Structure Determination of Carbohydrates

Internet tools for the interpretation of NMR spectra

Roland Stenutz

Department of Organic Chemistry,
Stockholm University
SE-106 91 Stockholm, Sweden

roland@organ.su.se

Polysaccharide structure:

Components	”type”	*Hex or HexNAc*
	relative configuration	*Glc or Man*
	absolute configuration	*D- or L-*
	ringsize	*-p or -f*
Linkages	position	*→4) or →6)*
	stereochemistry	*α- or β-*
Sequence		*→4)Glc(→4)Gal(→* or *→4)Gal(→4)Glc(→*

NMR can be used to perform all of the steps in a structure determination

- *except the determination of the absolute configuration*

Many oligo- and polysaccharides contain at least one sugar that occurs naturally only in one configuration.

e.g. D-Glc, D-GlcA, D-GlcNAc, D-GalNAc

The relative configuration of the other residues can be determined by NMR.

The interpretation of NMR spectra becomes difficult if the polysaccharide -

- is irregular (or a mixture)

- contains many similar residues

- is very large

Every polysaccharide has a unique 1D-NMR-spectrum -
i.e. all of the information about the structure is contained in a 1D-spectrum.

Most of the NMR-experiments are performed to assign the resonances and do not provide additional information about the structure.

NMR-experiments

Assignment of ^{1}H-resonances

^{1}H,^{1}H-COSY, ^{1}H,^{1}H-TOCSY

Assignment of 13C-resonances

^{1}H,^{13}C-HETCOR or HSQC

Sequence determination

NOESY (Short ^{1}H,^{1}H distances)

HMBC ($^{3}J_{COCH}$)

Methods for the interpretation of 1D-spectra can save much time and effort!

Current approaches:

comparison with a database (SugaBase)

simple and accurate but limited to known structures or sub-structures.

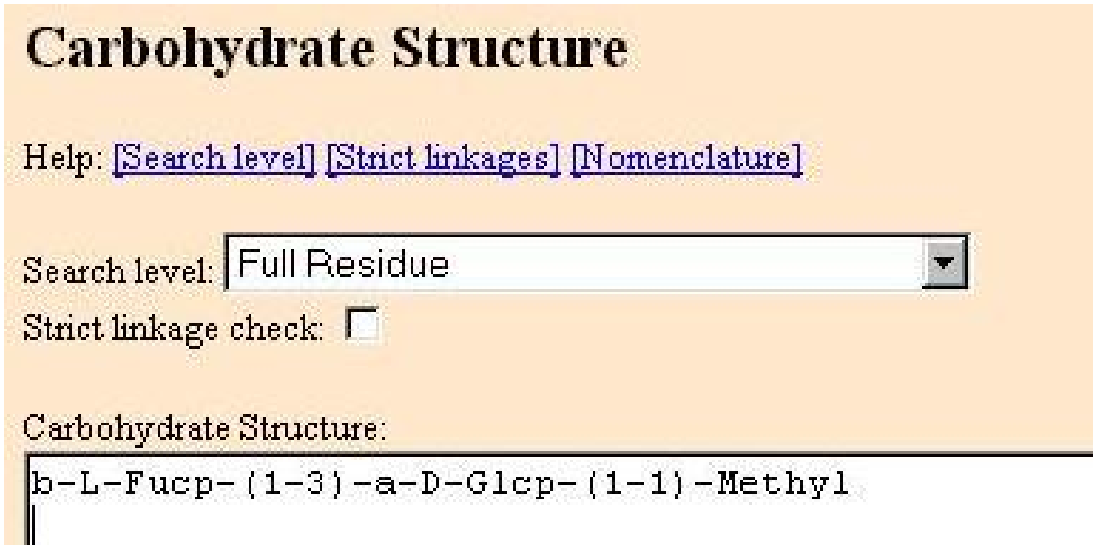
comparison with simulated NMR spectra (CASPER)

requires information about the components and linkages to limit the number of possible structures.

Artificial Neural Networks (ANN)

current application are limited to a single class of compounds.

SugaBase
(http://www.boc.chem.uu.nl/sugabase/databases.html)



SugaBase

C#: P-0201-A00404
CC: CCSD:A00404
MHz 75
Temp 333
Solv D20
Original Reference: Acetone
Reference Value : 31.45
Correction Applied: -0.37

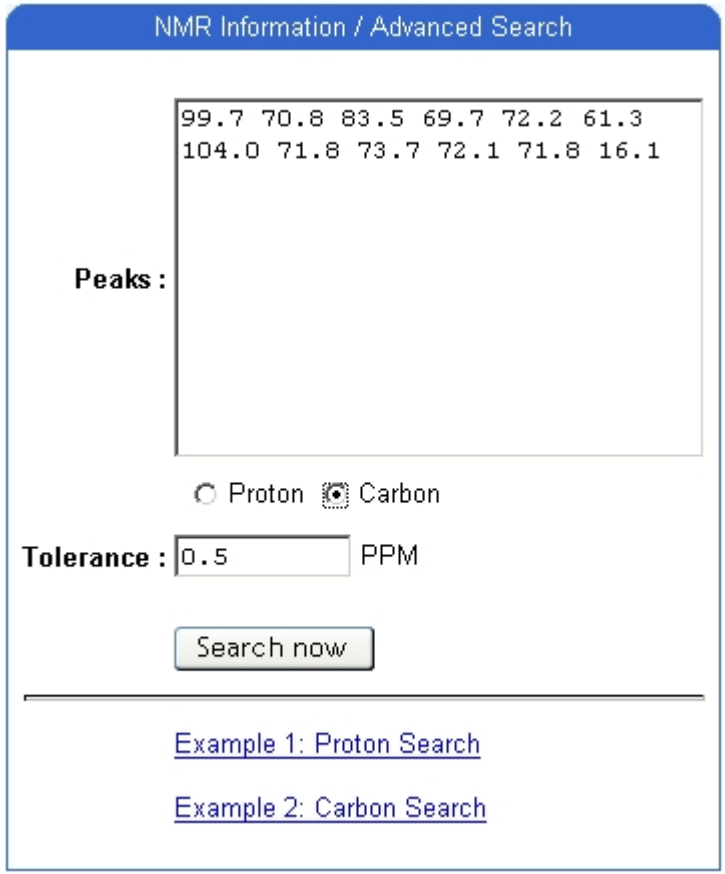
Residue	Linkage	Carbon	PPM	J	Hz	Note
a-D-Glcp		C-1	99.7			
		C-2	70.8			
		C-3	83.5			
		C-4	69.7			
		C-5	72.2			
		C-6	61.3			
b-L-Fucp	3	C-1	104.0			
		C-2	71.8			
		C-3	73.7			
		C-4	72.1			
		C-5	71.8			
		C-6	16.1			

Sweet-DB

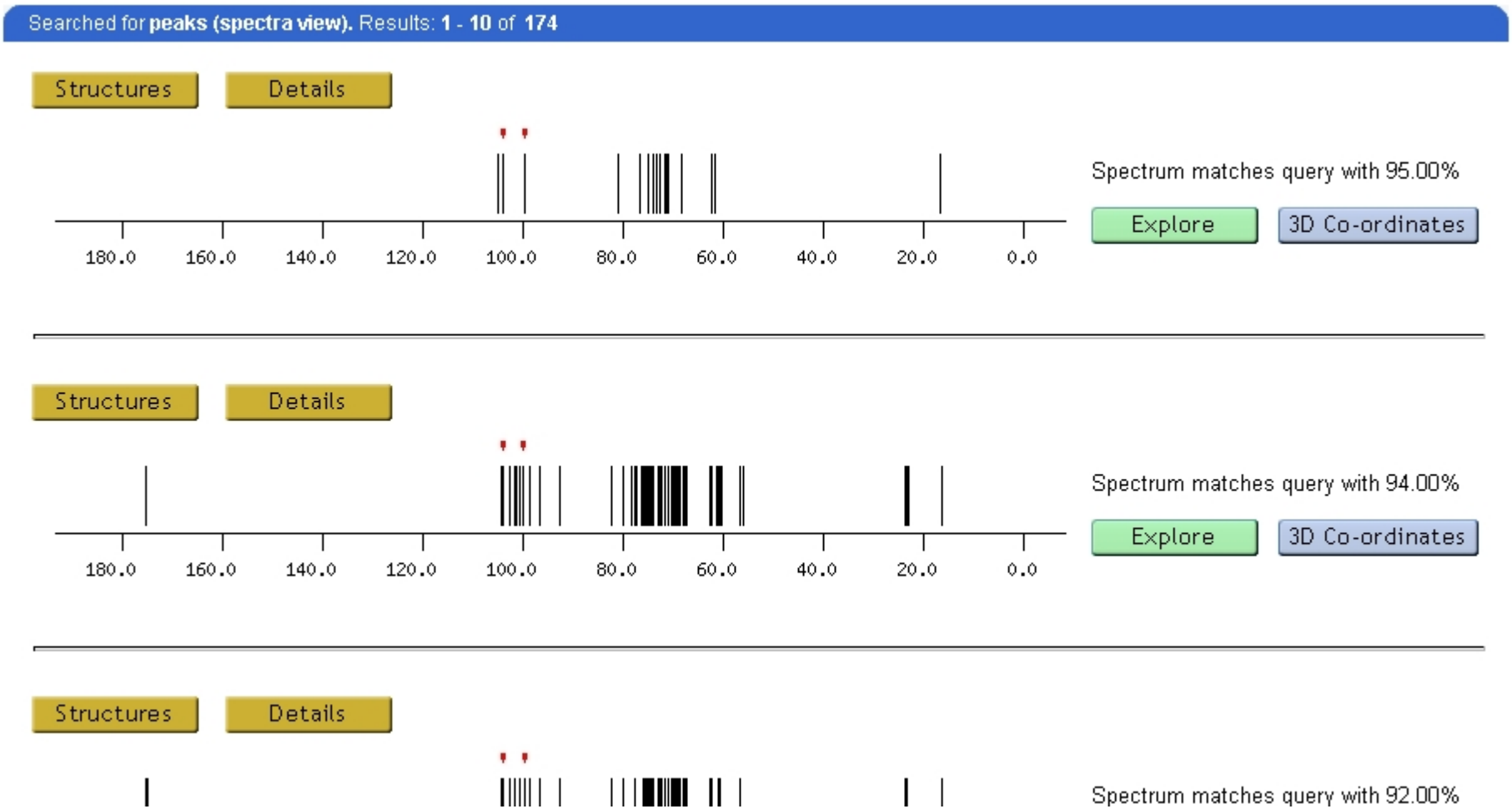
http://glycosciences.de hosts several tools and databases for the carbohydrate chemist

Sweet-DB is an interface to CarbBank/SugaBase

Search with NMR-data



Search results



Search results

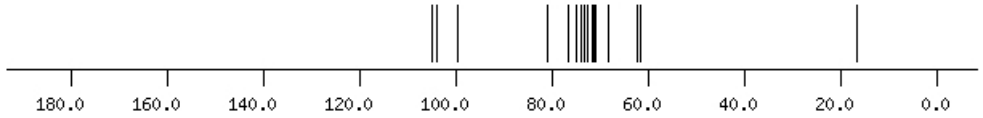
Searched for **nmr information.** Results: **1** - **10** of **174**

Spectra | Details

Structure matches query with 95.00%

```
b-L-Fucp-(1-3)+
              |
         b-D-Glcp-(1-1)-Methyl
              |
a-D-Manp-(1-2)+
```

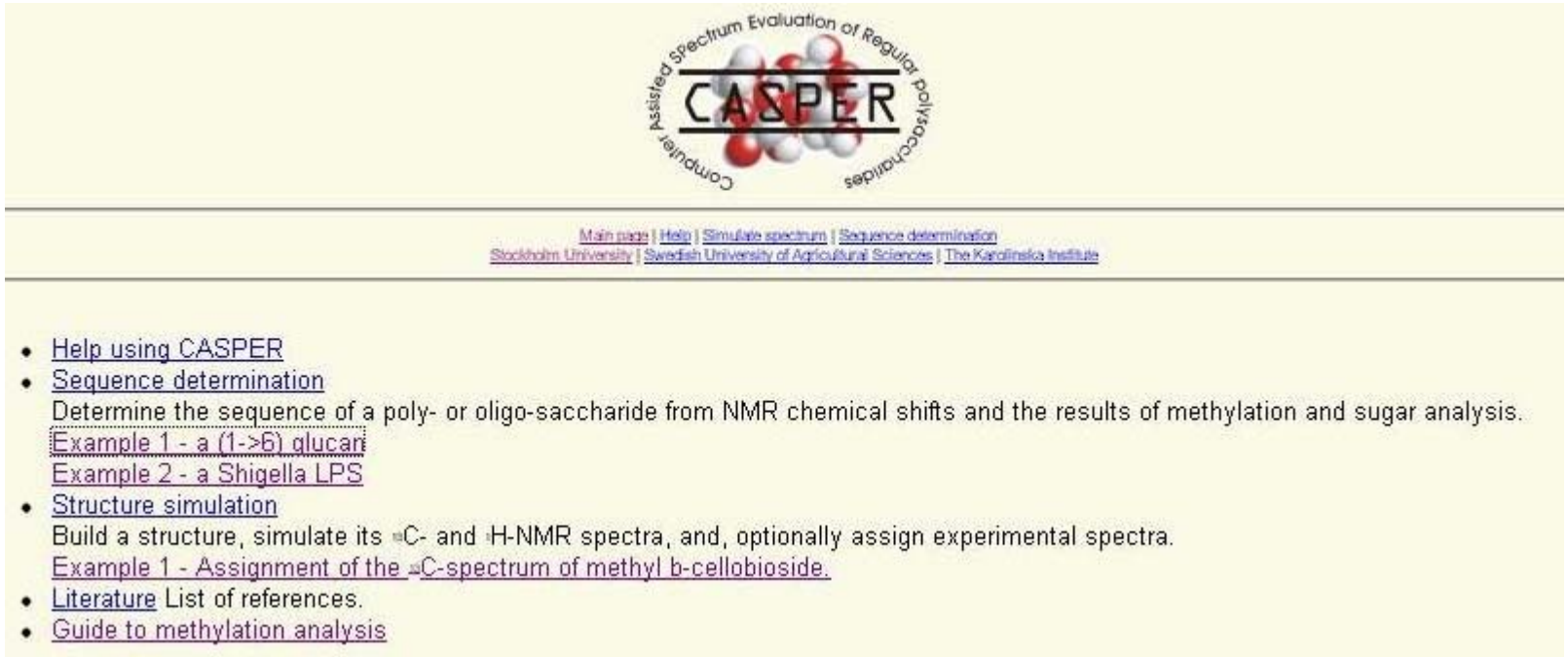
Explore | 3D Co-ordinates



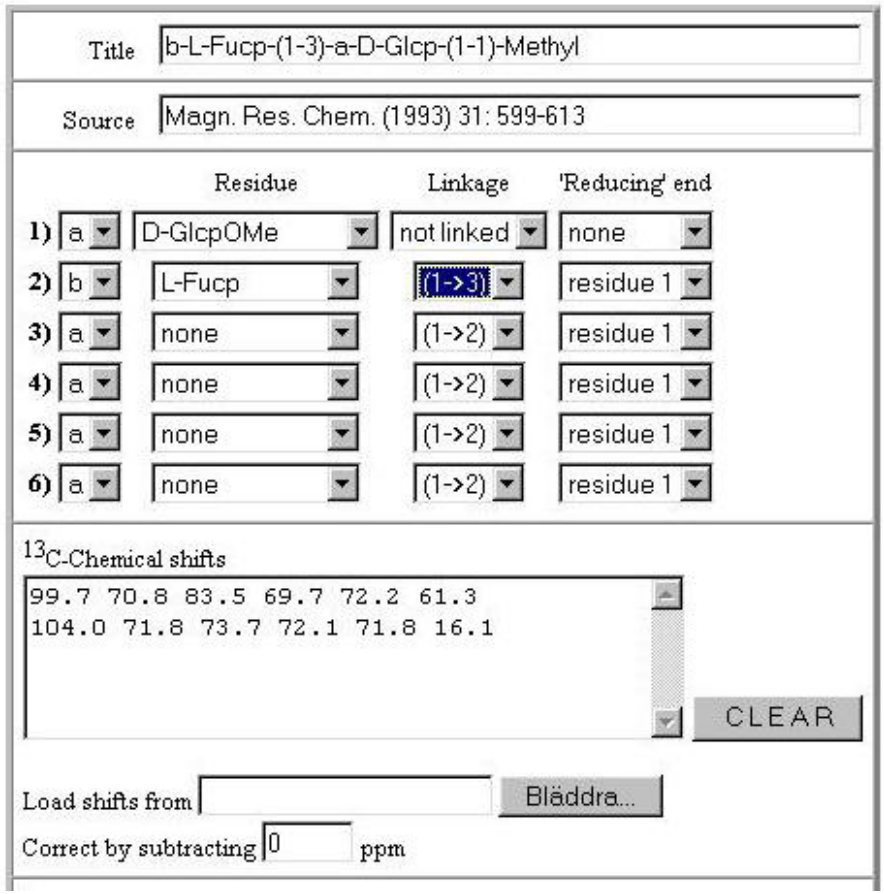
MHZ	75						
Temperature	333						
Solvent	D2O						
Residue	**Linkage**	**Carbon**	**PPM**	**JFrom**	**JTo**	**Hz**	**Note**
b-D-Glcp		C-1	105			0	
b-D-Glcp		C-2	74.800003			0	
b-D-Glcp		C-3	80.900002			0	
b-D-Glcp		C-4	71.400002			0	
b-D-Glcp		C-5	76.5			0	
b-D-Glcp		C-6	61.599998			0	
b-L-Fucp	3	C-1	104			0	
b-L-Fucp	3	C-2	72.400002			0	
b-L-Fucp	3	C-3	74			0	
b-L-Fucp	3	C-4	72.599998			0	
b-L-Fucp	3	C-5	71.400002			0	
b-L-Fucp	3	C-6	16.6			0	
a-D-Manp	2	C-1	99.599998			0	
a-D-Manp	2	C-2	71.199997			0	
a-D-Manp	2	C-3	71			0	
a-D-Manp	2	C-4	68.199997			0	
a-D-Manp	2	C-5	73.099998			0	
a-D-Manp	2	C-6	62.099998			0	

CASPER

(http://www.casper.organ.su.se)



CASPER

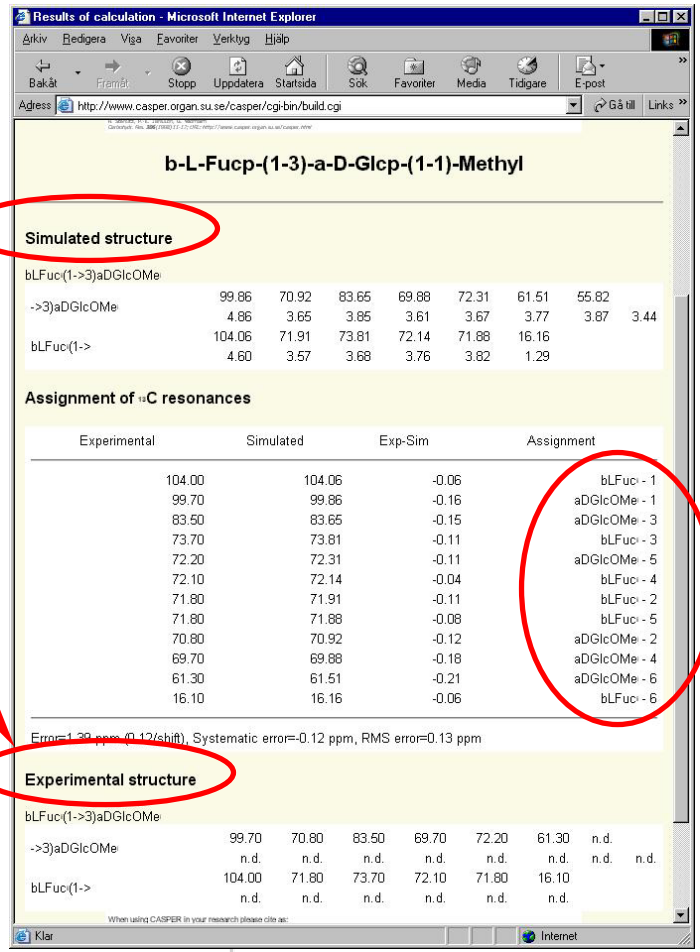

Title
b-L-Fucp-(1-3)-a-D-Glcp-(1-1)-Methyl
Source
Magn. Res. Chem. (1993) 31: 599-613
Residue
Linkage
'Reducing' end
1) a D-GlcpOMe not linked none
2) b L-Fucp (1->3) residue 1
3) a none (1->2) residue 1
4) a none (1->2) residue 1
5) a none (1->2) residue 1
6) a none (1->2) residue 1
13C-Chemical shifts
99.7 70.8 83.5 69.7 72.2 61.3
104.0 71.8 73.7 72.1 71.8 16.1
CLEAR
Load shifts from
Bläddra...
Correct by subtracting 0 ppm

CASPER

Calculated shifts

Experimental shifts

Assignment



CASPER

b-L-Fucp-(1-3)-a-D-Glcp-(1-1)-Methyl

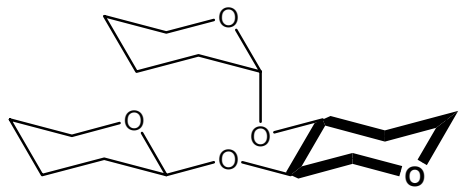
Simulated structure

bLFuc(1->3)aDGlcOMe

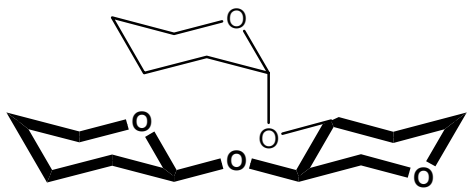
->3)aDGlcOMe	99.86	70.92	83.65	69.88	72.31	61.51	55.82	
	4.86	3.65	3.85	3.61	3.67	3.77	3.87	3.44
bLFuc(1->	104.06	71.91	73.81	72.14	71.88	16.16		
	4.60	3.57	3.68	3.76	3.82	1.29		

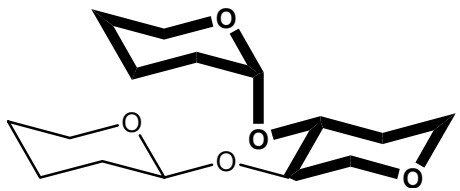
Chemical shift calculation

1) Start with monosaccharide

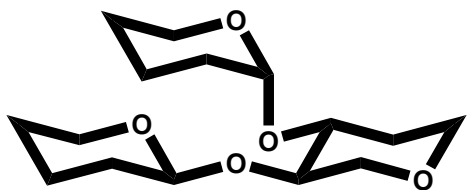


2) Add glycosylation shifts

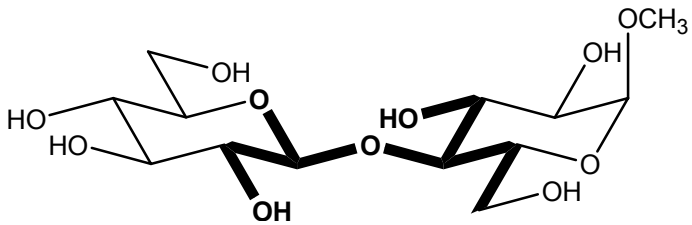




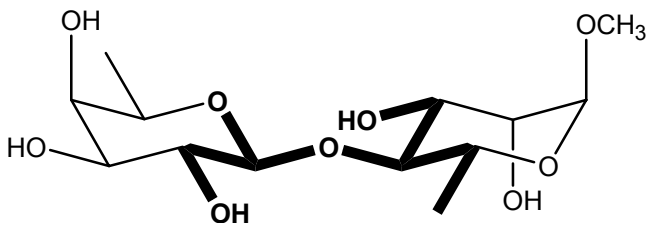
3) Add steric corrections



Glycosylation shifts



βDGlc(1→	6.45	-1.07	-0.18	-0.20	0.12	-0.19
→4)αDGlcOMe	-0.27	-0.25	-1.51	9.20	-1.36	-0.60



βDFuc(1→	6.85	-1.05	-0.13	-0.24	0.18	-0.23
→4)αDRhaOMe	-0.30	-0.66	-1.23	10.19	-1.35	-0.07

Sequence determination

1) Start with all possible components

2) Generate all topologies and sequences

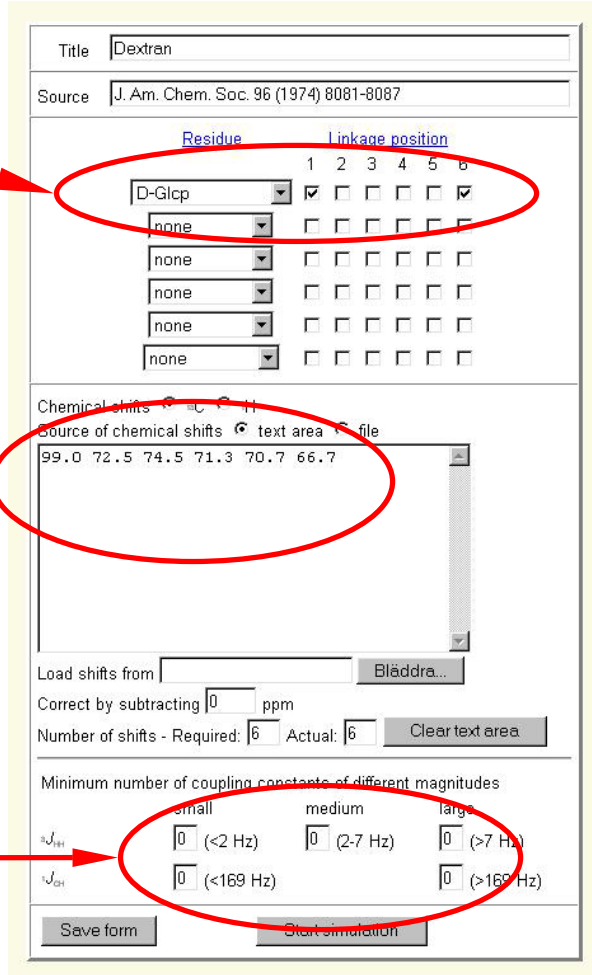
3) Calculate spectra and compare with experimental data

CASPER

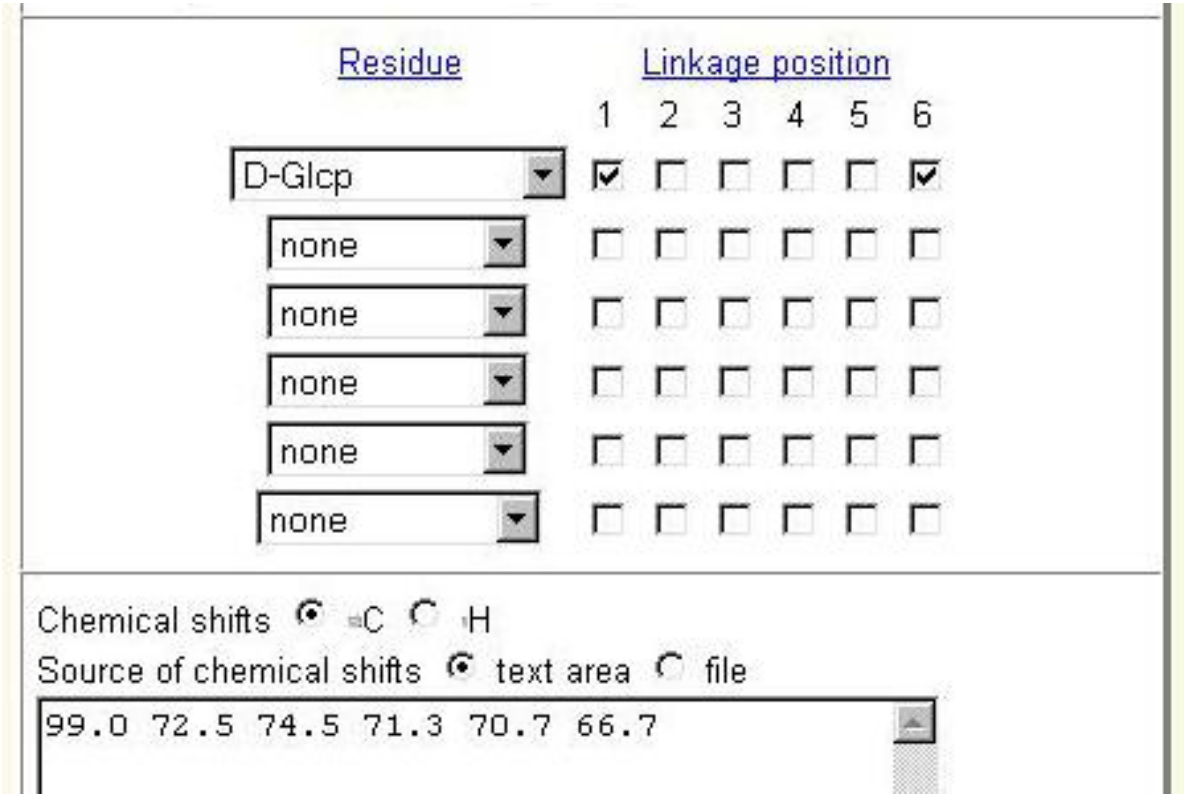
Components
&
linkage positions

Chemical shifts

Additional restrictions



CASPER


Residue
Linkage position
1 2 3 4 5 6
D-Glcp
none
none
none
none
none
Chemical shifts C H
Source of chemical shifts text area file
99.0 72.5 74.5 71.3 70.7 66.7

CASPER

Best fitting structures

1. Structure 1,error=0.65 (0.11)
   ->6)aDGlc(1->
2. Structure 2,error=15.81 (2.64)
   ->6)bDGlc(1->

Structure 1

->6)aDGlc(1->

Simulated structure

->6)aDGlc(1->	98.90 4.97	72.39 3.59	74.32 3.74	70.66 3.54	71.28 3.90	66.90 3.98	3.77

Main weakness of CASPER -

- requires data from chemical analyses to limit the number of structures that are simulated.

An Opportunity for ANN:s?

Caveats

It is assumed that the conformation in the PS is the same as in the disaccharide or trisaccharide fragments

In order to have transferable glycosylation shifts it is also assumed that the monosaccharides are rigid.

Extending the set of disaccharide and trisaccharide fragments used in the calculations may reduce these problems - but they are inherent to the approach.

JRP-2 members

Deutsches Krebsforschungszentrum

Stockholm University

Bijvoet Center for Biomolecular Research

Forschungzentrum Borstel

Swedish University of Agricultural Sciences