

CASPER Manual

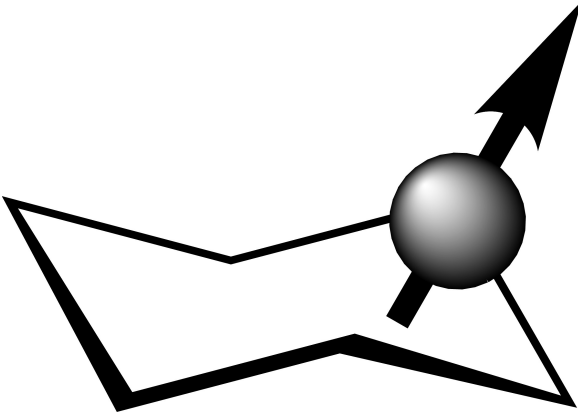


Table of Contents

1.Overview.....	3
2.Important Concepts.....	4
2.1.Structures in CASPER.....	4
2.2.Expected Error.....	4
2.3.CASPER reference data.....	4
3.Spectra calculation.....	6
3.1.Data input.....	6
3.2.Results.....	8
4.Structure determination	10
4.1.Data input.....	10
4.2.Results.....	13
5.Component Analysis.....	14
5.1.Data input.....	14
5.2.Results.....	14
6.Examples using CCPN projects.....	15
6.1.Structure determination of an E. coli O-antigen polysaccharide.....	15
6.2.Assignment verification.....	17
7.Tutorials.....	18
7.1.Calculating ¹³ C and ¹ H chemical shifts of the Escherichia coli O22 O-antigen polysaccharide....	18
7.2.Sequence determination using insufficient linkage data.....	19
Appendix A – CASPER structure representation.....	20

1. Overview

CASPER^{1,2,3} is a tool for calculating NMR chemical shifts of oligo- and polysaccharides. It can also be used for determining the primary structures of unknown glycans based on experimental NMR data. The structure determination is performed by calculating chemical shifts of possible structures and comparing them with the supplied experimental data to find the best match. The process requires experimental NMR data, coupled with results from sugar and methylation analyses if available. CASPER chemical shift calculations are based on the chemical shifts of the monosaccharides in the structure with the glycosylation shifts of the constituent disaccharides applied (how the chemical shifts change when the monosaccharides are linked together). If there are vicinal couplings trisaccharide corrections are also taken into account.

CASPER can use CCPN⁴ projects as data input and can also save results in that format. This means that a set of spectra can be opened in CCPN Analysis to pick the peaks and create resonances followed by loading that information into CASPER to quickly determine the structure and annotate the signals.

CASPER is available at: <http://www.casper.org.au/se/casper/>

CASPER is an academic software used free of charge with no explicit or implicit guarantees of performance or results. CASPER is used at your own risk.

-
- 1 M. Lundborg and G. Widmalm. Structure Analysis of Glycans by NMR Chemical Shift Prediction. *Anal. Chem.* **83** (2011), 1514-1517.
 - 2 M. Lundborg, C. Fontana and G. Widmalm. Automatic Structure Determination of Regular Polysaccharides Based Solely on NMR Spectroscopy. *Biomacromolecules* **12** (2011), 3851-3855.
 - 3 P.-E. Jansson, R. Stenutz and G. Widmalm. Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program CASPER. *Carbohydr. Res.* **341** (2006), 1003-1010.
 - 4 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **59** (2005), 687 - 696.

2. Important Concepts

2.1. Structures in CASPER

By default CASPER uses a text nomenclature for describing structures. Structures in CASPER use square brackets to signify branches, meaning that a simple undecorated N-glycan core would be written as:



This means that there is a branch at the β -D-Man. It is substituted at positions 3 and 6. See Appendix A for more information about structure representations in CASPER.

At the bottom of structure determination results and assignment tables there is a switch to turn on and off graphical (CFG) representation of structures, created using GlycanBuilder⁵. Some structures cannot be displayed graphically.

2.2. Expected Error

When viewing results from a calculation the structure is shown along with the chemical shifts of each atom. Each residue also has an 'Expected Error' value, see Fig. 3. This value shows how reliable the results are and is supposed to reflect the error in the calculated ¹³C chemical shifts in the residue, whereas the ¹H error is lower (dividing the expected error by 5 is often a good approximation). This expected error value is additive and comes from the following sources:

- Low quality of monosaccharide data. Not very common. Error generally low (<0.1)
- Low quality of di- or trisaccharide data, or reference data from conditions different to the CASPER standard. Errors often in the range of 0.1-1.
- Missing trisaccharide correction. Error 5.
- Errors from using a replacement residue (or multiple replacement residues) in a di- or trisaccharide correction. The added error depends on the similarities, around the substituted position, between the replacement residue and the preferred residue. Errors in the range 0.1-3.

The value of the expected error is coloured in varying degrees of red to indicate how trustworthy the calculated chemical shifts are. The lower the expected error value the better. It is not uncommon to have an expected error ~1 and still get very good results. If the value is higher than 2 the chemical shifts of that residue might not be reliable. In some cases it is possible to get very good agreement between calculated and experimental chemical shifts even if the expected error is higher than 5, but it should not be taken for granted.

2.3. CASPER reference data

CASPER calculations are based on a database of mono-, di- and trisaccharide chemical shifts. The monosaccharide chemical shifts are used as starting point for each residue in a structure, which means that all constituent monosaccharides must be present in the database. The disaccharide information is used to apply glycosylation shifts (changes in NMR chemical shifts caused by linking monosaccharides to each other) to the residues. The trisaccharide corrections are only used at vicinal disubstitutions to account for *e.g.* steric effects. If di- or trisaccharide data, of a specific combination, is not present in the database CASPER tries to replace the residues with similar ones, if possible (see section 2.2).

⁵ A. Ceroni, A. Dell and S.M. Haslam. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code for Biology and Medicine* (2007), 2:3.

The CASPER reference data is, with few exceptions, recorded in D₂O at 70 °C and referenced against TSP in D₂O (¹H 0.000 ppm), 1,4-dioxane in D₂O (¹³C 67.40 ppm) or acetone in D₂O (¹H 2.225 ppm, ¹³C 31.07 ppm). When using CASPER, the results will be most accurate if experimental spectra are recorded under the same, or similar, conditions.

3. Spectra calculation

3.1. Data input

The spectra calculation module is available from <http://www.casper.org.au/casper/calculate.php> and is used for calculating the chemical shifts of a known glycan structure. The data is provided by the user through a web interface (see Fig. 1 and 2). Brief help is available for all input fields by clicking the question marks in the user interface. When calculating chemical shifts the only information that is required to submit is the carbohydrate structure.

Calculate Chemical Shifts

Project Browse... ?

Title ?

Source ?

	Residue	Linkage	'Reducing' end
1)	a ▾ D-Galp ▾	not linked ▾	none ▾ ?
2)	a ▾ none ▾	(->2) ▾	residue 1 ▾
3)	a ▾ none ▾	(->2) ▾	residue 1 ▾
4)	a ▾ none ▾	(->2) ▾	residue 1 ▾
5)	a ▾ none ▾	(->2) ▾	residue 1 ▾
6)	a ▾ none ▾	(->2) ▾	residue 1 ▾
7)	a ▾ none ▾	(->2) ▾	residue 1 ▾
8)	a ▾ none ▾	(->2) ▾	residue 1 ▾
9)	a ▾ none ▾	(->2) ▾	residue 1 ▾
10)	a ▾ none ▾	(->2) ▾	residue 1 ▾
11)	a ▾ none ▾	(->2) ▾	residue 1 ▾
12)	a ▾ none ▾	(->2) ▾	residue 1 ▾

¹³C-Chemical shifts ?

¹H-Chemical shifts ?

Correct by subtracting ppm

Correct by subtracting ppm

CLEAR Search ECODAB CLEAR Search ECODAB ?

Show/hide advanced options

Save URL of form Start simulation

Fig. 1. The user interface for calculating NMR chemical shifts of a known glycan structure.

3.1.1. CCPN project

A project in the CCPN format can be uploaded and used as input data. The project must be tarred and gzipped (.tgz) before submitting it to CASPER. The user can manually override or complement data in the project by filling in the structure and respective chemical shifts fields. For example, if there is no ¹³C input in the form it is instead loaded from the project, provided it exists there. It is also possible to e.g. manually input ¹³C chemical shifts and load ¹H data from the CCPN project.

3.1.2. Title

This is the title of the project or the name of the compound.

3.1.3. Structure (required)

The structure of the sugar is specified residue by residue. Each residue line has four different properties:

- Anomeric configuration – alpha (a) or beta (b).
- Residue – the name of this carbohydrate residue (the protein residues Asparagine, Serine and Threonine are also possible to use to allow N- and O-glycans).
- Linkage – the position to which this residue links (to the residue specified by Reducing end).

- Reducing end – the residue number (in the list of residues) this residue links to.

3.1.4. ^{13}C chemical shifts and ^1H chemical shifts

Experimental ^{13}C and ^1H data is written here. This data will be compared to the calculated chemical shifts of the structure resulting in a score showing how well they match. The data can also be compared to NMR data from known *E. coli* O-antigen polysaccharide structures in the database ECODAB using the 'Search ECODAB' button, which can reveal similarities to already studied *E. coli* O-antigen structures.

3.1.5. Correct by subtracting

The experimental data can be corrected for systematic errors by subtracting a value from the submitted chemical shifts. This can be useful, *e.g.*, when different experimental conditions have been used.

3.1.6. Advanced options

The advanced options should only be used with caution.

It is possible to add substituents to a residue. The type of substituent is chosen, the position where it binds to the residue as well as which residue it is connected to. The substituent feature is very limited at this stage. There is not data for many combinations of substituents and residues and it is not thoroughly tested.

3.1.7. Save URL of form

To save your input you can get a URL, which can be added as a bookmark to a web browser or easily inserted into an e-mail. When entering this URL into a web browser the input form will be populated as it was when this button was pressed, **with the exception of the CCPN project input field**. The CCPN project will have to be specified manually every time as a safety measure to make sure the user is aware of which project is sent to CASPER.

3.1.8. Start simulation

Pressing this button starts the simulation. The results should be presented within a few seconds.

Calculate Chemical Shifts

Project Browse... ?
 Title ?
 Source ?

	Residue	Linkage	'Reducing' end
1)	b ▼ D-GalpNAc ▼	(->6) ▼	residue 4 ▼ ?
2)	a ▼ D-Galp ▼	(->3) ▼	residue 1 ▼
3)	b ▼ D-GalpNAc ▼	(->4) ▼	residue 2 ▼
4)	b ▼ D-Galp ▼	(->3) ▼	residue 3 ▼
5)	a ▼ L-Fucp ▼	(->2) ▼	residue 4 ▼
6)	a ▼ none ▼	(->2) ▼	residue 1 ▼
7)	a ▼ none ▼	(->2) ▼	residue 1 ▼
8)	a ▼ none ▼	(->2) ▼	residue 1 ▼
9)	a ▼ none ▼	(->2) ▼	residue 1 ▼
10)	a ▼ none ▼	(->2) ▼	residue 1 ▼
11)	a ▼ none ▼	(->2) ▼	residue 1 ▼
12)	a ▼ none ▼	(->2) ▼	residue 1 ▼

¹³ C-Chemical shifts	¹ H-Chemical shifts
104.7 52.8 78.2 69.6 75.9 62.4	4.53 3.99 3.90 4.05 3.68 3.8 3.8
96.9 69.6 70.5 78.0 71.9 61.5	5.08 3.68 3.87 4.11 3.82
102.6 51.9 77.2 65.4 76.1 62.4	4.59 4.03 3.81 4.15 3.66 3.8 3.8
103.3 77.3 74.9 70.0 74.2 69.3	4.61 3.65 3.81 3.90 3.77 3.79 3.79
100.4 69.5 70.9 73.1 68.0 16.5	5.24 3.78 3.66 3.73 4.22 1.21
23.6 23.8 175.25 175.35	2.045 2.045

Correct by subtracting ppm Correct by subtracting ppm ?
 ?

Show/hide advanced options

Fig. 2. Input for calculating the chemical shifts of the repeating unit of the O-antigen of E. coli O128.

3.2. Results

The calculated chemical shifts are shown and assigned to the atoms in the structure – listed residue by residue. If experimental NMR data was also provided CASPER presents a list of how the experimental and simulated data is matched along with errors for each signal. Finally the experimental chemical shifts are shown as assigned to the atoms in the structure in the same way as the input structure is displayed. An example of the output is shown in Fig. 3.

3.2.1. CCPN Save

This button saves the current assignments (and structure) in a new CCPN project. If a CCPN project was used as input this saved project will be based on that. A CCPN project can be opened, *e.g.*, in CCPN Analysis for further analysis of the assignments. This requires that the carbohydrate ChemComps have been installed (copied to \$CCPNMR_TOP_DIR/data/ccp/molecule/ChemComp) – these are not shipped with a default CCPN installation, but are available at <http://www.casper.org.au/casper/CarbohydrateChemComps.tgz>.

3.2.2. Get 3D Structure

3D structures of the glycan (pdb and psf files, if possible) are generated by CarbBuilder⁶. If the structure is a repeating unit it is possible to specify the number of repetitions.

3.2.3. Get pdf report

This will generate a pdf file containing the currently displayed assignment results.

6 M. Kuttel, G. Widmalm, Y. Mao and M. Lundborg. CarbBuilder: An adjustable tool for building 3D molecular structures of carbohydrates for molecular simulation. *Proc. 7th IEEE Int. Conf. eScience* (2011), 395-402.

Chemical Shift and Assignments

Calculated chemical shifts

--6)α-L-Fuc ^V (1→2)β-D-Gal ^{IV} (1→3)β-D-GalNAc ^{III} (1→4)α-D-Gal ^{II} (1→3)β-D-GalNAc ^I (1→									
	1	2	3	4	5	6	6	Me_2	CO_2
--3)β-D-GalNAc ^I (1→	102.70	51.20	77.30	64.99	75.62	61.87		23.06	174.95
Expected Calc. Error: 1.59	4.47	4.13	3.83	4.15	3.71	3.83	3.86	2.03	
--4)α-D-Gal ^{II} (1→									
	1	2	3	4	5	6	6		
Expected Calc. Error: 3.09	96.16	69.32	70.47	78.74	71.54	61.27			
	5.08	3.87	3.83	4.19	3.82	3.72	3.76		
--3)β-D-GalNAc ^{III} (1→									
	1	2	3	4	5	6	6	Me_2	CO_2
Expected Calc. Error: 3.74	105.05	53.02	76.67	68.69	75.87	62.27		23.15	175.67
	4.48	4.12	4.05	4.12	3.70	3.61	3.83	2.06	
--2,6)β-D-Gal ^{IV} (1→									
	1	2	3	4	5	6	6		
Expected Calc. Error: 2.24	101.97	77.34	74.39	70.12	73.87	69.59			
	4.68	3.61	3.82	3.90	3.73	3.92	4.05		
α-L-Fuc ^V (1→									
	1	2	3	4	5	6			
Expected Calc. Error: 0.83	100.28	69.09	70.48	72.76	67.29	16.12			
	5.21	3.80	3.71	3.78	4.30	1.20			

Assignment of ¹H resonances

Experimental	Calculated	Expt-Calc	Assignment
5.24	5.21	0.03	α-L-Fuc ^V - 1
5.08	5.08	0.00	α-D-Gal ^{II} - 1
4.61	4.68	-0.07	β-D-Gal ^{IV} - 1
4.59	4.48	0.11	β-D-GalNAc ^{III} - 1
4.53	4.47	0.06	β-D-GalNAc ^I - 1
4.22	4.19	0.03	α-D-Gal ^{II} - 4
4.15	4.15	0.00	β-D-GalNAc ^I - 4
4.15	4.13	0.02	β-D-GalNAc ^I - 2
4.11	4.12	-0.01	β-D-GalNAc ^{III} - 4
4.05	4.12	-0.07	β-D-GalNAc ^{III} - 2
4.03	4.05	-0.02	β-D-Gal ^{IV} - 6
3.99	4.05	-0.06	β-D-GalNAc ^{III} - 3
3.90	3.92	-0.02	β-D-Gal ^{IV} - 6'
3.90	3.90	0.00	β-D-Gal ^{IV} - 4

Fig. 3. Top part of the results page, of the O128 simulation, showing the simulated structure along with chemical shifts for all atoms followed by a list of assignments and their differences. The list above is cropped.

4. Structure determination

4.1. Data input

The user interface for determining the structure of an unknown carbohydrate (Fig. 4 and 5) is available at <http://www.casper.organ.su.se/casper/determine.php> and is different to the chemical shift calculation input (chapter 3).

Determine Glycan Structure

Project ?

Title ?

Source ?

Residue	Linkage position							Chemical shifts	
	1	2	3	4	5	6	*	¹³ C	¹ H
D-Galp	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>

Chemical shifts 1D 2D ?

¹³C chemical shifts

¹H chemical shifts

Correct by subtracting ppm

Number of chemical shifts - Anticipated: Actual:

Minimum number of coupling constants of different magnitudes ?

	small	medium	large
³ J _{HH}	<input type="text"/> (<2 Hz)	<input type="text"/> (2-7 Hz)	<input type="text"/> (>7 Hz)
¹ J _{CH}	<input type="text"/> (<169 Hz)		<input type="text"/> (>169 Hz)

Show/hide advanced options

Fig. 4. The interface for structure determination.

4.1.1. CCPN project

A project in the CCPN format can be uploaded and used as input data. The project must be tarred and gzipped (.tgz) before uploading. The user can manually override data in the project by filling in the structure and respective chemical shifts fields. Structure data from the CCPN project is only loaded from unlinked polymer residues in the first molecule (in the CCPN project) not belonging to a specific chain. It is not recommended to specify the residues in the CCPN project, especially if the previous sentence was difficult to understand, since it is more convenient to specify it manually in the CASPER interface. The chemical shifts are loaded from the project if they are not provided in the form. Coupling constants cannot currently be loaded from the project and must therefore be entered by hand, if they are available.

4.1.2. Title

This is the title of the project or the name of the compound.

4.1.3. Residues and Linkage positions (required)

Possible residues and their free linkage positions are listed here. This data comes from sugar and methylation analyses. If a residue is unknown or uncertain it is possible to select one of the different unknown alternatives, e.g. "Unknown" or "Unknown HexNAc". All selected linkage positions for all residues are used when building the possible structures. If linkage positions are unknown the '*' column

must be checked. This means that all selected linkage positions can be used, but they do not have to be. Combining unknown residues and unknown linkage positions increase the calculation times drastically since so many structure combinations are generated. It is therefore recommended to specify the sugar components and linkage positions as accurately as possible.

4.1.4. Chemical shifts (required)

Data from 1D ^{13}C , ^1H and 2D ^1H - ^{13}C HSQC/HETCOR, H2BC and HMBC as well as ^1H - ^1H TOCSY, spectra are typed in the respective fields. To switch between 1D and 2D input click the toggle buttons at the top. Regardless if 1D or 2D is selected when CASPER is started, data of both types are used if they are provided. Chemical shifts can be separated by comma or whitespace. 2D input should be in the order ^{13}C chemical shift followed by ^1H chemical shift pairwise. Below the input there is a counter showing how many chemical shifts are provided and how many would be anticipated based on the residues used. It is good to input as many values as possible, but if peaks overlap it is enough to input just one chemical shift. Providing more data to CASPER is adviceable (higher number of chemical shifts from different kinds of NMR experiments), but it is very important not to submit peaks originating from heterogeneities or noise.

4.1.5. Correct by subtracting

The experimental data can be corrected for systematic errors by subtracting a value from the submitted chemical shifts (this value may also be negative resulting in an addition instead). This correction can be useful *e.g.* when different referencing conditions have been used. Corrections for 1D will be used for correcting 2D chemical shifts too.

4.1.6. Minimum number of coupling constants of anomeric atoms

To limit the number of structures generated, experimental coupling constants can be given. CASPER will only use structures that have *at least* as many coupling constants in each magnitude span as specified here. This can dramatically reduce calculation times. Both $^3J_{\text{HH}}$ and $^1J_{\text{CH}}$ of anomeric atoms can be used.

4.1.7. Advanced options

The advanced options should only be used with caution.

It is possible to add substituents to a structure determination and also specify which residue it is linked to and at which position, if that is known.

If the linkages between any of the residues are known beforehand it is possible to specify that as well. If it is known only which residues are linked together the position can be set to unknown.

If the biological source of the sample is known the structure determination can be guided by rules based on structural features of the kind of glycan. Currently there are rules for N-glycans, O-glycans, WecA assembled repeating units of polysaccharides, *Shigella flexneri* polysaccharides and *Haemophilus influenzae* oligosaccharides. Applying the appropriate rule will shorten the calculation times as well as improve the results since irrelevant structures are discarded before the NMR chemical shifts are compared.

In addition to the options above there are alternatives to increase the diversity of the generated structures:

Free anomer in reducing end – If the reducing end is free there will be a mixture of anomeric forms. If this option is selected CASPER will rank the results list by the sum of the δ_{Rel} (see section 4.2.), of the two anomeric forms and group them together, while the score is reported separately for each anomeric form. This can be a bit confusing at first since the results list will not look sorted at first sight, since it is

sorted on two entries at a time. Since both structures are assigned independently some experimental signals may not be used if they are not close enough for either of the anomeric forms, resulting in an overestimation of the agreement between experimental and calculated chemical shifts. Use this option with caution.

Uncertain absolute configurations – Both D and L forms of the residues will be generated for each structure.

Uncertain ring size – In some cases it is not known beforehand if the residues are in pyranose or furanose ring form. By selecting this option both alternatives will be used in the structure permutations.

4.1.8. Save URL of form

See section 3.1.7.

4.1.9. Start determination

Pressing this button starts the calculations. The time required depends on how many structures must be generated and how complete the chemical shift information is. It takes longer if 2D data is used.

Determine Glycan Structure

Project Browse... ?

Title ?

Source ?

Residue Linkage position Chemical shifts 1D 2D ?

Residue	Linkage position						C-H chemical shifts	TOCSY H-H correlations from anomers
	1	2	3	4	5	6 *		
D-Manp	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	103.77 4.48 100.3 4.6	5.15 4.2 5.15 4.1
D-Manp	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	100.2 5.15 97.7 4.94 95	5.15 3.93
D-Manp	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	5.15	4.94 4.09 4.94 3.88
D-GlcpNAc	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	79.02 3.94 79.36 3.74	4.6 3.99 4.6 3.86
D-GlcpNAc	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	77.38 4.2	4.6 3.76 4.6 3.61
D-Galp	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	77.16 4.14 76.18 3.74	4.48 3.95 4.48 3.68
D-Galp	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	75.57 3.59 73.35 3.67	4.48 3.56
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	72.86 3.74	
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	71.8 3.67 71.8 3.55	
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	71.1 4.09 70.4 3.87	
none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	70.2 3.91 69.38 3.94	

Number of chemical shift pairs

HMBC C-H correlations from anomers

103.77 3.74 103.77 3.55
100.3 3.76
79.36 4.48
71.8 5.15
71.8 4.48
73.3 4.48

Number of chemical shift pairs

H2BC C-H correlations from anomers

103.77 3.55
100.3 3.76
71.8 4.48
55.72 4.6

Minimum number of coupling constants of different magnitudes ?

³J_{HH} small (<2 Hz) medium (2-7 Hz) large (>7 Hz)

¹J_{CH} small (<169 Hz) large (>169 Hz)

Show/hide advanced options

Fig. 5. Example input of a heptasaccharide, utilizing four types of 2D experiments and some linkage positions unknown.

4.2. Results

When the calculations are complete a list of the 10 best structures is shown along with their "score" (δ_{Rel}), which is derived from the matching errors of the experimental data (see Fig. 6). Different experiment types have different contributing factors when calculating δ_{Rel} , and the score is divided by the fraction of signals that could be assigned. The formula for calculating the score is $\delta_{\text{Rel}} = (0.1 \times \delta_{\text{C}} + 5.0 \times \delta_{\text{CH}} + 1.0 \times \delta_{\text{HH}} + 2.5 \times \delta_{\text{CHLR}}) / \text{Assignment Completeness}$. The δ_{Rel} score is then calculated relative to the best structure. The highest ranked structure always has $\delta_{\text{Rel}} = 1.00$ and the lower δ_{Rel} the better agreement between the submitted experimental data and the calculated chemical shifts.

"Show/hide score details" gives more detailed information about the actual differences for each experiment submitted as well as the assignment completeness.

"Toggle between text and graphical representation of structures" enables viewing the glycans in the CFG representation instead of the CASPER text structures. Sometimes graphical structures cannot be created.

Clicking on one of the structures will bring up the detailed assignment lists similar to the output from spectra simulation (see chapter 3).

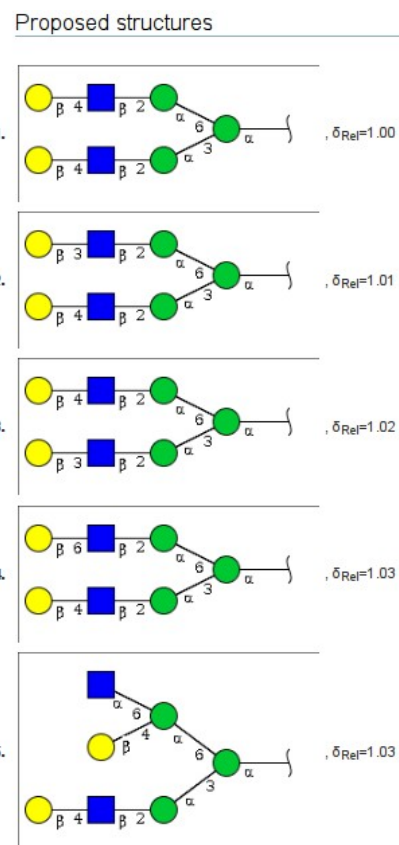


Fig. 6. Top 5 structures of the results from the calculations using the data in figure 5. Structures are represented graphically in the CFG format, by default the text representation is used.

5. Component Analysis

The sugar components, of a poly- or oligosaccharide, and their absolute configurations can be determined using the CASPER component analysis module, available at <http://www.casper.org.au.se/casper/component.php>. In order to facilitate determination of absolute configurations the sample must be butanolysed with optically active butanol. If only the types of residues, in the sample, need to be revealed the sample can be hydrolysed or methanolysed instead. The NMR chemical shifts of the derivatised components will be compared to reference chemical shifts in the CASPER database. The experimental data can be a combination of 1D ^1H and ^{13}C and 2D ^1H - ^{13}C HSQC/HETCOR experiments. In most cases HSQC/HETCOR or ^{13}C will give the best results. The experimental chemical shifts will be compared to both anomeric forms as well as both ring forms (pyranose and furanose) of each monosaccharide, with the specified derivatisation, in the CASPER database. The chemical shift comparisons are performed for one component at a time, which means that chemical shifts that do not match any signal might not be used. In the results there is a list of unused chemical shifts. If this list is very long there are probably unidentified components in the sample. In theory each component in the sample may be observed in at least four different forms after lysis. In practice often only two or three forms can be observed. CASPER does not need chemical shifts from all forms to be observed, but if they are they will be used in the analysis.

5.1. Data input

The input to the component analysis consists of chemical shifts from one or more types of experimental spectra and also how the sample was derivatised. The experimental chemical shifts can be submitted using a CCPN project.

5.2. Results

The results are presented as a list of components that are predicted to be present in the sample. Along each component can be seen the experimental chemical shifts that were used to identify it. CASPER does not require all experimental shifts to be used in the predictions, but rather matches the experimental shift that is in best agreement with a calculated chemical shift. This means that some experimental chemical shifts might not be used in the predictions, especially if there are many similar chemical shifts, and these are listed at the bottom of the results. This list should always be scrutinised – if there are many unused chemical shifts there might be components in the sample that CASPER could not identify. But if the list of unused chemical shifts is short or if most of them are very similar to chemical shifts that were matched to calculated chemical shifts the predictions can be expected to be of high quality.

After analysing the results it is possible to proceed to the structure determination with the predicted components automatically selected. It is important to keep in mind that the component analysis can not make any quantitative comparisons, which means that if a residue is present more than once in an oligosaccharide or the repeating unit of a polysaccharide it will have to be specified as separate units in the structure determination interface.

6. Examples using CCPN projects

6.1. Structure determination of an *E. coli* O-antigen polysaccharide

This is a step-by-step example illustrating the structure determination and resonance assignment workflow starting from processed NMR spectra of an unknown *E. coli* O-antigen polysaccharide.

6.1.1. Acquiring spectra

- A 1D ^1H spectrum and a 2D $^1\text{H},^{13}\text{C}$ -HSQC spectrum were acquired at 30 °C.

6.1.2. Creating a CCPN project in CCPN Analysis

- The spectra were opened in CCPN Analysis (see Fig. 7).
- Peaks were picked and resonances assigned (see Fig. 8).
 - First the resonances in the HSQC spectrum were created and then they were assigned to the corresponding peaks in the 1D spectrum.
 - If there is any uncertainty whether a peak is a true signal or only noise it is important not to create any resonances for it since CASPER will try to use all signals in structure determination and assignments.
- The project was saved and the project directory was archived using tar and gzip in order to facilitate uploading it to CASPER in the next step.
 - `'tar -czf <archive name>.tgz <CCPN project>'`

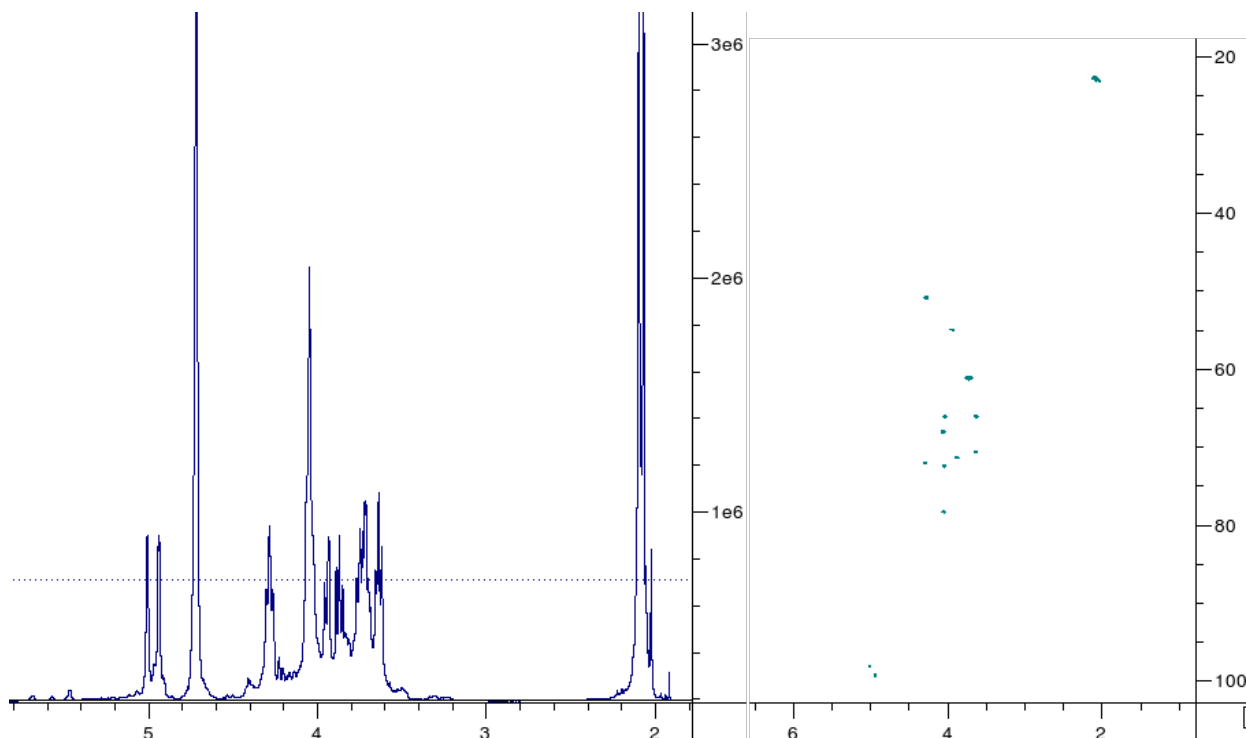


Fig. 7: ^1H and $^1\text{H},^{13}\text{C}$ -HSQC NMR spectra in CCPN Analysis before picking peaks and creating resonances.

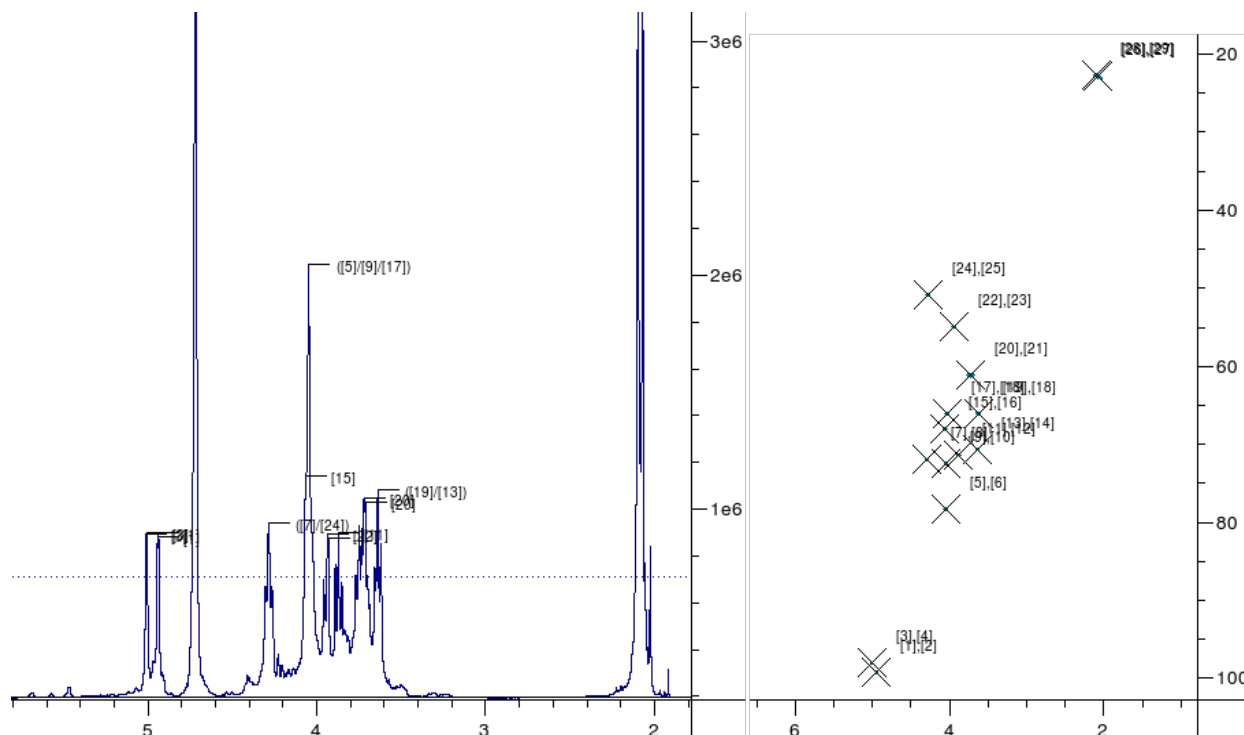


Fig. 8: Resonances are not yet assigned to atoms.

6.1.3. Structure determination using CASPER

In the CASPER structure determination interface the newly created archive of the project was specified in the Project field and two unknown residues with unknown linkage positions were selected in the Residue fields, based upon the number of signals in the anomer region of the NMR spectra as well as the total number of signals. Since it was known that this was an *E. coli* O-antigen the WecA Repeating Unit biological rule was selected (available after clicking Show/hide advanced options). No more information was entered before starting the structure determination. The calculations took approximately one minute, due to the unknown residues and substitution positions. The list of results can be seen in Fig. 9.

After clicking on the first structure in the list the assignments were reviewed and finally the project was saved by clicking the 'CCPN Save' button. After this the assignments can be studied in CCPN Analysis by loading the project containing the CASPER assignments (see Fig. 10).

Proposed structures

1. $\rightarrow 6$ - α -D-GlcNAc-(1 \rightarrow 4)- α -D-GalNAc-(1 \rightarrow , δ_{Re} =1.00
2. $\rightarrow 4$][α -D-GlcNAc-(1 \rightarrow 6)] α -D-GalNAc-(1 \rightarrow , δ_{Re} =1.69
3. $\rightarrow 4$][α -D-GalNAc-(1 \rightarrow 6)] α -D-GlcNAc-(1 \rightarrow , δ_{Re} =1.90
4. $\rightarrow 3$][α -D-GalNAc-(1 \rightarrow 6)] α -D-GlcNAc-(1 \rightarrow , δ_{Re} =1.93
5. $\rightarrow 6$ - α -L-GlcNAc-(1 \rightarrow 3)- α -D-GalNAc-(1 \rightarrow , δ_{Re} =2.01
6. $\rightarrow 6$ - α -L-GalNAc-(1 \rightarrow 4)- α -D-GlcNAc-(1 \rightarrow , δ_{Re} =2.02
7. $\rightarrow 6$ - α -D-ManNAc-(1 \rightarrow 4)- α -D-GalNAc-(1 \rightarrow , δ_{Re} =2.02
8. $\rightarrow 4$][α -L-GlcNAc-(1 \rightarrow 6)] α -D-GalNAc-(1 \rightarrow , δ_{Re} =2.11
9. $\rightarrow 6$ - α -D-GalNAc-(1 \rightarrow 4)- α -D-GlcNAc-(1 \rightarrow , δ_{Re} =2.17
10. $\rightarrow 4$][α -D-GalNAc-(1 \rightarrow 6)] α -D-GalNAc-(1 \rightarrow , δ_{Re} =2.25

Fig. 9: The structure of the *E. coli* O101 O-antigen is correctly determined with a significant margin to the second best match.

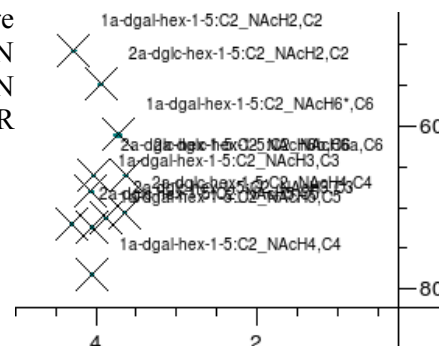


Fig. 10: All signals in the HSQC spectrum have been assigned by CASPER. Only a limited region of the spectrum is shown.

6.2. Assignment verification

In the normal case CASPER only uses unassigned chemical shifts as input for either structure determination or comparison of experimental chemical shifts to calculated chemical shifts of a known structure. However, if using a CCPN project as input to the 'Calculate chemical shifts' functionality CASPER will take existing assignments into account (meaning that the structure also needs to be specified in the CCPN project). This means that it is possible not only to verify assignments, but also to assign some signals and let CASPER assign the more difficult ones.

6.2.1. Input data

In this example the *E. coli* O101 O-antigen is used again. To illustrate the power of this approach the project from section 6.1.3. was used, but the anomeric assignments were switched and the α -D-GlcNAc H3,C3 and H4,C4 assignments were removed. This means that all crosspeaks except two are assigned and two of the assignments are wrong. The project was archived as explained in section 6.1.2.

6.2.2. Results

In the CASPER Calculate chemical shifts interface the project archive was specified in the Project field and the start button was pushed. Viewing the results (Fig. 11) it can be seen that the missing assignments are now fixed by CASPER and it is quite obvious that the two anomeric assignments should be switched.

Experimental	Calculated	Expt-Calc	Assignment
98.03 - 5.01	99.89 - 4.95	0.38	α -D-GlcNAc ⁱⁱ - 1
99.20 - 4.94	97.86 - 5.04	0.29	α -D-GalNAc ⁱ - 1
78.25 - 4.05	79.03 - 4.16	0.19	α -D-GalNAc ⁱ - 4
72.43 - 4.04	72.14 - 4.09	0.07	α -D-GalNAc ⁱ - 5
71.33 - 3.88	71.84 - 3.77	0.15	α -D-GlcNAc ⁱⁱ - 3
72.01 - 4.30	71.70 - 4.25	0.08	α -D-GlcNAc ⁱⁱ - 5
70.63 - 3.64	71.05 - 3.62	0.09	α -D-GlcNAc ⁱⁱ - 4
68.01 - 4.07	68.40 - 4.05	0.08	α -D-GalNAc ⁱ - 3
66.06 - 4.03	66.61 - 4.01	0.11	α -D-GlcNAc ⁱⁱ - 6
66.06 - 3.63	66.61 - 3.71	0.14	α -D-GlcNAc ⁱⁱ - 6'
61.09 - 3.72	61.23 - 3.92	0.20	α -D-GalNAc ⁱ - 6
61.09 - 3.72	61.23 - 3.91	0.19	α -D-GalNAc ⁱ - 6'
54.92 - 3.93	55.18 - 3.93	0.05	α -D-GlcNAc ⁱⁱ - 2
50.81 - 4.28	51.37 - 4.26	0.11	α -D-GalNAc ⁱ - 2
22.85 - 2.07	22.91 - 2.06	0.02	α -D-GalNAc ⁱ - Me_2
22.68 - 2.10	22.87 - 2.06	0.05	α -D-GlcNAc ⁱⁱ - Me_2

Fig. 11: The assignment table shows that two resonances are not correctly assigned.

Finally, the CCPN project was saved and the needed corrections were made in CCPN Analysis to obtain the correct assignments.

7. Tutorials

7.1. Calculating ^{13}C and ^1H chemical shifts of the *Escherichia coli* O22 O-antigen polysaccharide

You have a sample of a bacterial polysaccharide that you think is the *Escherichia coli* O22 O-antigen and you have the chemical shifts from ^{13}C NMR. Now you shall check if your experimental spectrum agrees with the simulated spectrum of that structure or if your structure assumption might be wrong. The structure of the polysaccharide looks like this:



The experimental ^{13}C chemical shifts (at 80 °C) are:

99.3 73.0 74.1 70.5 72.0 69.1
 104.4 74.7 77.4 78.0 77.6 174.3
 103.9 52.4 74.2 74.3 75.5 61.9
 96.6 68.3 80.9 70.2 72.2 62.3
 102.9 52.1 76.8 65.4 76.0 62.1

Procedure:

1. Set the Title to "Escherichia coli O22 O-antigen"
2. Specify the structure. Keep in mind that:
 - Linkage is the position to which the anomeric bond links to the residue at the Reducing end.
 - There are two residues linking to the $\beta\text{-D-GlcpNAc}$.
3. Enter the experimental chemical shifts in the ^{13}C -Chemical shifts text field.
4. Add the OAc substituent.
5. Press the 'Start simulation' button and then wait a few seconds.
6. View the results. Does it seem reasonable that this is the suggested structure? Keep in mind that the experiments are performed at slightly higher temperature than the experiments in the CASPER database.

Hint: The structure can be specified as

b D-GalpNAc (->6) residue 5
 a D-Galp (->3) residue 1
 a D-GalpNAc (->3) residue 2
 a D-GlcpA (->4) residue 3
 a D-Glcp (->4) residue 4

To add the substituent click "Show/hide advanced options" and in the menu that appears and then select
 O-Ac (->3) residue 3

7.2. Sequence determination using insufficient linkage data

This exercise demonstrates the possibilities to determine structures even if you do not know everything about the residues. This means CASPER will have to build, and calculate the chemical shifts of, *very many* structures so it takes a little longer.

You want to analyse the contents of a sample, but the results from methylation analysis have not been very accurate. You have data from HSQC and HMBC experiments.

The sugar and methylation analyses show:

1 D-GalpNAc with possible linkage positions 1, 2, 3, 4 and 6 (not necessarily all of them).

1 D-GlcpNAc with possible linkage positions 1, 2, 3, 4 and 6 (not necessarily all of them).

Results from NMR experiments:

HSQC:

99.227 - 4.947 98.049 - 5.006 78.256 - 4.057 72.453 - 4.042 71.982 - 4.291

71.337 - 3.873 70.703 - 3.635 68.062 - 4.048 66.179 - 4.029 66.179 - 3.640

61.137 - 3.715 61.137 - 3.715 54.929 - 3.942 50.851 - 4.281 22.866 - 2.072

22.695 - 2.098

HMBC:

99.227 - 4.057 99.227 - 3.942 98.049 - 4.029

78.256 - 4.947 72.453 - 5.006 71.982 - 4.947

71.337 - 4.947 68.062 - 5.006 66.179 - 5.006

1. Specify the residues and the possible linkage positions (to indicate that the linkage positions are not certain click the star column for that residue – otherwise all of the linkage positions must be used).
2. Enter the experimental NMR data.
3. Start the simulation and wait until it is finished – this can take a few seconds.
4. View the results list. Is there any structure that gets a better score than the other structures?
5. Consider the top five carbohydrate structures. Are they similar to each other?
6. View the assignments of the top ranked structure by clicking on it (assignments open up in a new tab in the browser window). Are the chemical shift assignments convincing?
7. Click the 'Back' button in your web browser to return to the results table.
8. Study the assignments of the second best structure to see what those look like. Try to find where the largest errors are. If you compare this structure with the first one, can you draw any conclusions regarding which structure elements cause most errors in the second assignments?

Appendix A – CASPER structure representation

Carbohydrate residue are represented by anomeric configuration (a/b), absolute configuration (D/L), three letter residue abbreviation, if the ring is furanose 'f' is appended:

aDGlc
bLGalf

2-NAc and 1-OMe are appended at the end without position specifier:

aDGlcNAcOMe

Linkages are described like this:

aDGal(1->4)aDGlcOMe

A label can be (but does not have to be) placed next to a residue. These are placed in square brackets (unfortunately the same as branches):

aDGal[b](1->4)aDGlcOMe[a]

in the CASPER web interface these labels are shown as lower case Roman numbers in superscript:

aDGalⁱⁱⁱ(1->4)aDGlcOMeⁱ

Branches are also represented by square brackets:

aDGal(1->4)aDGlc(1->2)[aLRha(1->4)]aDGlcOMe

If there is more than one branch to a residue they are sorted according to substitution position:

aDGal(1->4)aDGlc(1->2)[aLRha(1->4)][aDGlc(1->3)]aDGlcOMe

Repeating units are represented by "open" linkages at either end:

->4)aDGlc(1->6)aDGal(1->

or with a branch:

->4)[aDGlc(1->6)]aDGal(1->

Substituents can be added. They are added in alphabetical order with substitution positions and multiplicity specified. They are added after e.g. NAc and OMe, which are considered part of the residue name by CASPER.

aDGlc2Ac
aDGlc2,4diAc
aDGlc2,4diAc3Me

Some example structures:

A heptasaccharide:

bDGal(1->4)bDGlcNAc(1->2)aDMan(1->3)[bDGal(1->4)bDGlcNAc(1->2)aDMan(1->6)]aDMan

O128 O-antigen PS:

->6)[aLFuc(1->2)]bDGal(1->3)bDGalNAc(1->4)aDGal(1->3)bDGalNAc(1->

Shigella flexneri X:

->2)[aDGlc(1->3)]aLRha(1->2)aLRha(1->3)aLRha(1->3)bDGlcNAc(1->

Methyl beta-cellobioside:

bDGlc(1->4)bDGlcOMe

Xylose containing N glycan:

bDXyl(1->2)[aDMan(1->3)][aDMan(1->6)]bDMan(1->4)bDGlcNAc(1->4)bDGlcNAc

Repeating unit of E. coli O128 O-antigen:

->6)[aLFuc(1->2)]bDGal(1->3)bDGalNAc(1->4)aDGal(1->3)bDGalNAc(1->

Repeating unit of E. coli O105 O-antigen:

->4)aDGlcA2,3diAc(1->2)[bDRibf(1->3)]aLRha4Ac(1->3)bLRha(1->4)bLRha(1->3)bDGlcNAc6Ac(1->