On-line tools for the interpretation of NMR and MS-spectra

Roland Stenutz

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

roland@organ.su.se

Overview

What is the problem?

How do you ask the right question?

What answer do you get?

The problem

You have a spectrum of a (pure?) compound and want to know its structure…

…because other information about a compound can often be found only once its structure (identity) is known.

The question

What compounds have similar spectra?

Try to narrow down the problem using additional information – source, physical and chemical properties etc.

The answer

You may

…identify the compound

…identify the class of compounds it belongs to (partial structure)

…or just find out that it is unusual and worth investigating

The answer

How well you succeed depends on…

…the amount and quality of the data

…the type of data (MS vs. NMR)

…the additional restraints you may impose on the answers

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*

MS is insensitive to stereochemistry
- e.g. can't distinguish between Glc & Man

Part 1: NMR

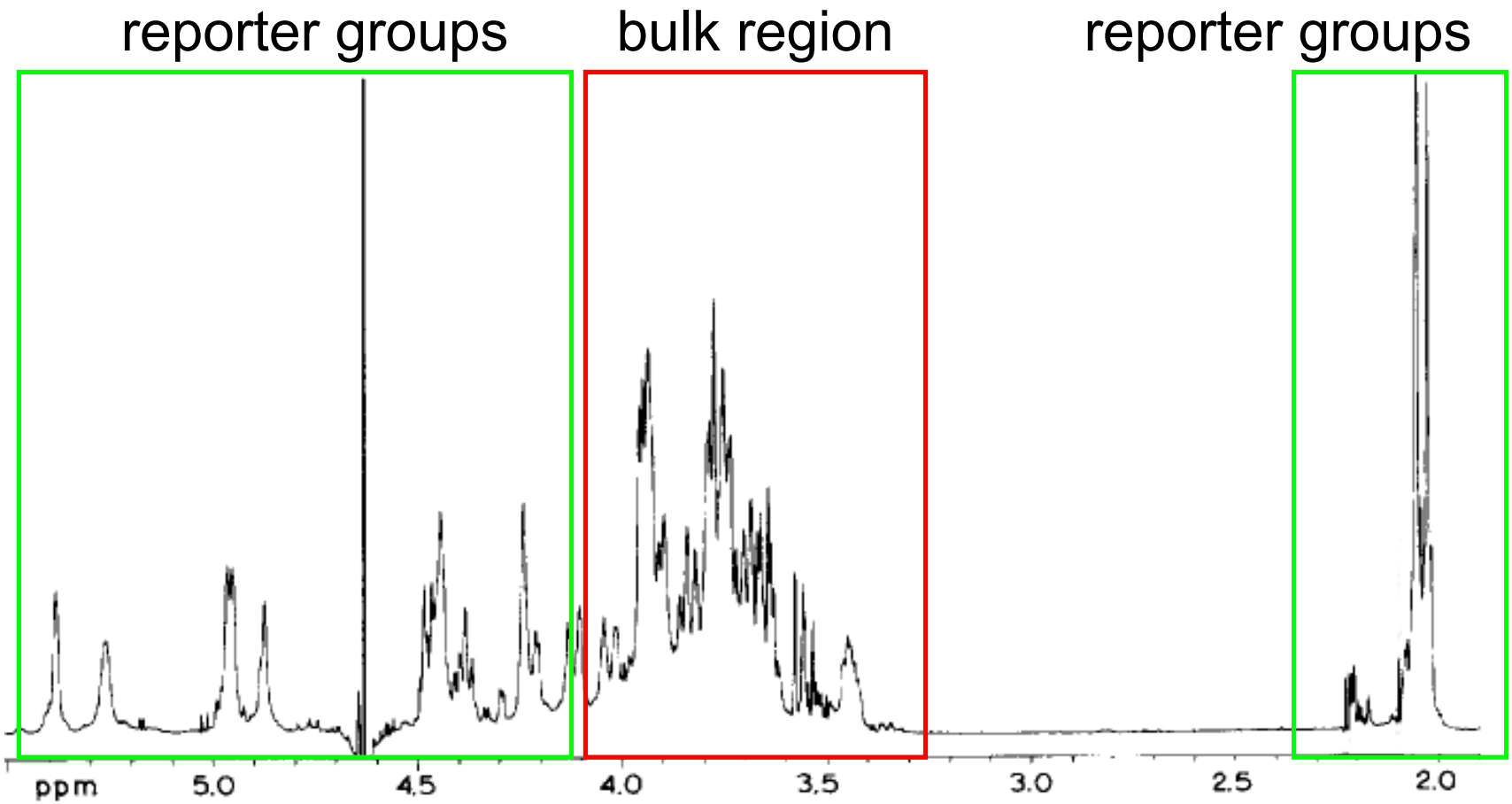
Structure of bacterial polysaccharides

- ^{13}C-NMR
- Polymers of repeating units
- → CASPER

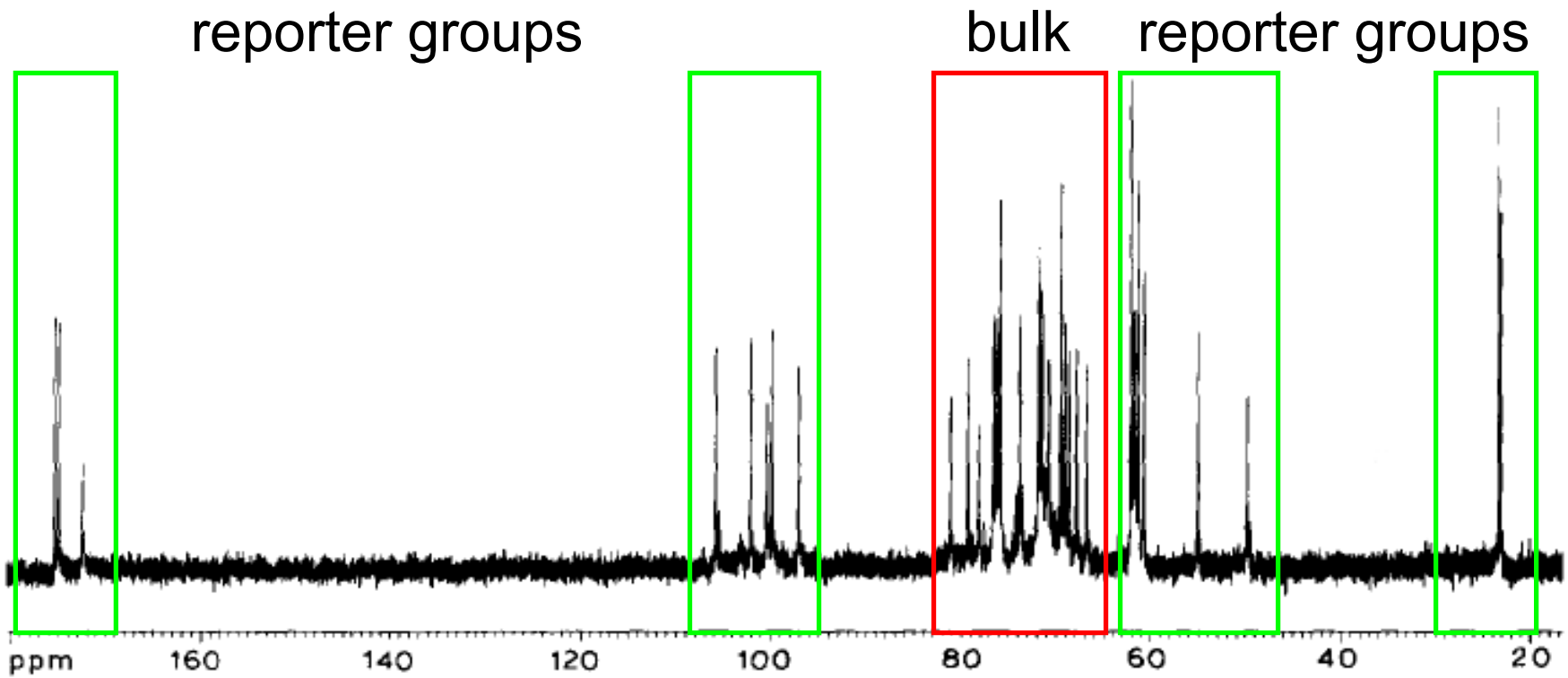
Structure of oligosaccharides from GP

- ^{1}H-NMR
- → SugaBase

Typical ^{1}H spectrum (EC O113)



Typical ^{13}C spectrum (EC O113)



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.

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.

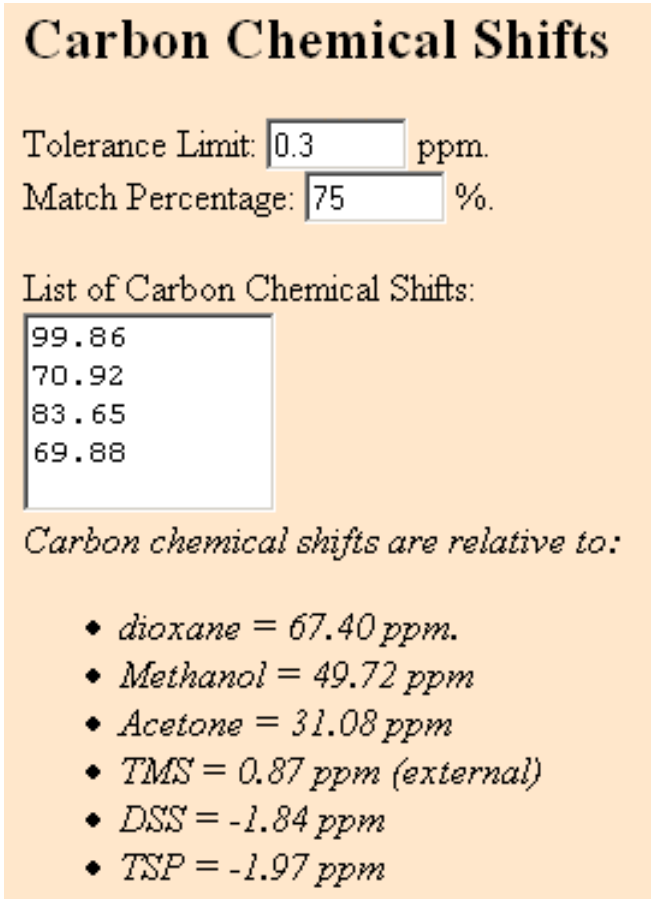
comparison with simulated NMR spectra (CASPER)

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



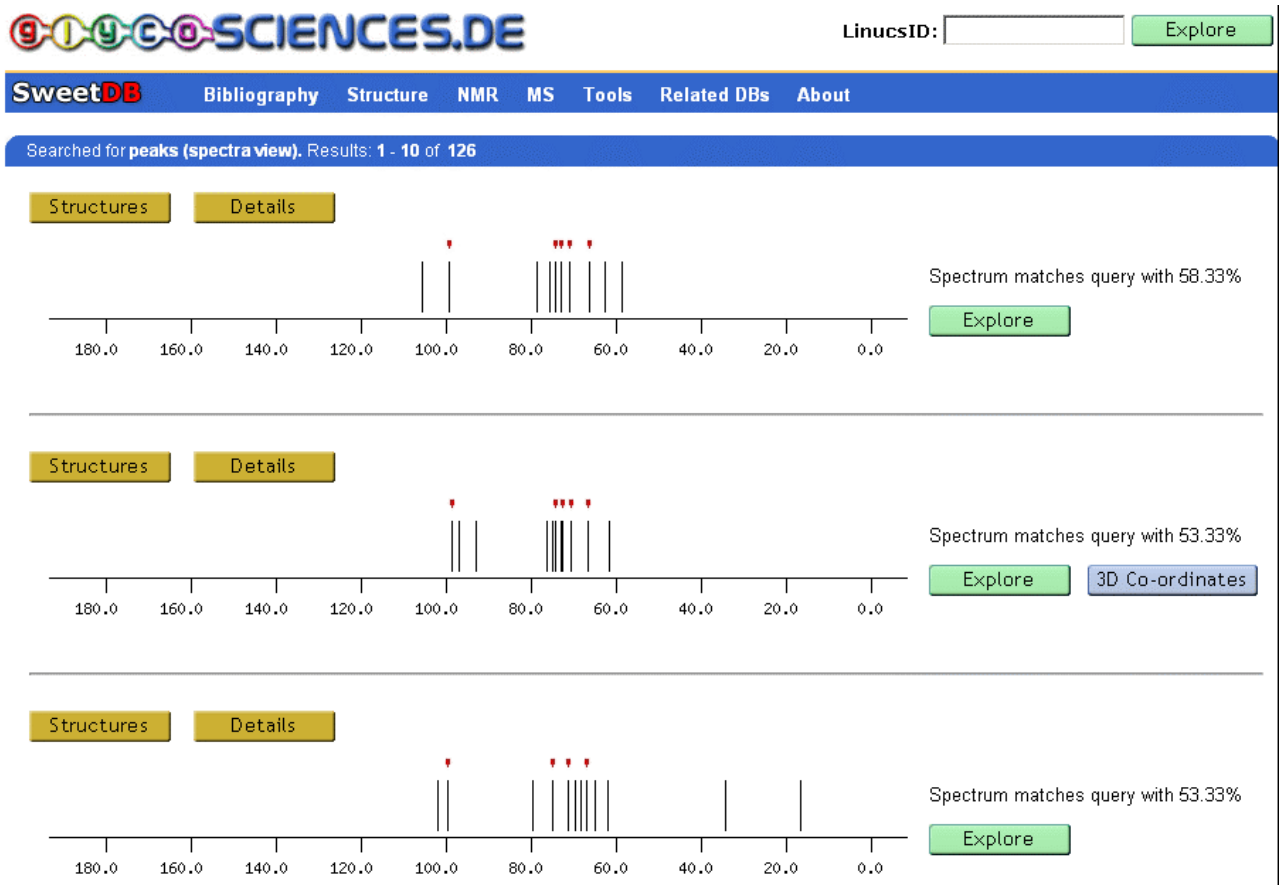
SugaBase

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

CCSD-data + NMR

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			

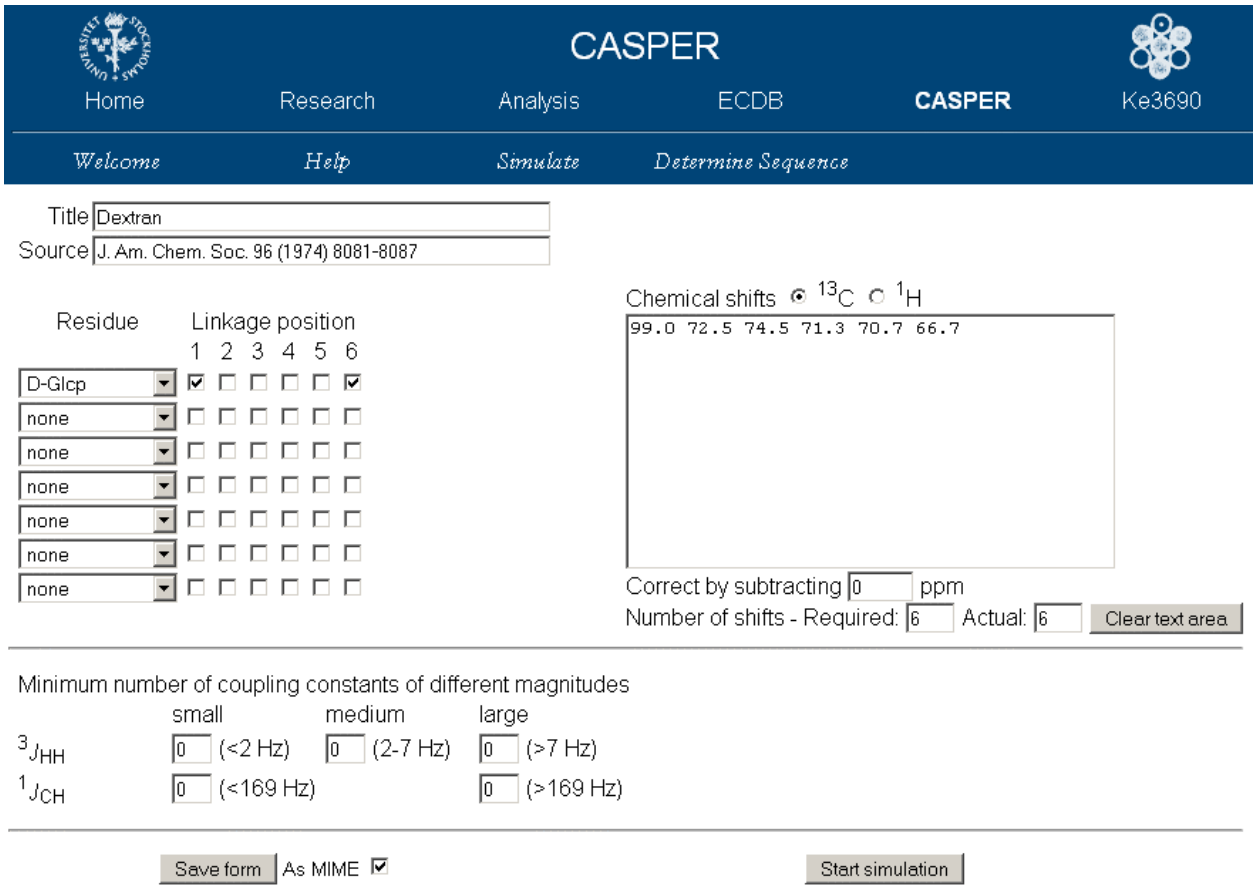
SweetDB – *glycosciences.de*



On-line tools for the interpretation of NMR and MS-spectra

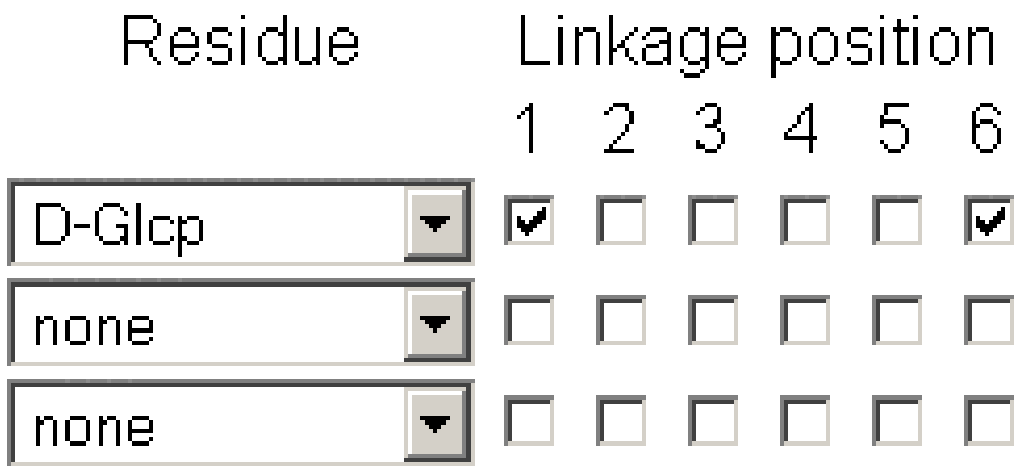
CASPER

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



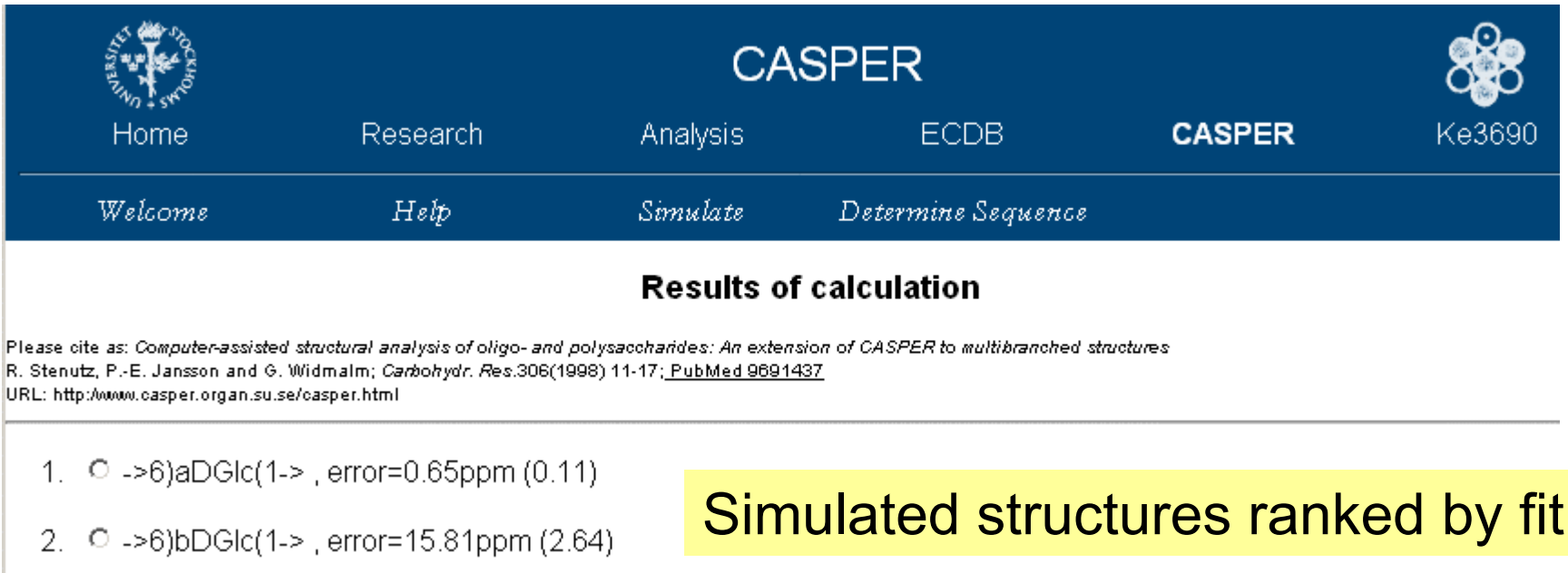
On-line tools for the interpretation of NMR and MS-spectra

CASPER



Data from methylation analysis is used to limit the number of structures generated.

CASPER



Simulated structures ranked by fit

CASPER

Simulated structure

->6)aDGlc^{i}(1->

->6)aDGlc^{i}(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

Assignment of ^{13}C resonances

Experimental	Simulated	Exp-Sim	Assignment
99.00	98.90	0.10	aDGlc^{i} - 1
74.50	74.32	0.18	aDGlc^{i} - 3
72.50	72.39	0.11	aDGlc^{i} - 2
71.30	71.28	0.02	aDGlc^{i} - 5
70.70	70.66	0.04	aDGlc^{i} - 4
66.70	66.90	-0.20	aDGlc^{i} - 6

Error=0.65 ppm (0.11/shift), Systematic error=0.04 ppm, RMS error=0.13 ppm

Experimental structure

->6)aDGlc^{i}(1->

->6)aDGlc^{i}(1->	99.00 n.d.	72.50 n.d.	74.50 n.d.	70.70 n.d.	71.30 n.d.	66.70 n.d.	n.d.

JCAMP-format

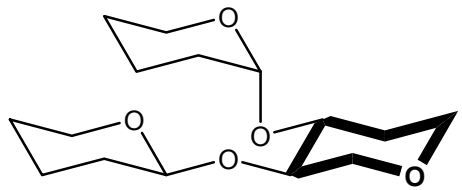
CASPER

Experimental	Simulated	Exp-Sim	Assignment
99.00	98.90	0.10	aDGlc^{i} - 1
74.50	74.32	0.18	aDGlc^{i} - 3
72.50	72.39	0.11	aDGlc^{i} - 2
71.30	71.28	0.02	aDGlc^{i} - 5
70.70	70.66	0.04	aDGlc^{i} - 4
66.70	66.90	-0.20	aDGlc^{i} - 6

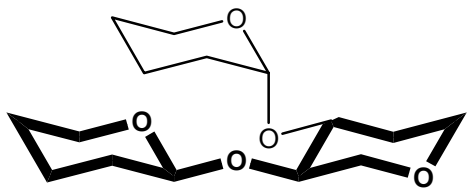
Simulation of NMR

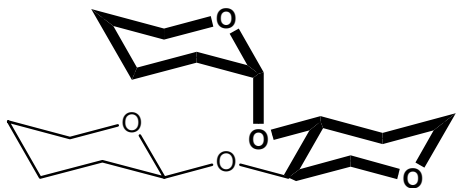
Chemical shift calculation

1) Start with monosaccharide

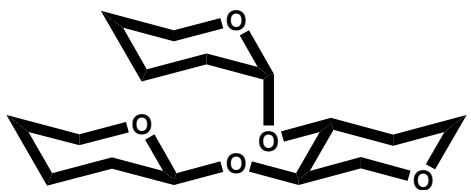


2) Add glycosylation shifts

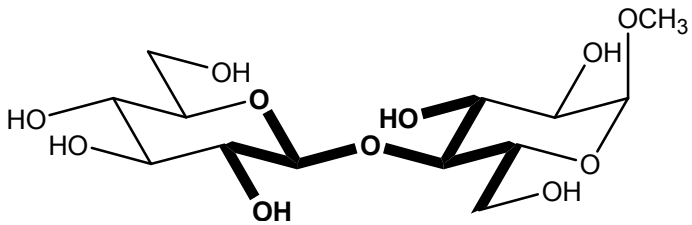




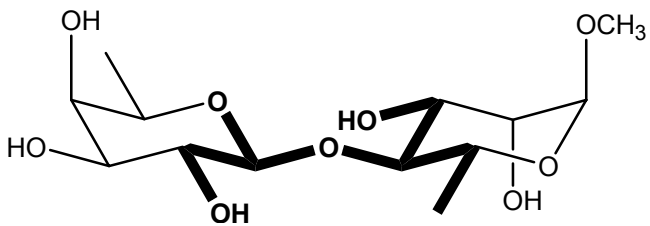
3) Add steric corrections



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

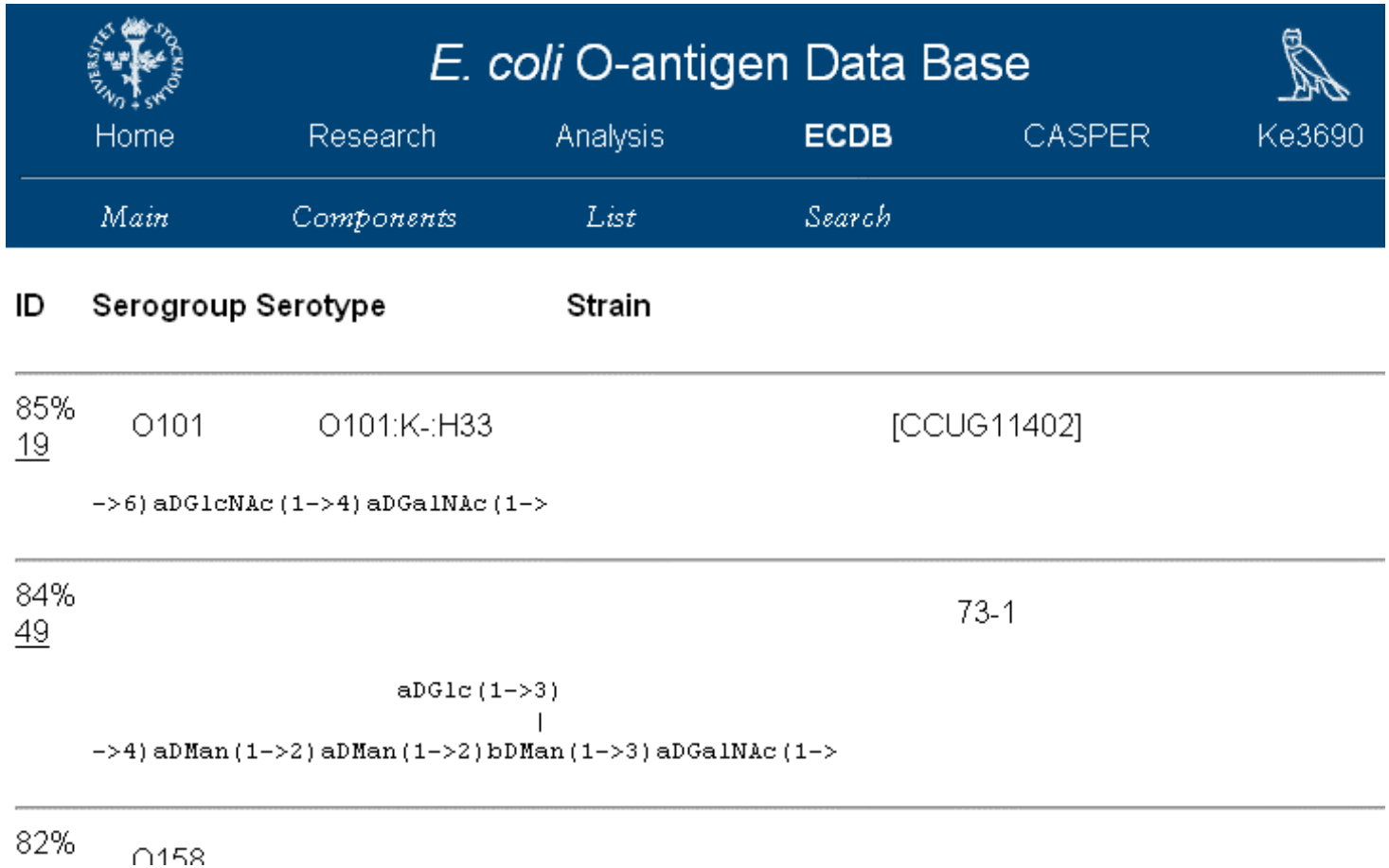
Caveats

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.

ECDB – *E. coli O-antigens*

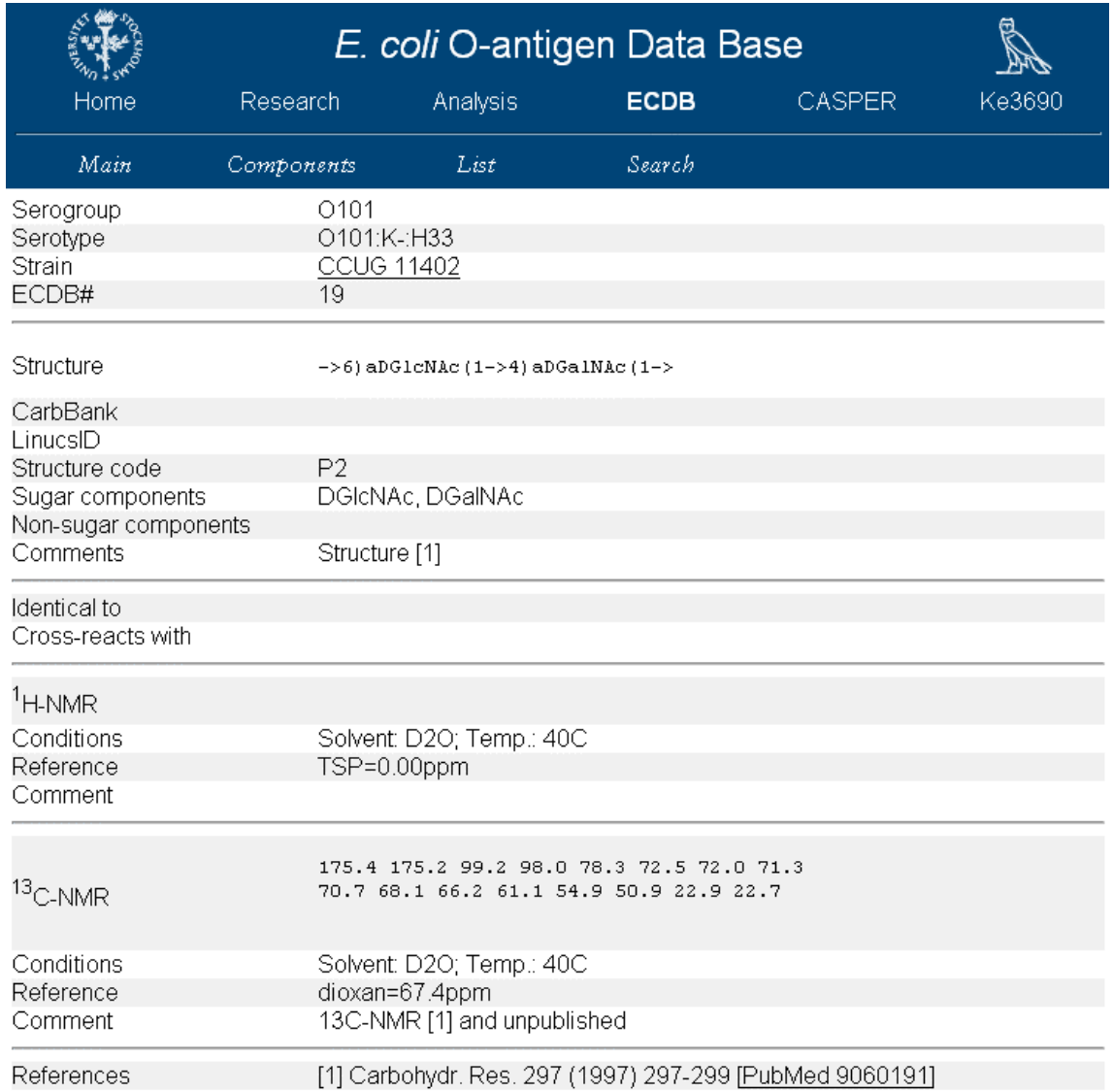


On-line tools for the interpretation of NMR and MS-spectra

NMR

structure

bibliography



E. coli O-antigen Data Base

Home | Research | Analysis | **ECDB** | CASPER | Ke3690

Main | *Components* | *List* | *Search*

Serogroup	O101
Serotype	O101:K-:H33
Strain	CCUG 11402
ECDB#	19
Structure	->6)aDGlcNAc(1->4)aDGalNAc(1->
CarbBank	
LinucsID	
Structure code	P2
Sugar components	DGlcNAc, DGalNAc
Non-sugar components	
Comments	Structure [1]
Identical to	
Cross-reacts with	
^{1}H-NMR	
Conditions	Solvent: D2O; Temp.: 40C
Reference	TSP=0.00ppm
Comment	
^{13}C-NMR	175.4 175.2 99.2 98.0 78.3 72.5 72.0 71.3 70.7 68.1 66.2 61.1 54.9 50.9 22.9 22.7
Conditions	Solvent: D2O; Temp.: 40C
Reference	dioxan=67.4ppm
Comment	13C-NMR [1] and unpublished
References	[1] Carbohydr. Res. 297 (1997) 297-299 [PubMed 9060191]

SDBS –small molecules

www.aist.go.jp/RIODB/SDBS/menu-e.html

Welcome to SDBS

Integrated Spectral Data Base System for Organic Compounds

Japanese

Last updated : March 25, 2004

National Institute of Advanced Industrial Science and Technology

Tsukuba, Ibaraki, Japan

NMR: *T.Saito, K.Hayamizu, M.Yanagisawa and O.Yamamoto*

MS: *N.Wasada*

ESR: *K.Someno*

IR: *S.Kinugasa, K.Tanabe and T.Tamura*

Raman: *K.Tanabe and J.Hiraishi*

Introduction

How to use

Search Compounds / Search NMR & MS / Display Spectra

Access to this database is free of charge, however we would request users not to download more than 50 spectra and/or compound structure files in one day. All accesses are logged. If, for some specific purpose, more spectra are required the user should consult us, and explain the intended usage for the data. We also request that when the database data is used in a publication or presentation, a proper acknowledgement is given such as:
SDBSWeb: http://www.aist.go.jp/RIODB/SDBS/ (access date)