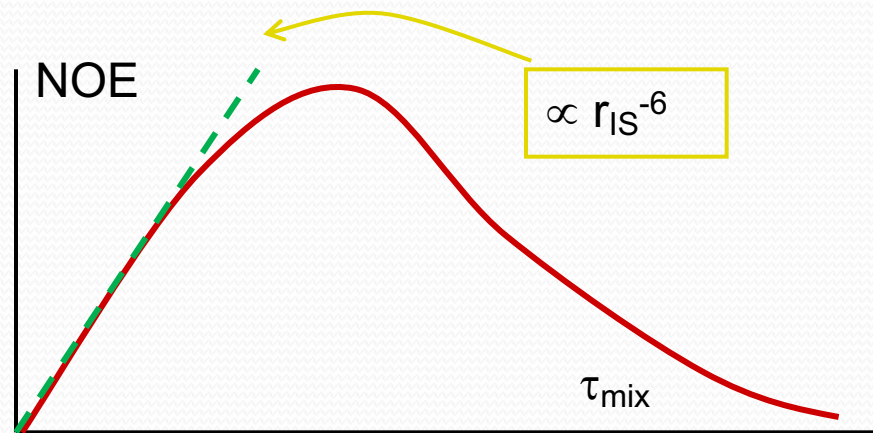
Intra-residue NOE contacts in monosaccharides: relative configuration of sugar residues

gluco, galacto configuration
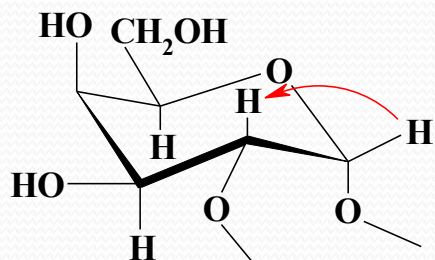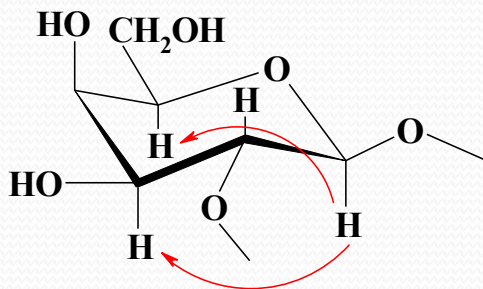
α-linkage: H1/H2

manno configuration
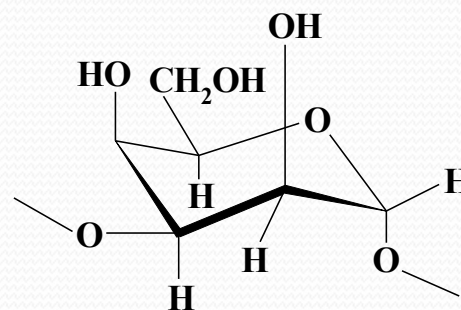
α-linkage: no contact

β-linkage: H1/H3, H1/H5
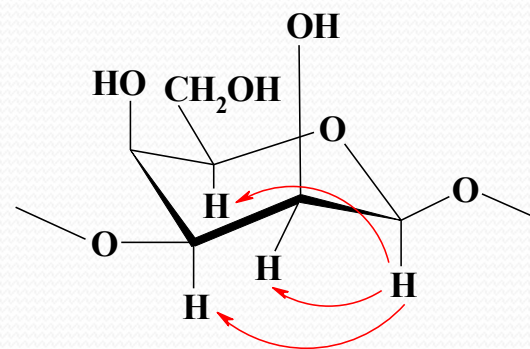
β-linkage: H1/H2, H1/H3, H1/H5

Monosaccharide Sequence

❖***Inter*-residue NOE**

❖Glycosylation shift (HSQC spectrum)

❖Inter-residual long range correlation (HMBC spectrum)

Inter-residue NOE contacts in saccharides (Linkage analysis)
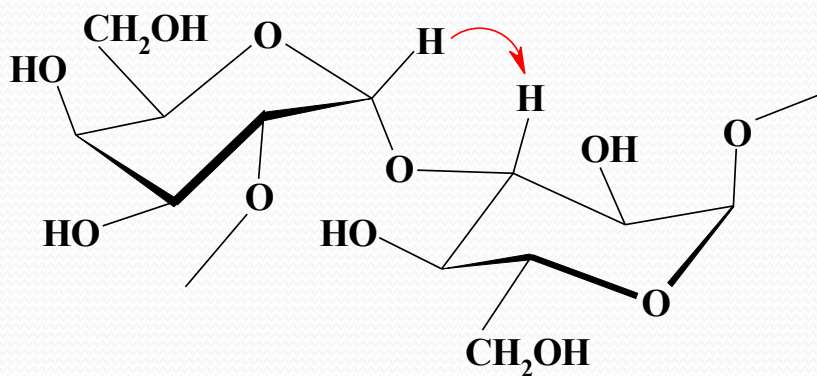
β-(1-3) linkage

α-(1-3)linkage
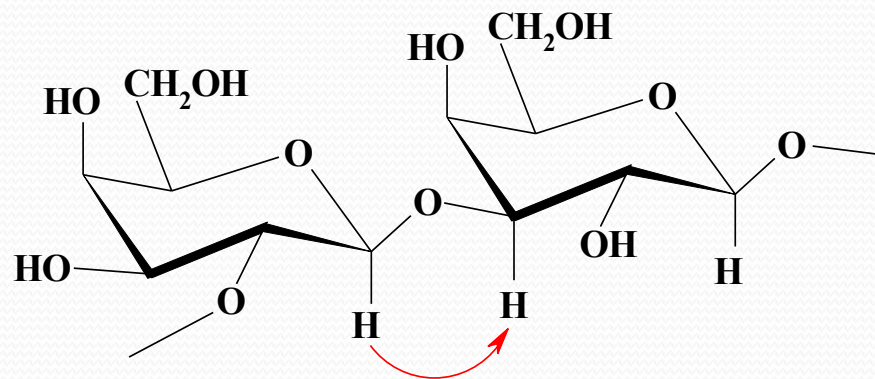
NOE in disaccharides may occur not only at the linkage protons but also at the neighbouring protons
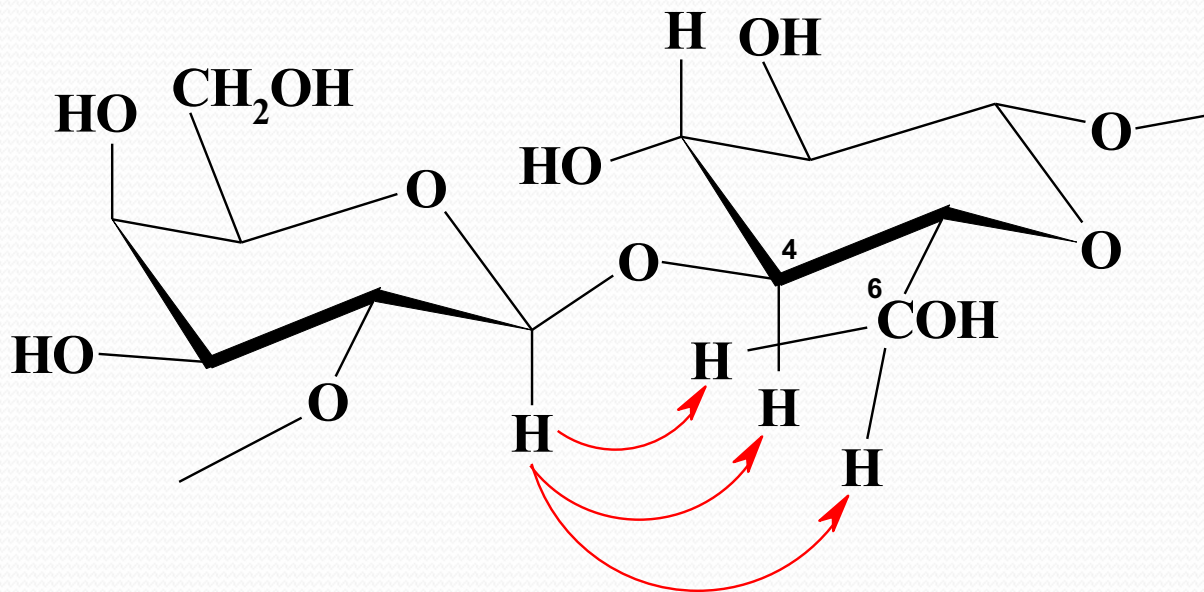
....Saccharide conformation...

Sucrose NOESY and TOCSY
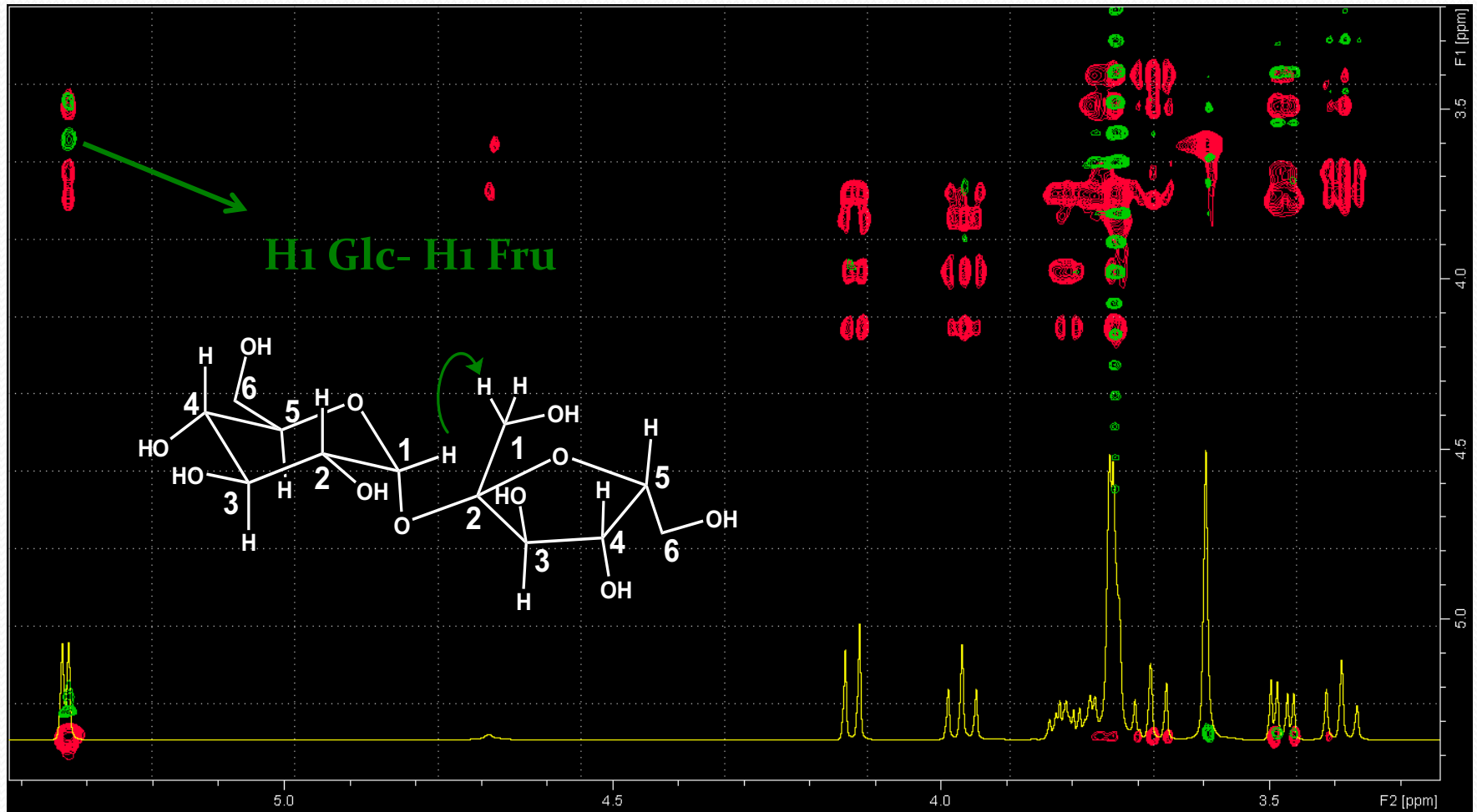
Application of various NMR techniques to carbohydrates

- HOMONUCLEAR (^{1}H-^{1}H)
- HETERONUCLEAR (^{1}H-^{13}C)
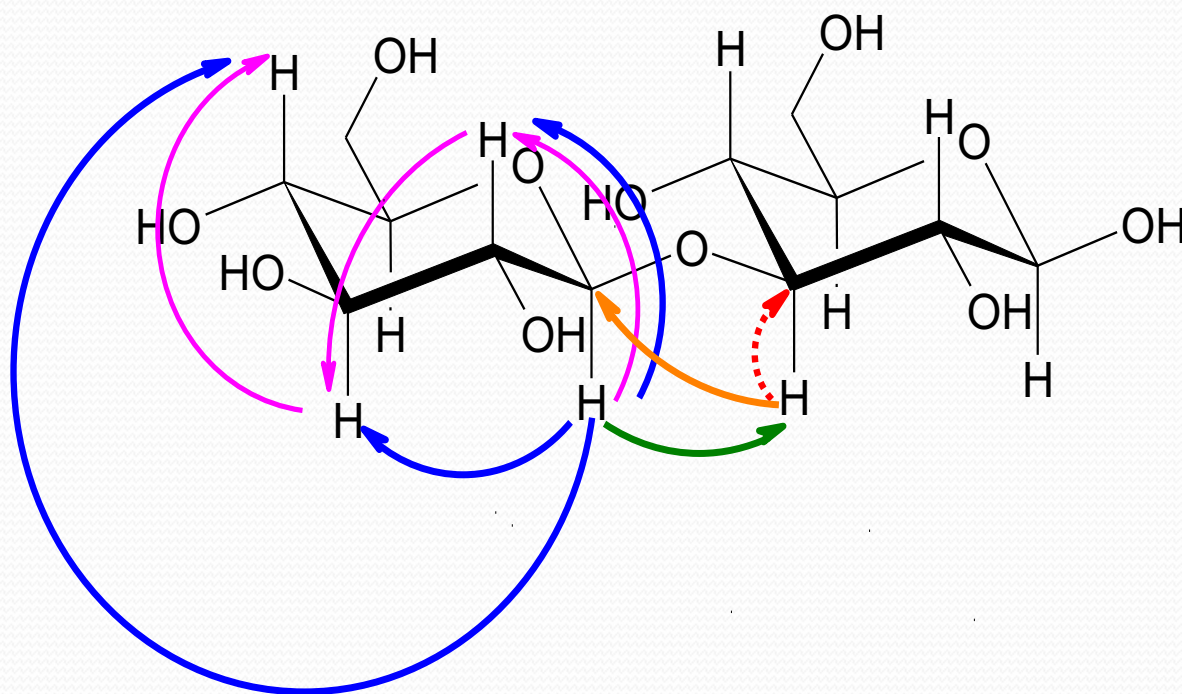
HOMONUCLEAR ^{1}H-^{1}H
COSY
TOCSY
NOESY/ROESY
HETERONUCLEAR ^{1}H-^{13}C
HMQC/HSQC
HMBC

A BACTERIUM'S ENEMY ISN'T YOUR FRIEND

a Strong immune response against WTA

b Weak immune response against WTA

Alanyl-shifted polyols protons

β-GlcNAc H-1 protons

R = H or Ala
R'= H or GlcNAc
P = Phosphate

ΔtarP WTA

ΔtarS WTA

Wild type WTA

ΔtarPΔtarS WTA

N-Acetyl
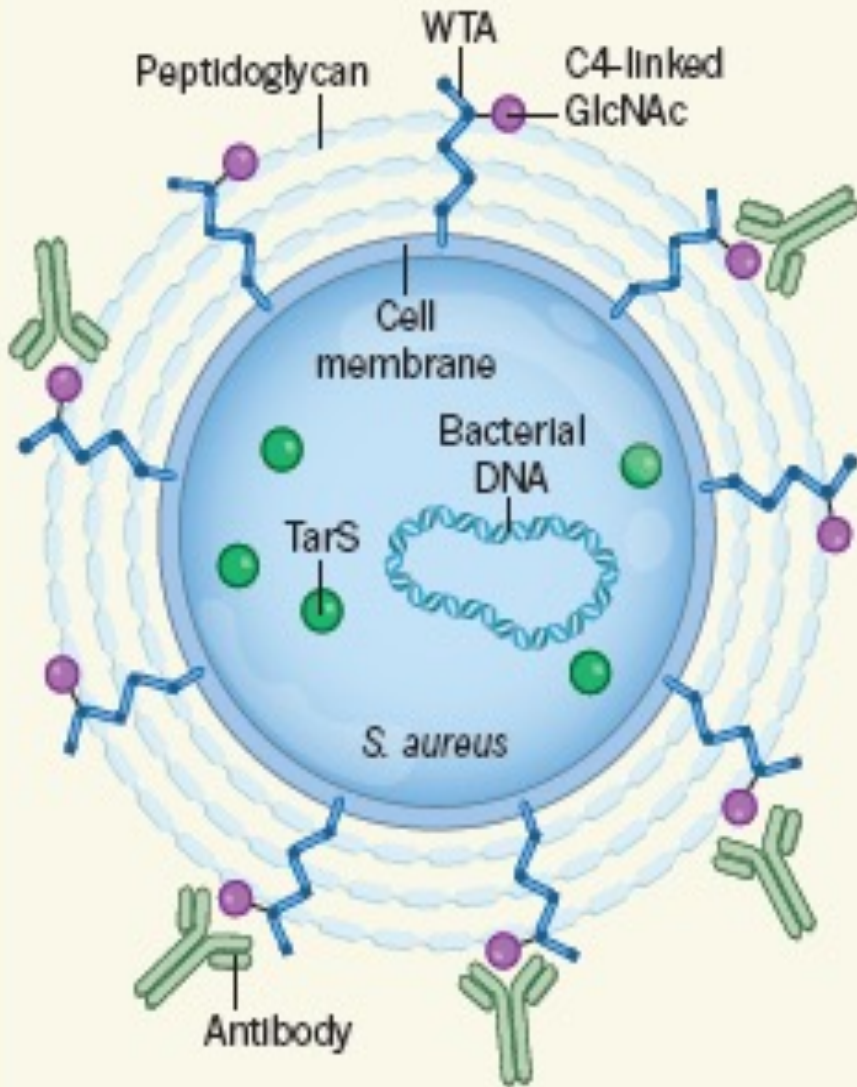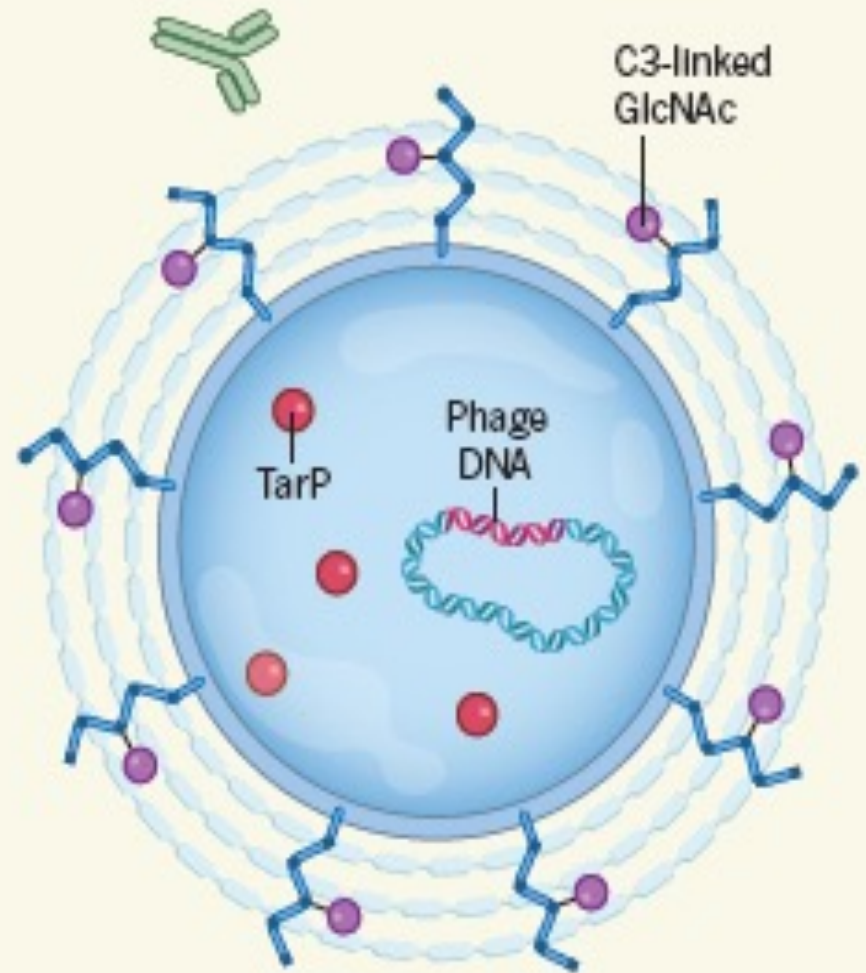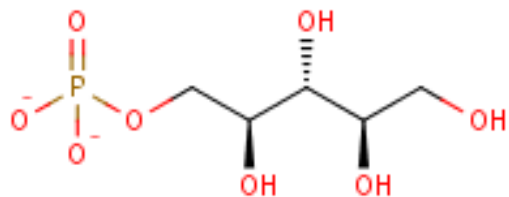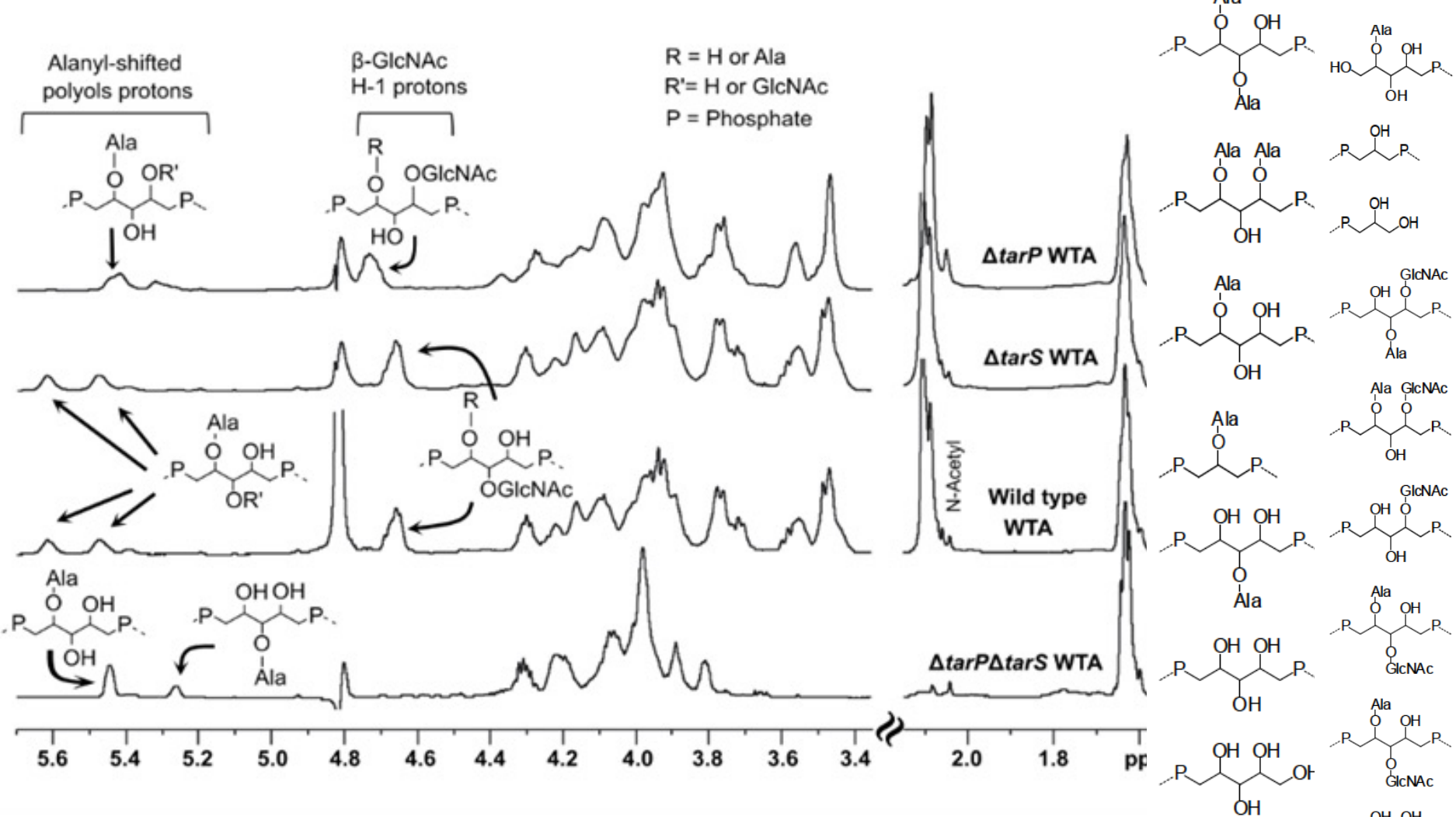
Methicillin-resistant Staphylococcus aureus alters cell wall antigen glycosylation to subvert protective host defense

Nature 2019

MOLECULAR INTERACTION by NMR

The ligand-based approach

- STD NMR
- TrNOESY
- WaterLOGSY
- Relaxation experiments
- Diffusion experiments

- Other nuclei: ^{19}F
- Paramagnetic tagging
- Other methods

The receptor based approach

- Isotope labelling
- Chemical shift perturbation mapping
- Paramagnetic tagging
- Other variations

Representation of protein-ligand interactions

Chemistry Open, 2016
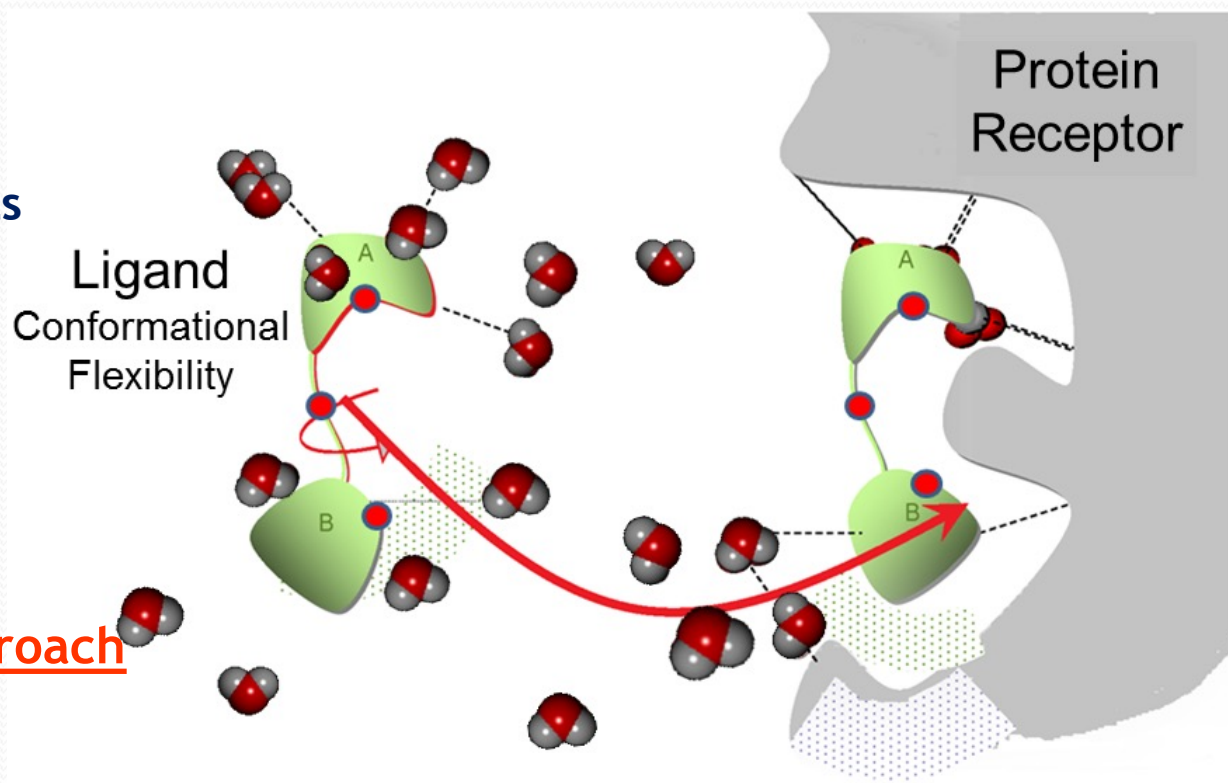
Ligand observation

$$L + R \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} LR$$
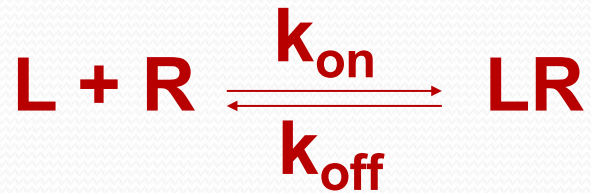
Two states equilibrium

L_{free} L_{bound} Molar fractions

FAST EXCHANGE

$$K_d = \frac{k_{off}}{k_{on} \; \textit{diffusion controlled}}$$

$$k_{off} \; = \; > 10^2 \; (s^{-1})$$

$$k_{on} \; = \; > 10^7 \; (s^{-1}M^{-1})$$

$$R_{Lobs} = L_f * R_{Lf} + L_b * R_{Lb}$$

$$\Delta R = L_b * (R_{Lb} - R_{Lf})$$

Experimental procedure: $L_0 >> R_0$;

$L_0/R_0 > 10 - 100....$ $L_f >>> L_b$

Necessary condition: $|(R_{Lb} - R_{Lf})| >> 0$

R_{Lb} Strong dependency on molecular size

NMR observable parameter R : **NOE; Diffusion; Line Shape**

Recapitulation

- K_{off} fast in the relaxation time scale, dissociation must occur before relaxation.

- K_{on} related to the efficacy of the interaction.

- Consider the molar fraction of ligand free and bound to the protein.

- Fast exchange in the chemical shift time scale

- **$L_0 >> R_0$ means excess of free ligand but since we are in conditions in which the exchange is high (rate) the system is dominated by the bound state**

- K_{off} Dissociation before relaxation takes place; dissociation rate high since the molecule relax in the binding site

TRANSFERRED   NOE

Information on the ligand bioactive conformation
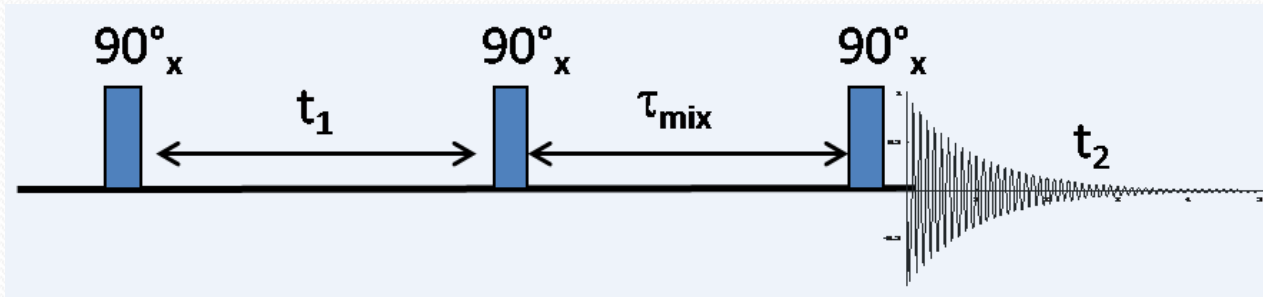
❖ **During the mixing time *inter* and *intra*-molecular NOE effects build up**
❖ ***Inter*-molecular tr-NOE effects are visible, intermolecular trNOEs are usually much larger than intramolecular effects**

Important notes

❖The mixing time must be short enough so that the contribution of the free ligand is negligible and long enough to allow visualization of the signal in the spectrum.

❖The molar ratio of ligand to receptor. It should be emphasized that the trNOESY experiment works well for ligands that have K_D in the range 10^{-3} – 10^{-6} M / mM- mM range

❖Small amount of purified receptor

❖Routinely used to probe ligand-receptor interaction

TRANSFERRED NOE and MOLECULAR MOTION
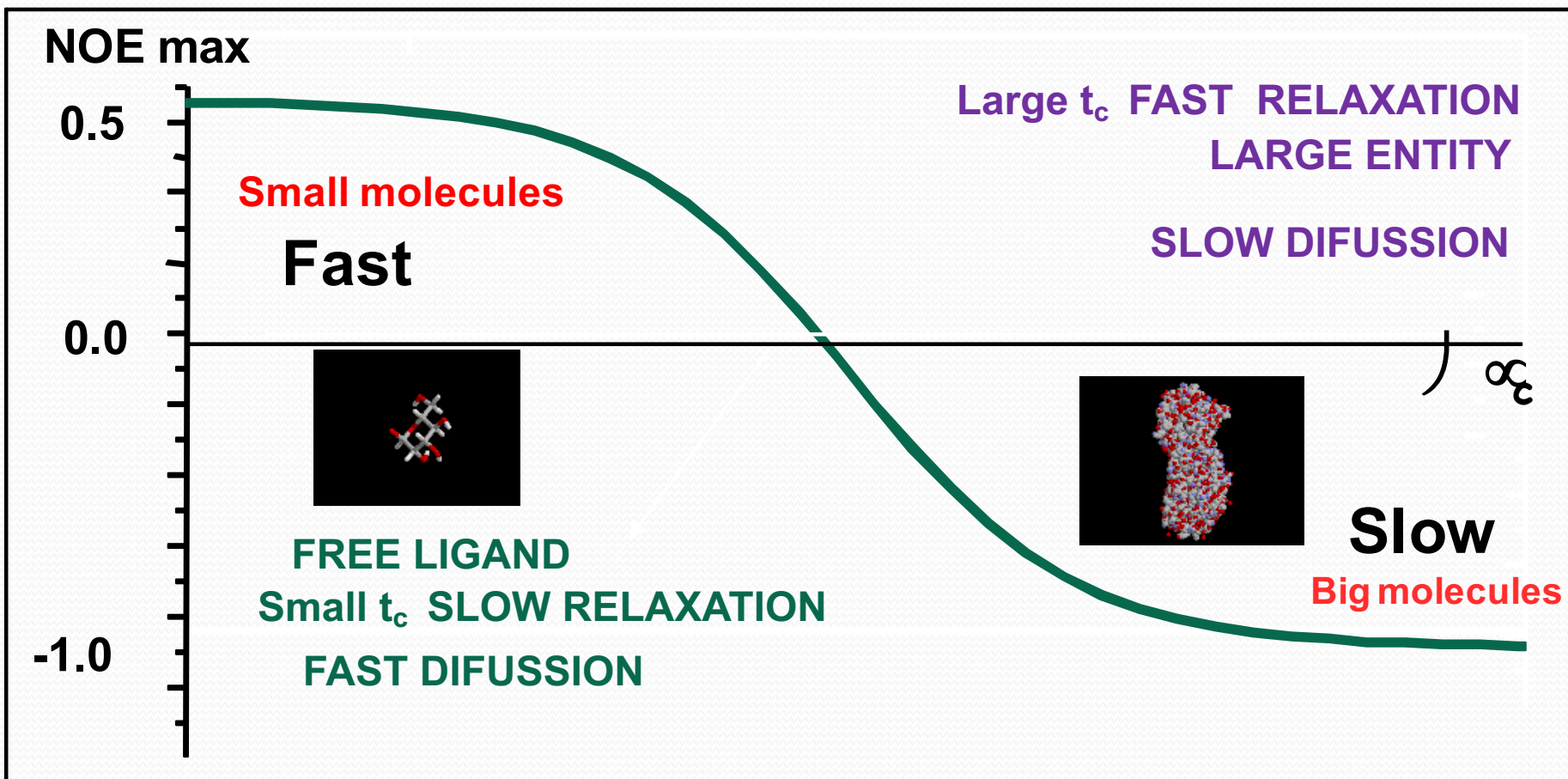
The bioactive conformation: Transfer NOESY

Kon

Koff

FREE

BOUND

THE BIOACTIVE CONFORMATION

NOE

small molecule

mixing time

large molecule

NOESY

TR-NOESY

TR-ROESY
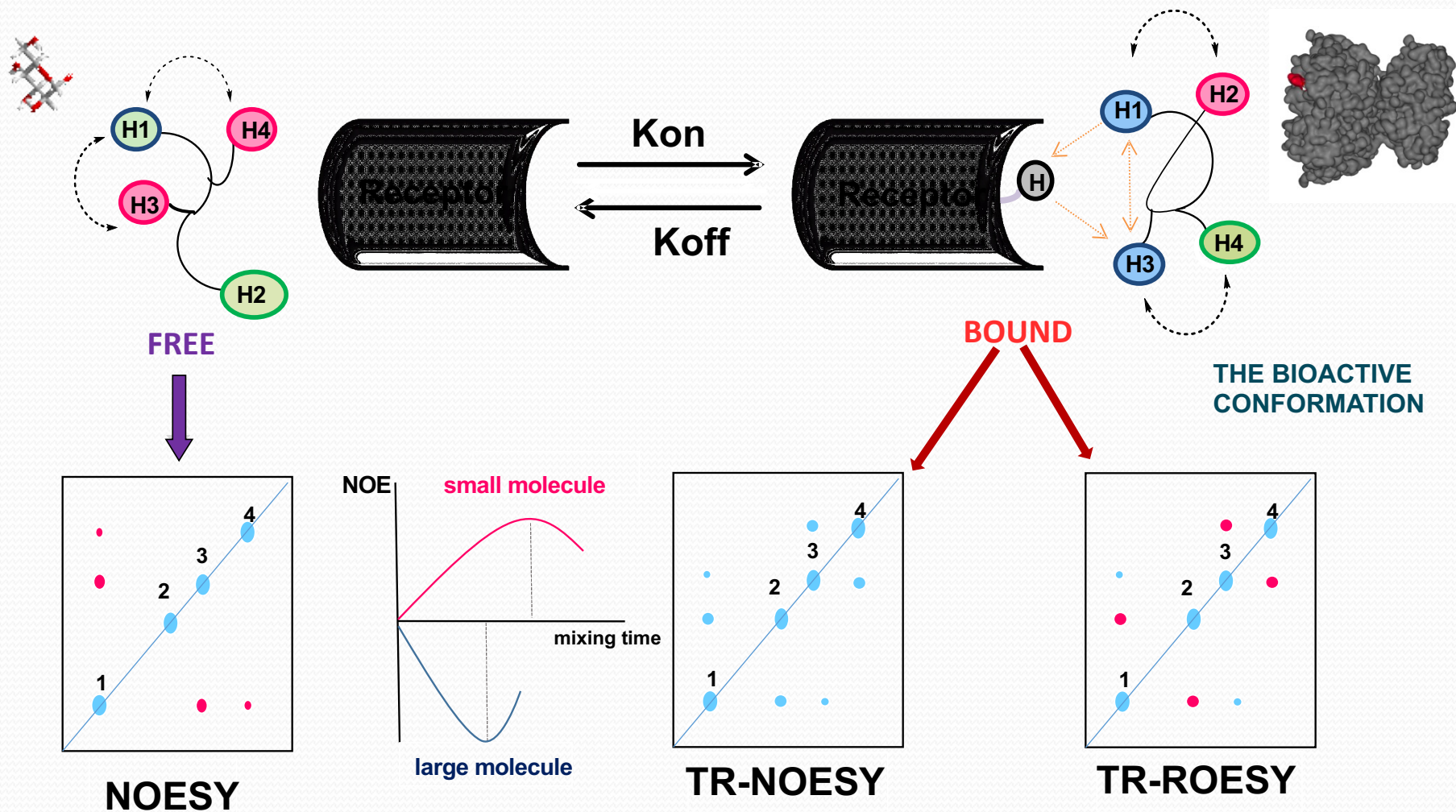
Is There Any Binding?

Which Is The Ligand Bioactive Conformation?

Saturation Transfer Difference NMR Spectroscopy – STD NMR

IS THERE ANY BINDING FOR ANY GIVEN COMPOUND?
WHICH IS THE BINDING EPITOPE?

IRRADIATION at the aromatic or aliphatic NMR regions

Single Compound or Library
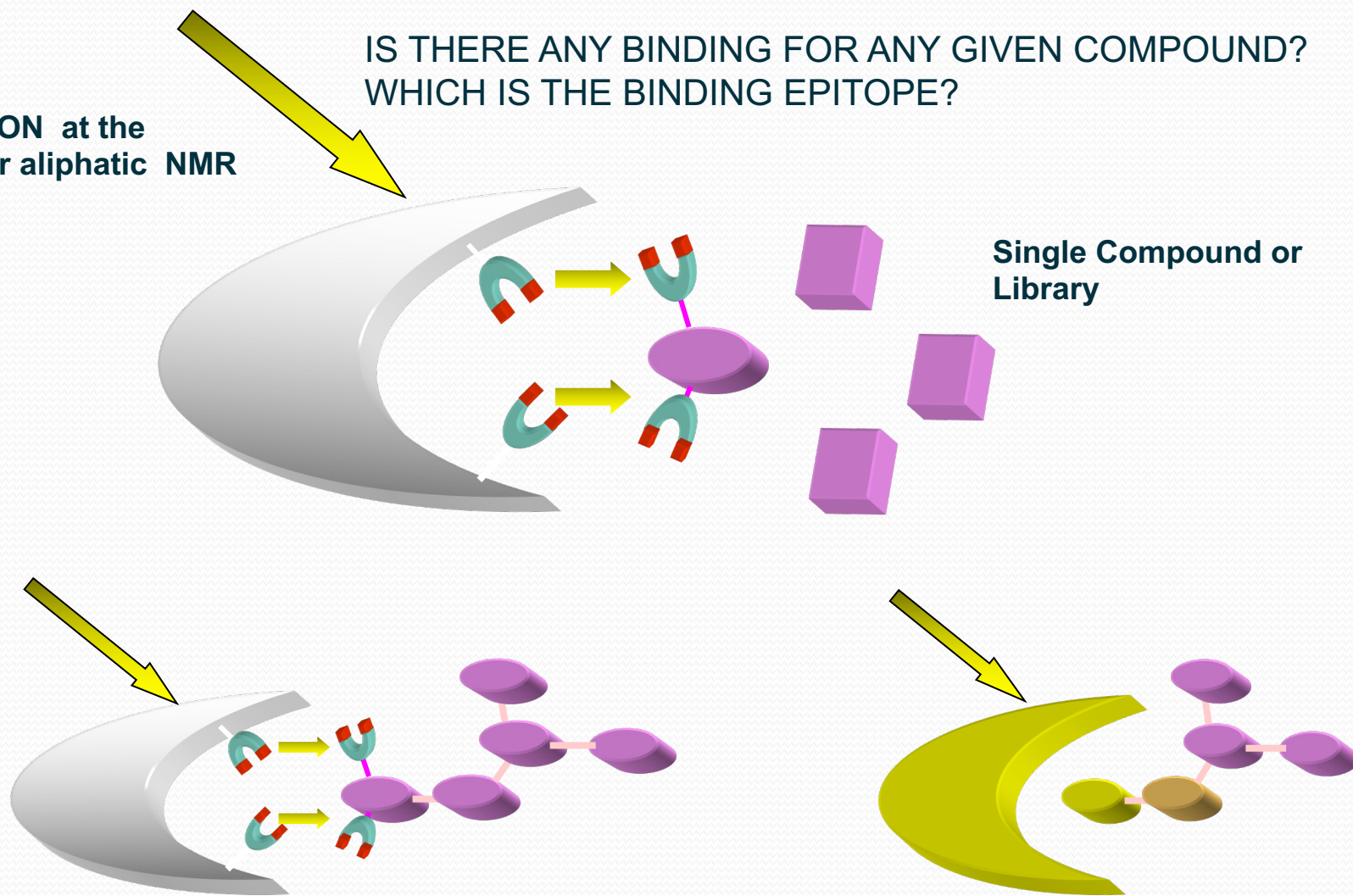
At long irradiation times, the saturation is transferred to the bound ligand, first to the protons belonging to the ligand epitope, then to the rest of the ligand

(Meyer and Mayer, Peters, 1999, 2000)

Schematic representation of STD NMR method.

on resonance

r.f.

or

binder

Ha Hb Hc

non binder

Ha' Hb' Hc'

Hb Hc
Ha

Relaxation

Ha Hb Hc

Kon

Ha Hb Hc

Koff

Ha Hb Hc

off resonance

r.f.

binder

Ha Hb Hc

non binder

Ha' Hb' Hc'

Hb Hc
Ha

off resonance → I_0

Ha Hb Hc   Ha' Hb' Hc'

ppm

on resonance → I_{SAT}

ppm

STD NMR → $I_{STD} = I_0 - I_{SAT}$

ppm
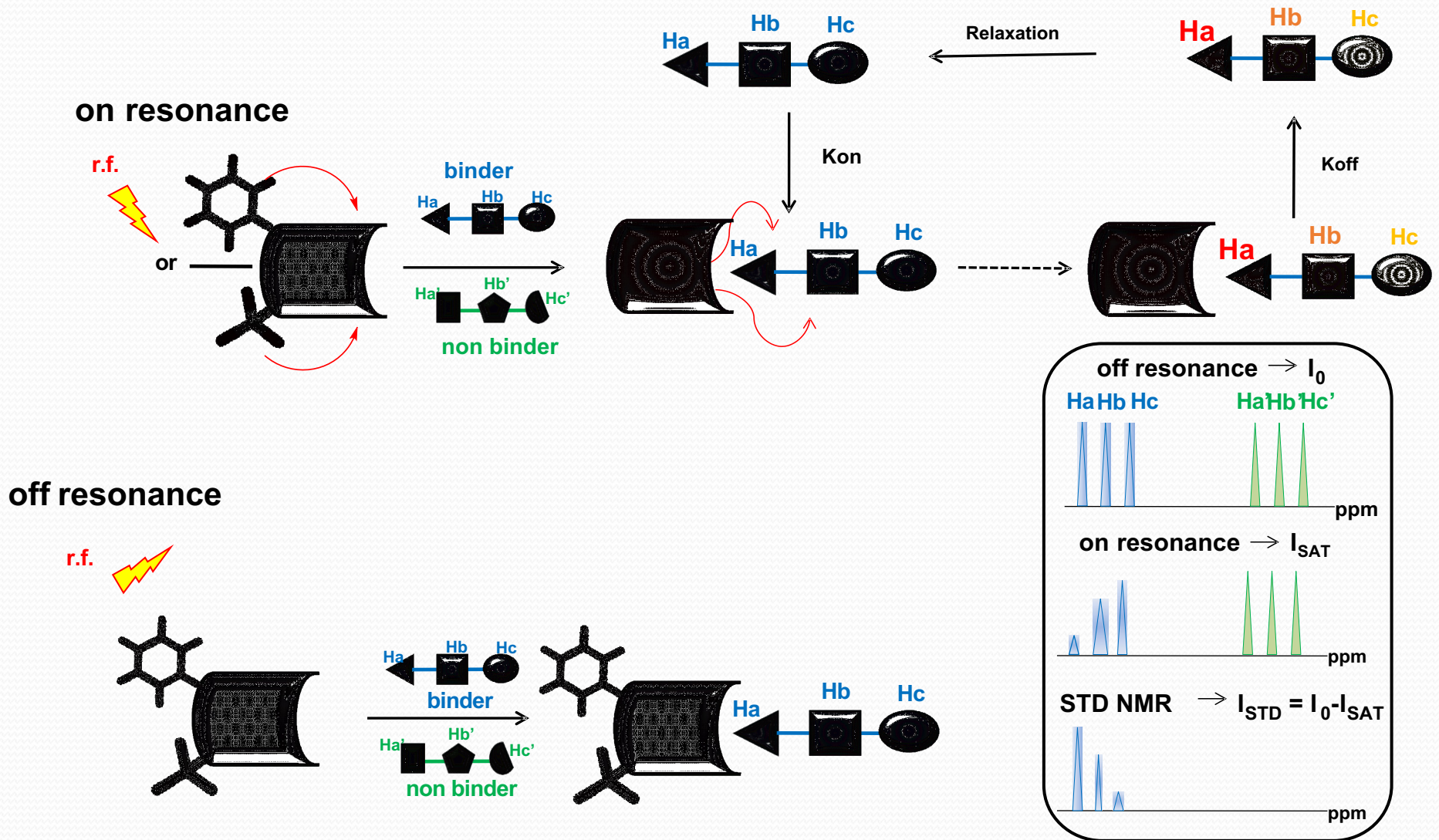
Key elements of protein-substrate binding

H close to receptor

H far from receptor

H furthest away from receptor

STD (Meyer and Mayer, Peters, 1999, 2000)

RECAPITULATION...

- **The saturation is transferred first to the protons belonging to the ligand epitope, then to the rest of the ligand**

- STD involves selective perturbation of protein-specific methyl proton resonances (0.5 p.p.m.–1 p.p.m.). This perturbation rapidly diffuses throughout the protein.

- This is usually done via a selective 180-degree pulse and results in a transfer of magnetization from the protein to any transiently bound ligands via the nuclear Overhauser effect (NOE).

- If the fragment binds to the target protein, the buildup of NOE that is transferred to the ligand results in enhanced signal corresponding to the resonances of that ligand in the STD spectrum.

- A number of factors affect the signal strength in an STD experiment, including protein size, duration of on-resonance irradiation, the frequency of irradiation, the dissociation constant of the ligand, the ligand/protein ratio and the field of the spectrometer.

- As such, it is useful to vary parameters such as ligand and protein concentration, the frequency of irradiation and the duration of irradiation

NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors

β-Gal-Ome and RCA120 (Ricinus communis agglutinin), 50 μM and 1.2 mM, 1:40

A) Reference 1D NMR spectrum of the 120 kDa protein RCA120

B) Corresponding STD NMR spectrums, ON at -2 ppm

C) 1D NMR spectrum recorded with a T11 filter (ms spin-lock pulse)

D) Reference 1D NMR spectrum of RCA120 in presence of 1.2 mm b-GalOMe

E) Corresponding STD NMR spectrum, which shows the β-GalOMe resonance signals and therefore its binding to the receptor

F) STD NMR spectrum as in (E) but with the $T_{1\rho}$ filter eliminating all protein background signals.

STD effect: $(I_0 - I_{sat})/I_0$

M. Mayer, B. Meyer, J. Am. Chem. Soc. 2001;123(25):6108-17; Bernd Meyer, Thomas Peters, Angewandte Chemie International Edition, 42, 8, 864-890

NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors

β-Gal-Ome and RCA120 (Ricinus communis agglutinin), 50 μM and 1.2 mM, 1:40
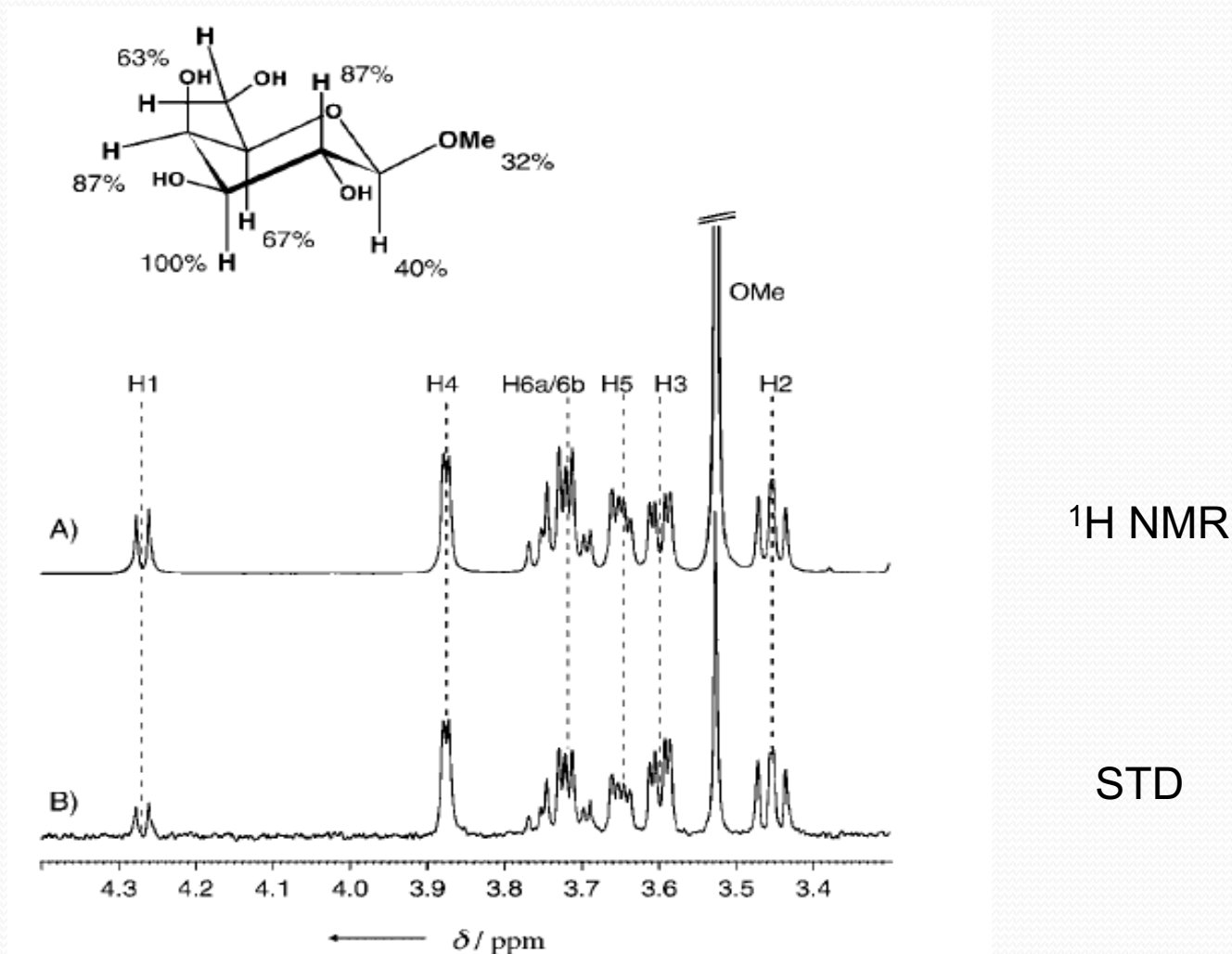
In order to determine the magnitude of the STD effects, the intensity of the signal in the STD NMR spectrum are compared with the signal intensities of a reference spectrum (off-resonance).
The STD signal with the highest intensity is set to 100% and the others are normalized to this signal.

M. Mayer, B. Meyer, J. Am. Chem. Soc. 2001;123(25):6108-17; Bernd Meyer, Thomas Peters, Angewandte Chemie International Edition, 42, 8, 864-890

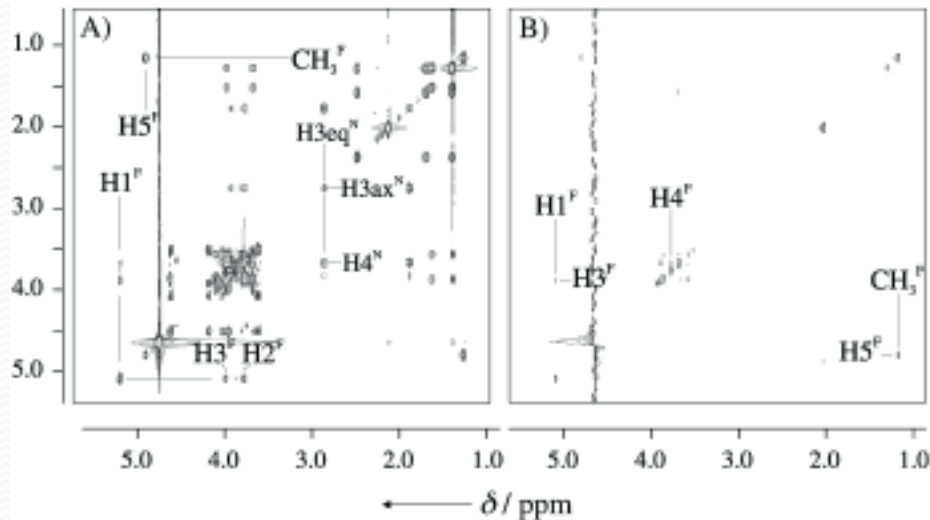# Epitope Mapping with STD NMR Spectroscopy

Epitope mapping for sialyl LewisX bound to the the lectin *Aleuria aurantia* agglutinin (AAA)
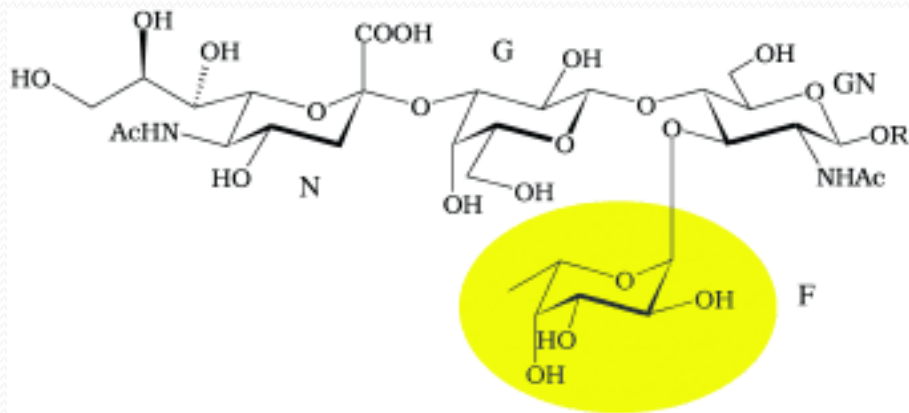
STD TOCSY spectrum of sialyl LewisX in the presence of AAA (molar ratio 100:1): only the spin system of the l-fucose residue (F, yellow circle) is visible
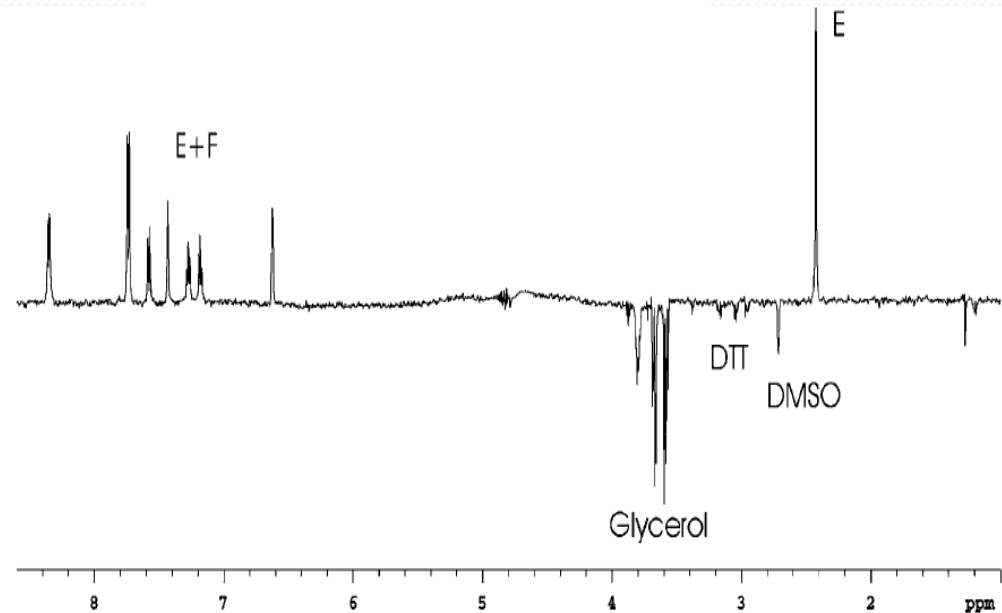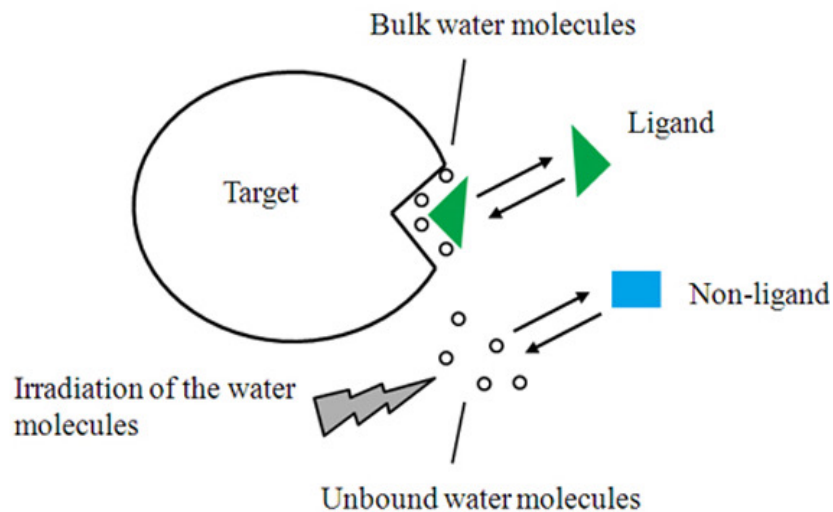
1D and 2D STD spectra shows unambiguously that only the fucose interact with the protein

M. Mayer, B. Meyer, Angew.Chem. 1999, 111, 1902 – 1906 ;Angew.Chem.Int.Ed. 1999, 38, 1784 – 1788.

Water-LOGSY NMR

The resonances of non-binding compounds appear with opposite sign and tend to be weaker than those of the interacting ligands.
Experimentally, the first step is the selective water excitation; during the mixing time (to be optimized based on the size of the complex) the water magnetization that has migrated to the protein is transferred to the ligand via direct of relay processes.

Dalvit C., Fogliatto G., Stewart A., Veronesi M, Stockman B., J Biomol NMR. 2001, 21(4):349-59.

Carr Purcell Meiboom Gill (CPMG) experiments

CPMG, is a relaxation-time-edited NMR experiment that exploits differences in transverse relaxation time (*T2*). *Proteins (and bound ligands)* have a small *T2 while free ligands have a large T2. Thus monitoring T2, binding* can be detected when the signal of the ligand decreases.
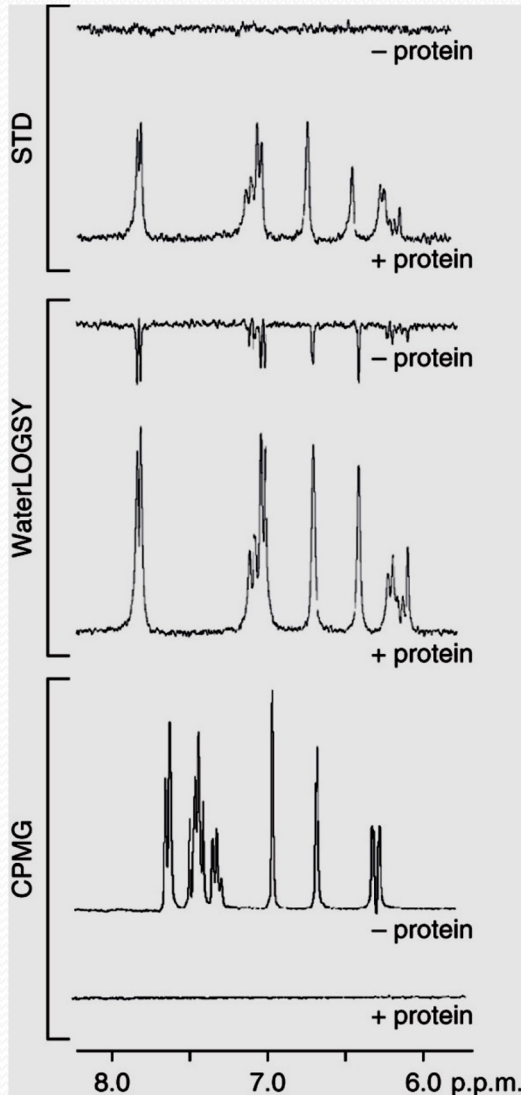
In the CPMG experiment, we should see the signal decrease (or disappear) in the presence of protein and a binding fragment
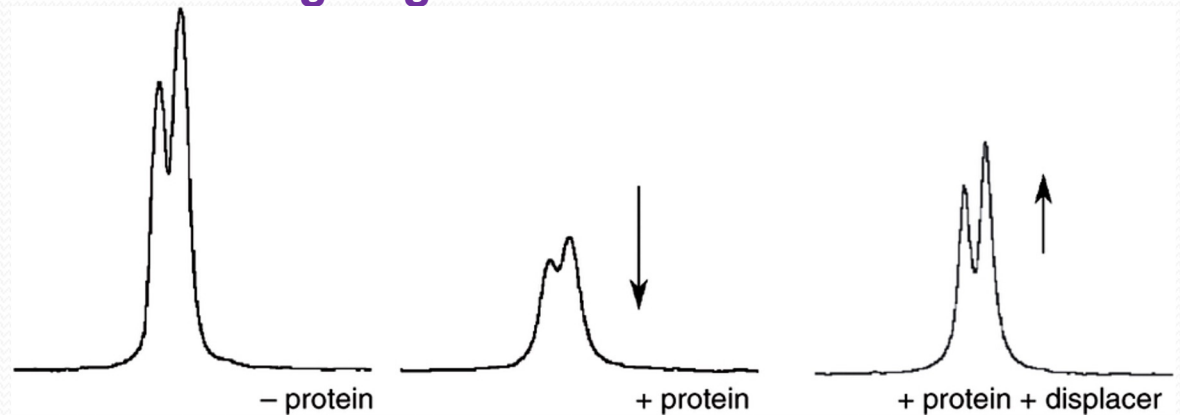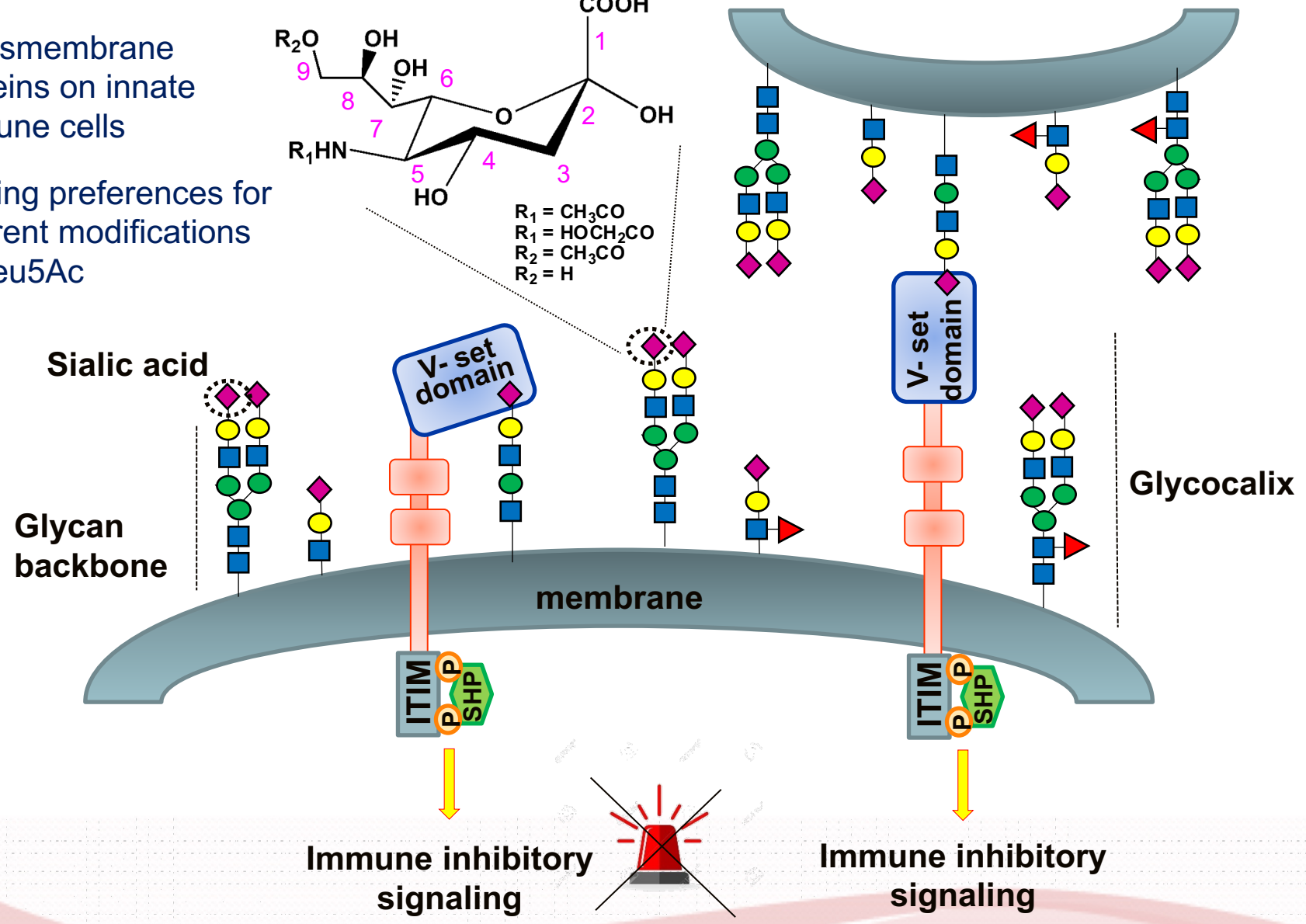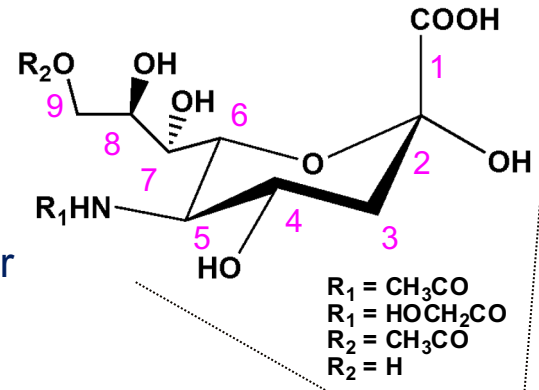

In the presence of protein, the signal decreases, which is indicative of fragment binding. In the presence of both protein and a molecule known to bind the target (displacer), the signal is restored by ~30%. The protein target is *M. tuberculosis* CoaBC (10 µM). Fragment and displacer are present at 0.5 and 0.25 mM, respectively
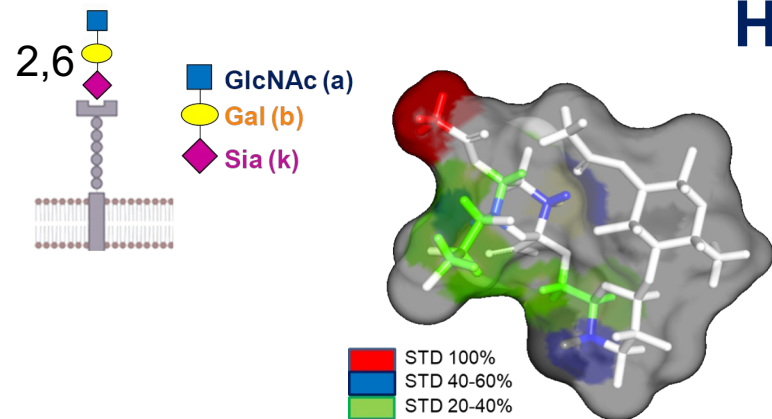
Carr, Y.H. & Purcell, M.E. Phys. Rev. 94, 630–638 (1954).; Hajduk, P.J., Olejniczak, E.T. & Fesik, S.W. J. Am. Chem. Soc. 119, 12257–12261 (1997).; E.H Mashalidis,   P. Śledź, S. Lang, C. Abell, Nature Protocols, 8, 2309–2324 (2013)

Siglecs: sialic-acid binding immunoglobulin-like lectins

SIGLECS – SIALIC ACID AXIS

- Transmembrane proteins on innate immune cells

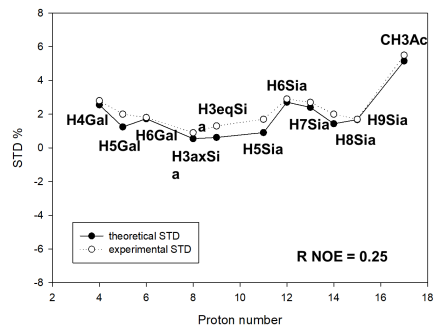- Binding preferences for different modifications of Neu5Ac

$R_1 = CH_3CO$
$R_1 = HOCH_2CO$
$R_2 = CH_3CO$
$R_2 = H$

Sialic acid

V- set domain

Glycan backbone

membrane

Glycocalix

ITIM P P SHP

ITIM P P SHP

Immune inhibitory signaling

Immune inhibitory signaling

H-CD22- sialoglycans studies
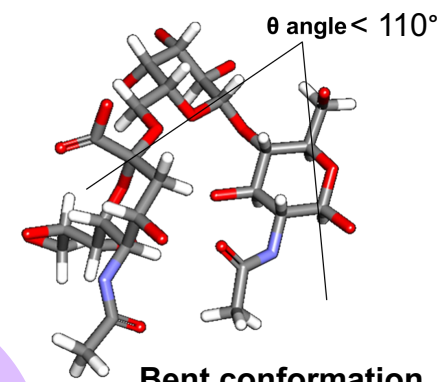
ChemBioChem, 2019
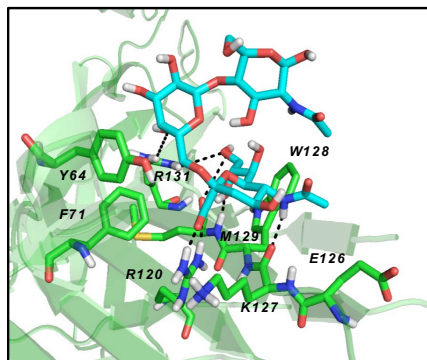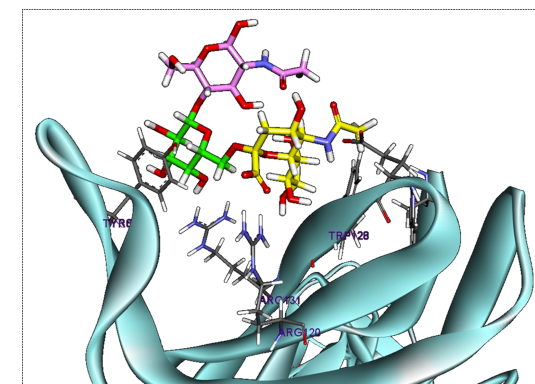
h-CD22- trisaccharide interaction: toward bioactive conformation

ω = O5-C5-C6-O6

ω

gt ω **60°**

gg ω **-60°**
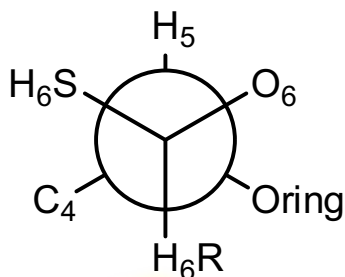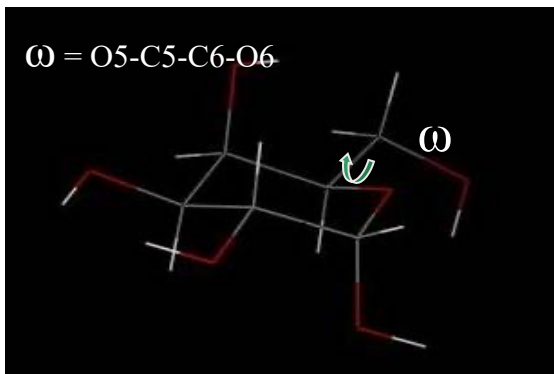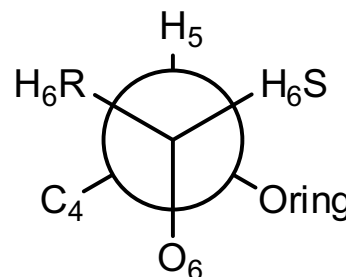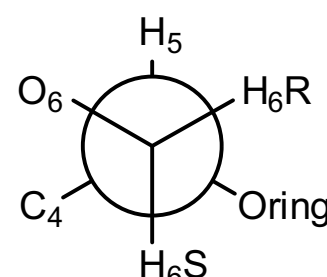
tg ω **180°**

HSQC-TOCSY (blue-green)
HSQC-HECADE (red)

PBS/D$_2$O

-4.1 Hz

Sia

Gal

GlcNAc
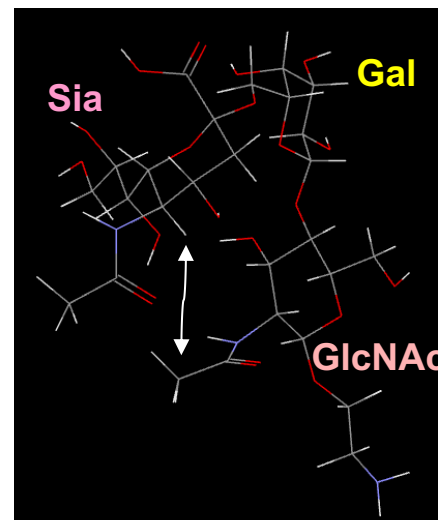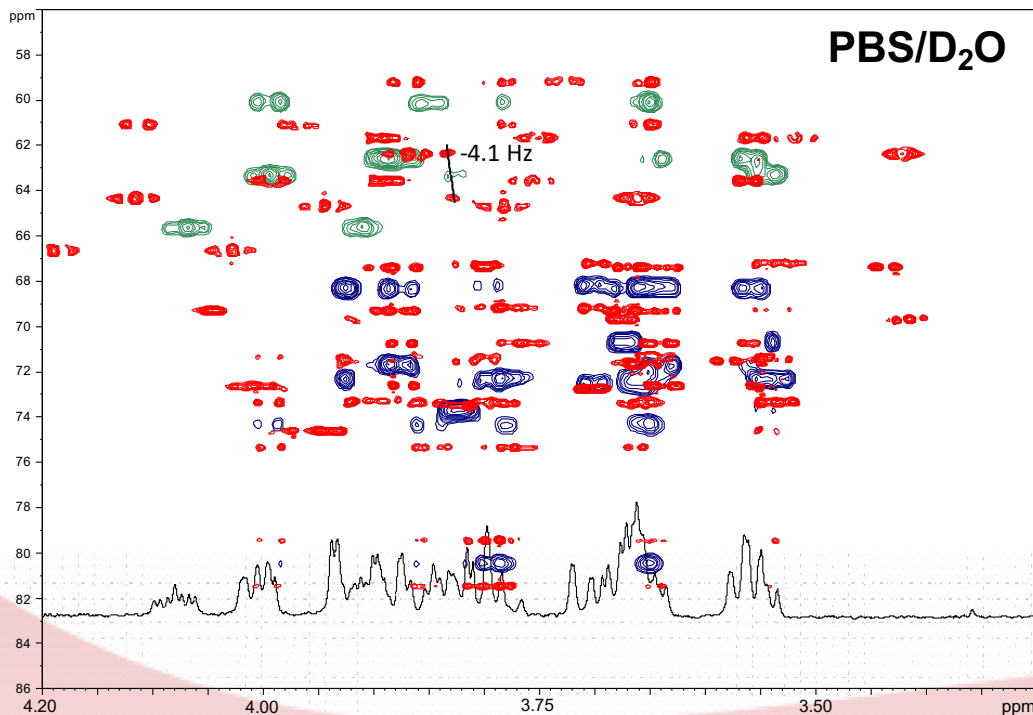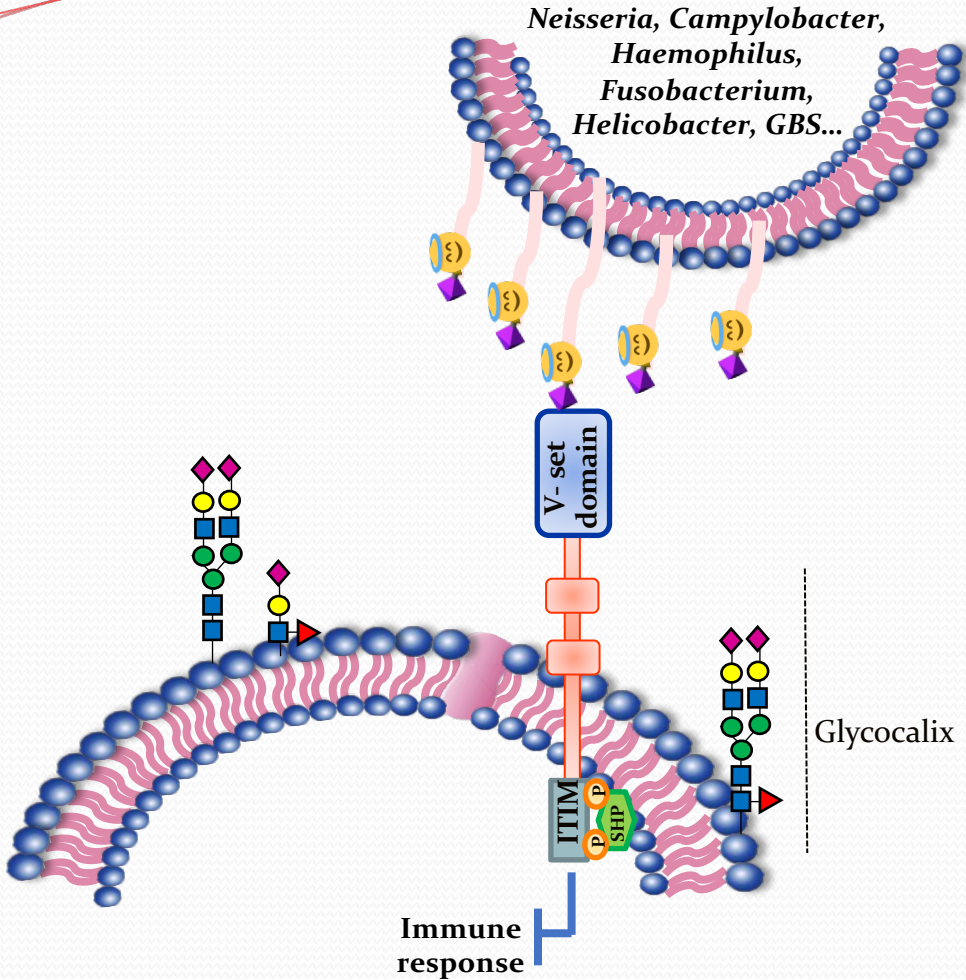
$$^2J_{(C6\,H5)} = -1.29 + 1.53\cos(\omega) - 3.68\sin(\omega)$$

*Kozmiński W, et al. *J Magn Reson.* 2000, 142, 294-299.

Siglecs interaction with pathogens

Neisseria, Campylobacter, Haemophilus, Fusobacterium, Helicobacter, GBS...

V- set domain

ITIM

SHP

P P P

Glycocalix

Immune response

- ➢ Cell signalling
- ➢ Cell adhesion
- ➢ Tumour sialic acids engagment
- ➢ Targets for cancer immunotherapy
- ➢ Uptake of sialylated pathogens
- ➢ Mechanisms to escape immune recognition

Dynamic changes of structures and composition of microbial cell envelope glycans, their **camouflage with host self-antigens (molecular mimicry), their cross-talk with host immune Siglecs**, is central in GLYTUNES. The understanding of the role of sialic acid–Siglec interactions in the context of pathogen-associated immune suppression will pave the way to the **development of novel therapeutics glycomimetic to be used in clinical and diagnostic applications**

Questions?