Available online at www.sciencedirect.com
SCIENCE DIRECT
Carbohydrate Research 341 (2006) 1003–1010
Carbohydrate RESEARCH

Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program CASPER

Per-Erik Jansson,[a] Roland Stenutz[b,*] and Göran Widmalm[b]

[a]*Centre for Medical Innovations, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[b]*Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden*

Received 15 December 2005; accepted 27 February 2006
Available online 27 March 2006

Abstract—A WWW-interface to a program for structure elucidation of oligo- and polysaccharides using NMR data, CASPER, is presented. The interface and the underlying program have been extensively tested using published data and it was able to simulate ^{13}C NMR spectra of >200 structures with an average error of about 0.3 ppm/resonance. When applied to the repeating units of *Escherichia coli* O-antigens the published structures were found among the five highest ranked structures in 75% of the cases. The average deviation between calculated and experimental ^{13}C chemical shifts was 0.45 ppm. Oligosaccharide spectra were calculated with even better accuracy (0.23 ppm/resonance) and the correct structure was ranked 1st or 2nd in all the cases examined. Additional NMR experiments that may be required to distinguish between candidate structures are aided by the assignments provided by the program. This computational approach is also suitable for use in structural confirmation of chemically or enzymatically synthesized oligosaccharides. The program is found at http://www.casper.organ.su.se/casper.

Keywords: NMR prediction; Computer aided structure elucidation (CASE); Database; Glycomics

1. Introduction

Oligo- and polysaccharides are compounds of fundamental importance in biochemistry and medicine. Since the cell surface of most organisms display carbohydrates, they are important for processes that involve cell recognition such as adhesion and immunological reactions.[1] They also serve as receptors for viruses and bacteria as well as bacterial toxins. Unlike nucleic acids and proteins their structures cannot be predicted from analysis of the genome nor is it easy to sequence them by chemical methods. Lack of structural information is one of the reasons that the role of carbohydrates in biological systems has not been fully appreciated until recently. The methods currently employed for determining the structure of complex carbohydrates depend largely on the level of details desired and the amount of materials available. Mass spectrometry is often used in the study of glycoproteins, where the amount of material is limited and the number of possible structures can be restricted by biosynthetic considerations. In the case of polysaccharides of bacterial origin the amount of available material is often considerably larger and both chemical analysis and ^{1}H and ^{13}C NMR spectroscopy are routinely performed. This allows all the details of the structure, such as ring size, the configuration of the glycosyl residues, the linkage positions, and possible substituents to be determined. However, this is a relatively time demanding process in particular when it is considered that about 15% of the structures of repeating units of lipopolysaccharides are found to be identical to those from other bacterial genera.[2] Only in rare cases is it possible to recognize that a structure is already known prior to a complete structure determination. Since in principle all structures give unique 1D NMR

* Corresponding author. Tel.: +46 8 6747263; fax: +46 8 15 49 08; e-mail: roland@organ.su.se

0008-6215/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2006.02.034

spectra several attempts have been made to extract structural data from 1D spectra. The methods can be divided into (i) databases, (ii) artificial neural networks, and (iii) computer programs for structure generation/spectrum calculation.

One of the database methods, the structural reporter group approach, has been successfully applied in the structure determination of complex oligosaccharides from glycoproteins.[3] The reporter groups used are well resolved ^{1}H NMR resonances such as the signals of anomeric protons and of methyl groups. Since the number of well resolved resonances, as well as the dispersion, in ^{1}H NMR spectra is limited it is necessary to record the spectra under highly standardized conditions. Since no chemical information is necessary, this method is suitable for the analysis of small amounts of material. In principle, ^{13}C NMR chemical shifts can be used but they are less sensitive to structural changes. A substantial database of known structures and substructures is necessary and the approach is generally not applicable to novel oligosaccharides or bacterial polysaccharides for which the variation in monosaccharide components and linkage types is large.

Artificial neural networks have been used to interpret the NMR spectra of xyloglycans[4] and of glucuronoxylomannans.[5] They have the advantage that spectra of compounds with irregular structure or in the presence of contaminants can be interpreted. A training set of known structures is required and it is difficult to know how far the results can be extrapolated and therefore they have not found wider use.

The calculation of NMR spectra can be used for structure determination if it is combined with a suitable mechanism for generating trial structures. The program CASPER, Computer Assisted SPectrum Evaluation of Regular polysaccharides, belongs to this category.[6] All structures consistent with data from chemical analyses are generated, their NMR spectra calculated and finally ranked according to their agreement with the experimental spectrum. Chemical information is used in this approach to limit the number of possible structures. The computation of the chemical shifts is based on schemes with substituent induced chemical shifts, referred to as glycosylation shifts below, and is in favorable cases accurate to about 0.1 ppm in ^{13}C NMR. Additional NMR experiments, which may be required to distinguish between the suggested structures, are aided by the tentative assignments that are the results of this procedure. Until recently this approach required a complete NMR spectrum, that is, all chemical shifts had to be determined (but not assigned), which is difficult in ^{1}H NMR spectroscopy because of spectral overlap. Although this problem has been overcome,[7] ^{13}C is still the preferred nucleus because of the greater dispersion of the resonances and the better predictability of the chemical shifts. This type of approach has been used to elucidate the structure of several bacterial polysaccharides using CASPER[8] and similar programs.[9,10]

2. Results: the CASPER program

The current version of CASPER is a C-program that runs on a Linux-system. The user interface to the CASPER program is a simple command line interface with support for scripts. Basic flow control in the form of conditional execution is provided. From the command line all of CASPER's features are accessible but most changes in the program also affect the interface. To make CASPER more user-friendly we decided to add a graphical interface. Using HTML-forms and the common gateway interface (CGI) seemed particularly attractive since CASPER would be available globally and instantaneously without the need for installation and maintenance at more than one site. The WWW-interface to the program can be accessed using the following link: http://www.casper.organ.su.se/casper.

2.1. Calculation of NMR spectra

The NMR spectra of oligo- and polysaccharides can be approximated from the chemical shifts of the constituent monosaccharides and the glycosylation shifts that are caused by substitution. The size of the glycosylation shifts depends on the stereochemistry at and near the glycosidic bond.[11,12] This allows the same glycosylation shifts to be used for the linkage between several different monosaccharides that have similar stereochemistry at the glycosidic linkage. Provided that only short range interactions are present, additivity of glycosylation shifts is observed. Notable exceptions are observed for residues with vicinal substitution, for example, for branch points as well as for the substituting sugars. The vicinal substitution causes steric interactions, not present in the corresponding disaccharide elements. To some extent this can be compensated by the inclusion of corrections based on data from similar trisaccharides. Preferably, the NMR data should be recorded under the same standardized conditions used to create the database (see below) but this may not always be possible.

2.2. Databases

The databases in CASPER contain a range of information about the structure of the residues. Atom labels, connectivities in the spin systems and information about allowed linkage positions are on the lowest level in a hierarchy of data structures. This information is used for displaying the chemical shifts, restricting the possible connections between residues and for the calculation of 2D NMR spectra (^{13}C,^{1}H- or ^{1}H,^{1}H-correlated spectroscopy). Translation tables allow data for one type

of residue to be used for another as discussed above. This enables CASPER to use glycosylation shifts from, for example, hexopyranoses to be used for 2-acetamido-2-deoxy-hexopyranoses. The next level of data structures contains the chemical shifts for the unsubstituted glycosyl residues together with links to relevant glycosylation shifts. Information about the configuration of the atoms at the linkages, which is needed to find suitable corrections for steric crowding (see below), is also stored here. Since a complete set of glycosylation shifts would be very large there is also a list of residues that are used to approximate missing data. Finally, there are glycosylation shifts obtained from disaccharide fragments and sets of chemical shift corrections from sterically crowded trisaccharides.

It is not practical to obtain experimental values for all glycosylation shifts because of the large number that would be required. This is true even more so for the steric corrections. However, the configurational arrangement at and near to the glycosidic linkage is the most important factor and data can be transferred from one case to another. CASPER therefore contains a scheme of default values that will be used in case the databases do not contain an exact match for the linkage sought. In that case a search will be made for the enantiomer of the fragment. If this is not found, the first residue will be replaced by a similar one and finally both will be varied. The steric corrections are selected so that the substitution positions, anomeric and absolute configurations of the three residues match.

In order to facilitate maintenance of the databases and the glycosylation shifts, steric corrections can be extracted automatically utilizing chemical shifts of di- or trisaccharides. Storing chemical shifts rather than the glycosylation shifts in the external databases may seem like a trivial difference but it reduces the likelihood of errors significantly, since the literature values can be entered unaltered. It also allows changes, that is, resulting from changing the chemical shifts of a monosaccharide residue to be passed on to glycosylation shifts and steric corrections automatically. In total, the databases contain about 150 monosaccharides (counting both anomers), 150 sets of glycosylation shifts, and approximately 70 sets of sterical corrections. The web-interface currently allows 50 of the residues most common in bacterial polysaccharides to be used (Table 1). About half of the data contained in the databases originate from our laboratory and were recorded using standardized conditions (D_2O, pD ≈ 9, 70 °C, internal dioxane $\delta_C = 67.40$ or TSP $\delta_H = 0.00$). The remaining data were taken from the literature and contain an estimate of its accuracy (expected per resonance error in ^{13}C), which if possible, was obtained by comparison with a CASPER simulation of the fragment. In most cases, the reference was also adjusted.

2.3. Interfaces

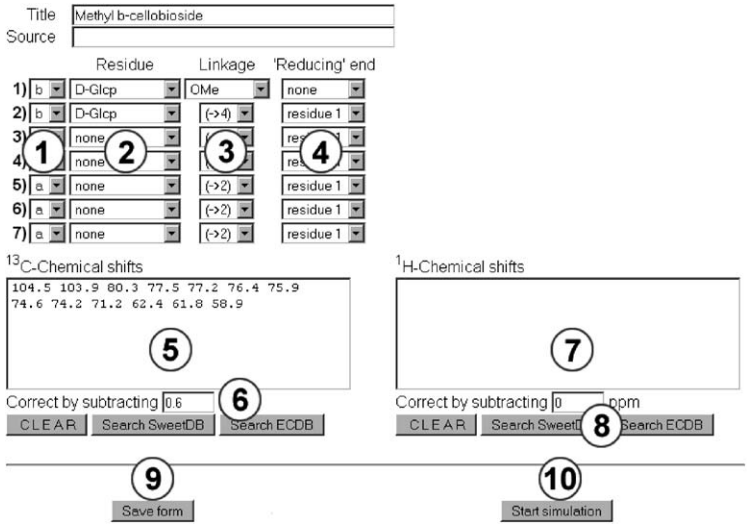
The user can access the program using an interface based on HTML-form. Simple JavaScript procedures make a preliminary check of the entered data. The forms are designed to reduce the possible input errors such as typographic errors and invalid choices. On the server the data are checked again, mainly to ensure that all parameters are in the correct format, before an input script for CASPER is prepared. This verification is essential for security reasons since it is easy to bypass client-side controls. CASPER may still fail during a simulation, for example, when the time required for completion exceeds 3 min (time out with the present server settings), but the number of errors is greatly reduced. Since the content of a form is not normally saved by the browser, a special function is also available to convert the entered data into an HTML-form with appropriate default values for all fields. This filled form can then be saved locally without loss of data. The interface greatly simplifies the use of CASPER but in order to keep it simple only some of the program options are presently available.

2.3.1. Chemical shift calculation. When the structure of an oligo- or polysaccharide is known, CASPER can be used to simulate the complete set of ^{13}C and ^{1}H-chemical shifts for this compound. The glycan structure is constructed by selecting up to seven monosaccharide residues, defining their anomeric configurations, and the connections between them. For the first residue it is also possible to construct a methyl glycoside by linking to a methyl group. When ^{13}C or ^{1}H-chemical shifts are entered CASPER will assign them. The chemical shifts can also be used to search SugaBase[13] at the SweetDB[14] site or a local database of *Escherichia coli* O-antigens.[2] Scheme 1 shows how the structure and ^{13}C chemical shifts for methyl cellobioside[15] are entered.

The JavaScript procedures in the input form prevent the selection of invalid combinations of residues and linkages, such as linking to the positions in deoxysugars that lack hydroxyl groups, and prevent non-numeric data to be entered as chemical shifts. If the input data

Table 1. Sugar residues supported by the WWW-interface (all residues are in their pyranoid form)

Hexoses	D-Gal*p*, D-Glc*p*, D-Man*p*
6-Deoxyhexoses	D-Fuc*p*, L-Fuc*p*, D-Rha*p*, L-Rha*p*, D-Qui*p*
3,6-Dideoxyhexoses	Abe*p*, Col*p*, Par*p*, Tyv*p*, Asc*p*
2-Acetamido-2-deoxyhexoses	L-Fuc*p*NAc, D-Gal*p*NAc, D-Glc*p*NAc, D-Man*p*NAc, Mur*p*NAc, L-Rha*p*NAc, D-Qui*p*NAc
Uronates	D-Gal*p*A, D-Gal*p*ANAc, D-Glc*p*A, D-Man*p*A, D-Man*p*ANAc



Scheme 1. Spectrum simulation: (1) anomeric configuration of the residue; (2) glycosyl residue; (3) position to link to (or OMe for methyl glycosides); (4) residue to link to; (5) ^{13}C chemical shifts to be assigned (optional); (6) correction to apply to chemical shifts; (7) ^{1}H-chemical shifts to be assigned (optional); (8) use chemical shifts to search SweetDB or a local database of *E. coli* O-antigens; (9) save filled-in form; (10) start simulation and assignment.

pass these first tests they are submitted to the server. There the values are extracted, checked for consistency, and inserted into a script, which is then submitted to CASPER. The program calculates the chemical shifts for the structure, which is then filtered through a script that adds HTML formatting before returning the result to the client. The results of a spectrum simulation are shown in Scheme 2.

Simulated structure

β-D-Glc^{ii}-(1→4)-β-D-GlcOMe^{i}

→4)-β-D-GlcOMe^{i}	103.96	73.77	75.32	79.88	75.64	61.26	57.91	
	4.39	3.33	3.63	3.61	3.60	3.82	3.99	3.58
β-D-Glc^{ii}-(1→	103.36	74.14	76.60	70.53	76.90	61.67		
	4.52	3.34	3.53	3.43	3.51	3.74	3.92	

1

Assignment of ^{13}C resonances

Experimental	Simulated	Exp-Sim	Assignment
103.90	103.96	-0.06	β-D-GlcOMe^{i} - 1
103.30	103.36	-0.06	β-D-Glc^{ii} - 1
79.70	79.88	-0.18	β-D-GlcOMe^{i} - 4
76.90	76.90	0.00	β-D-Glc^{ii} - 5
76.60	76.60	0.00	β-D-Glc^{ii} - 3
75.80	75.64	0.16	β-D-GlcOMe^{i} - 5
75.30	75.32	-0.02	β-D-GlcOMe^{i} - 3
74.00	74.14	-0.14	β-D-Glc^{ii} - 2
73.60	73.77	-0.17	β-D-GlcOMe^{i} - 2
70.60	70.53	0.07	β-D-Glc^{ii} - 4
61.80	61.67	0.13	β-D-Glc^{ii} - 6
61.20	61.26	-0.06	β-D-GlcOMe^{i} - 6
58.30	57.91	0.39	β-D-GlcOMe^{i} - OMe

2

Error=1.44 ppm (0.11/shift), Systematic error=0.00 ppm, RMS error=0.15 ppm 3

Experimental structure

β-D-Glc^{ii}-(1→4)-β-D-GlcOMe^{i}

→4)-β-D-GlcOMe^{i}	103.90	73.60	75.30	79.70	75.80	61.20	58.30	
	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
β-D-Glc^{ii}-(1→	103.30	74.00	76.60	70.60	76.90	61.80		
	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	

4

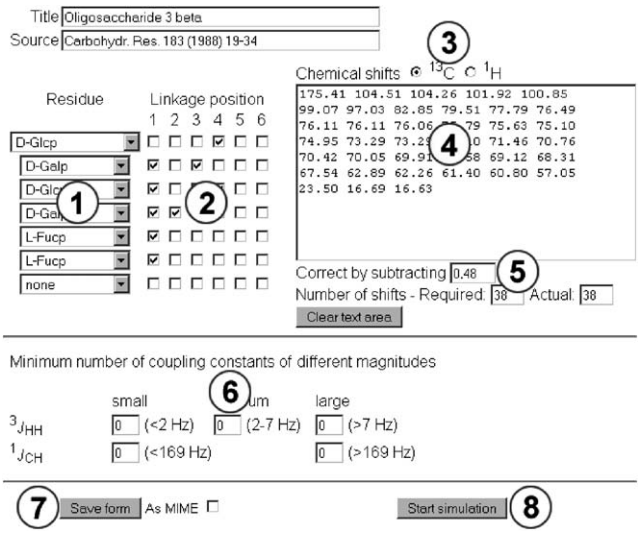
Scheme 2. Results of simulation: (1) simulated NMR data; (2) assignment of experimental chemical shifts in chemical shift order; (3) difference in chemical shifts between experimental and simulated spectra; (4) assigned experimental data by residue/atom (n.d. = not determined).

2.3.2. Sequence determination. A sequence determination of an oligo- or polysaccharide using CASPER requires information about the residues and their linkages. The former is obtained from sugar analysis and absolute configuration determination, and the latter, from methylation analysis. All structures consistent with this information are generated. The ^{1}H and ^{13}C NMR spectra for each structure are calculated, compared with the experimental data, and ranked according to similarity.

A comparison between the simulated and experimental spectrum can be performed in two different ways. The original and simplest method is a simple line-by-line comparison of the sorted spectra. The sum of the absolute values of the differences is then used for ranking. For this to succeed the spectra have to contain the same number of resonances. If the spectra differ in the number of resonances then, the best (lowest error) assignment is found but this procedure is slower, in particular for large structures. Root-mean-squared errors are also calculated for comparison with other programs[9] but they seem to be linearly related to and approximately 50% larger than the sum of errors.

The anomeric $^{3}J_{H,H}$ or $^{1}J_{C,H}$ values can be used to reduce the number of generated structures. Only the unambiguously classified coupling constants are used as lower limits for the number of residues of a given type. This allows the use of $^{3}J_{H,H}$ or $^{1}J_{C,H}$ values even when some are unresolved in the spectrum or fall between two ranges. Structures violating these restrictions are not generated. In the case of structures containing many residues the use of these constraints may significantly reduce the time required for the calculations. However, the chemical shifts are normally sufficient to determine the anomeric configurations.

A typical input is shown in Scheme 3. CASPER returns a list of the 10 structures with the lowest calculated deviation from the experimental spectrum. From this list structures can be selected for closer examination. The results are then presented as for the simulations above (Scheme 2).



Scheme 3. Sequence determination: (1) glycosyl residues; (2) linked positions; (3) selection of experimental ^{1}H or ^{13}C NMR data; (4) experimental NMR data; (5) adjustment of chemical shifts; (6) coupling constants to be used as restraints (optional); (7) save filled-in form; (8) start sequence determination.

3. Discussion

Since the first version of CASPER, published in 1987, it has been used as an aid in structure elucidation in our laboratory. However, the interface has been a weakness, and because the program was run locally, it would also have been necessary for anyone wanting to run a CASPER calculation to install the program. The new interface has made it possible for anyone anywhere to use the program without lengthy preparations. Designing a simple interface was more complicated than anticipated and there is no doubt room for further improvement. The current version is, however, sufficiently advanced to be useful in a number of different applications.

The new interface and the underlying program have been tested extensively on three different datasets. Two kinds of problems related to the current implementation were noted. About half of the *E. coli* O-antigens contained residues not supported by the web-interface although they are supported by the program. A doubling of the number of glycosyl residues would be needed in order to cover them all. Incomplete experimental spectra, especially of larger structures, often caused time-out errors. Using coupling constants as restraints decreases the time required and thus reduces the problem. Simulations that require more than a few minutes to execute are probably not suited for an interactive approach.

The results of the four sets of calculations are summarized in Table 2. In about 50% of the examined *E. coli* O-antigens (dataset I) the correct structure is ranked first or second by ^{13}C and 75% are ranked no worse than 5th place (Fig. 1). However, most of the remaining structures are not among the top 10. The majority of the latter structures contain oligosaccharide elements with steric interactions but there are also cases where there presumably are typographical errors in the reported chemical shifts or even uncertainty regarding the reported structure. The ^{1}H NMR data gave significantly less reliable results. In part this is due to the smaller dispersion of chemical shifts and the limited number of readily identifiable ^{1}H-chemical shifts in a standard 1D spectrum. Additional ^{1}H-chemical shifts must then be obtained from 2D experiments. There is also the problem of assigning resonances of the prochiral protons, for example, H-6_{proR} and H-6_{proS} in hexoses, and

Table 2. Summary of the results for the different datasets

Dataset	Topology	Number of structures	Average error ($\Delta\delta$, ppm)	Average rank
I, ^{13}C NMR	Linear	14	0.36	2.1
E. coli O-antigens	1 Branch	24	0.65	2.7[a]
	2 Branches	6	0.53	—[b]
	All	**44**	**0.54**	
I, ^{1}H NMR	Linear	9	0.05	3.7[c]
	1 Branch	20	0.06	4.3[d]
	2 Branches	3	0.06	—[e]
	All	**32**	**0.06**	
II	Linear	7	0.14	1.0
Oligosaccharides	1 Branch	12	0.28	1.6
	All	**19**	**0.23**	
III	Linear	144	0.32	
Bacterial PS	1 Branch	81	0.34	
	2 Branches	13	0.39	
	All	**238**	**0.33**	

[a] Average for 13 ranked structures (4 ranked >10th and 7 timed out).
[b] 2 Ranked >10th and 4 timed out.
[c] Average for 6 ranked structures (2 ranked >10th and 1 timed out).
[d] Average for 3 ranked structures (14 ranked >10th and 6 timed out).
[e] All three determinations had time-out errors.

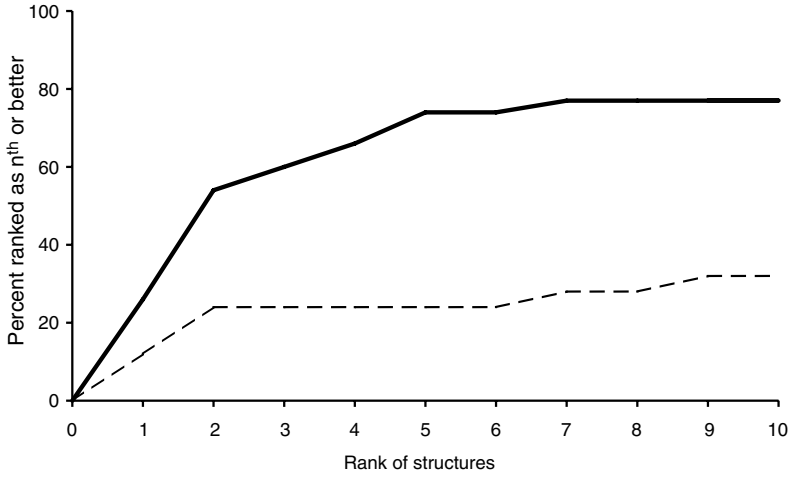


Figure 1. Ranking of *E. coli* O-antigen structures from CASPER simulations based on ^{13}C NMR data (solid line) and ^{1}H NMR data (broken line).

the greater sensitivity of ^{1}H-chemical shifts to experimental conditions. Only 25% of the structures were ranked first or second using ^{1}H-chemical shifts and only 30% were among the top 10 (Fig. 1). The calculation of the chemical shifts of oligosaccharides from mammalian sources (dataset II) gave very good results. All structures were ranked third or better. Presumably this is because the coverage of the component mono-, di-, and trisaccharides in the databases is better than for bacterial polysaccharides that contain many unusual residues. Dataset III, obtained mainly from literature data for bacterial polysaccharides, was also simulated with an average error of 0.3 ppm/signal. As in the cases above there was a slight increase in average error going from linear to branched structures.

The web-interface has made the use of CASPER much easier although some features of the program have not been fully implemented. A more flexible interface that gives access to additional functions is desirable as well as interfacing with other databases. The latter problem is currently being addressed by a newly formed consortium for the integration of European carbohydrate databases,[21] which will develop data standards for carbohydrate structures as well as experimental data. As part of this effort, a link from the chemical shift estimation in glycoSCIENCES.DE[22] to CASPER has been implemented. This interface also allows a wider range of residues and larger structures to be simulated. Currently the number of residues in the oligosaccharide or repeating unit to be simulated is limited by the interface

to seven. This allows the spectra of almost all bacterial polysaccharides to be calculated but is too restrictive for many N-linked glycans. The reason for this threshold is that larger structures often will exceed the time limit of the server (≈3 min) during sequence determinations, especially when spectral data is incomplete.

The likelihood of a correct structure being proposed depends in a complex way on the number of possible permutations of the components, the accuracy of the chemical shift calculation and the *difference in the sum of errors* between proposed structures. Only in a few cases is it possible to say with certainty that the proposed structure is the only reasonable one. New structures should therefore always be supported by additional experimental data. In practice, selected NMR experiments may be used to distinguish between highly ranked structure suggestions. It is also possible to select the correct structure using additional experimental data, such as specific degradations or biosynthetic considerations. The ability of CASPER to distinguish between several candidate structures could be further improved by including the possibility to analyze 2D NMR spectra or using a combination of ^{1}H- and ^{13}C-data.

Considering that the program was originally developed for the structural elucidation of bacterial polysaccharides it is particularly satisfying that it performs even better for oligosaccharides of mammalian origin. CASPER analysis results in a highly probable structure and as such it forms the basis of a rapid approach to a final structure using a limited amount of additional information or just a candidate that can be used in biochemical, genetics, or bioinformatics investigations. We anticipate that increased use of the program will lead to feedback from users that we can use for future improvements.

4. Experimental

Polysaccharide structures and NMR data were taken from the literature. Information about linkage positions was taken from the published structures and all chemical shifts were adjusted to a common reference (1,4-dioxane in D_2O $\delta_C = 67.40$ and TSP $\delta_H = 0.00$). Four sets of calculations were performed with three different sets of experimental spectra.

Dataset I was taken from a database of repeating units of *E. coli* O-antigens.[2] The database contained a total of 89 structures of which 41 contained residues not supported by the web-interface. Two sets of calculations were performed with this dataset using either ^{13}C or ^{1}H NMR data. Of the remaining structures three had no published ^{13}C NMR data. Calculations with incomplete NMR data take longer time than with complete spectra and therefore two structures had time-out errors during the assignment of their chemical shifts. Nine had time-out errors during sequence determination, leaving 44 assigned spectra and 35 sequence determinations using ^{13}C NMR data. The calculations above were repeated using ^{1}H NMR data. Of the 48 structures that could be simulated 12 had no published ^{1}H NMR data. Four structures caused time-out errors during spectrum simulation and nine during sequence determination leaving 32 and 27 spectra for analysis, respectively.

Dataset II consisted of the ^{13}C chemical shifts of 19 oligosaccharides related to milk,[16,17] complex-type,[18] and high-mannose glycans[19] and related compounds[20] were taken from the literature.

Dataset III contains 238 structures from various literature sources that have been used to test the performance of CASPER during database development. For this dataset only the deviation between the calculated ^{13}C NMR spectra and the published spectra was calculated.

Acknowledgments

This work was supported by a grant from the European Union (EUROCarbDB, RIDS Contract number 011952, to R.S.) and in part by a grant from the Swedish Research Council (to G.W.).

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Stenutz, R.; Weintraub, A.; Widmalm, G. *FEMS Microbiol. Rev.*, in press. http://www.casper.organ.su.se/ECODAB.
3. Vliegenthart, J. F. G.; van Halbeek, H.; Dorland, L. *Pure Appl. Chem.* **1981**, *53*, 45–77.
4. Meyer, B.; Hansen, T.; Nute, D.; Albersheim, P.; Darvill, A.; York, W.; Sellers, J. *Science (Washington, DC)* **1991**, *251*, 542–544.
5. Cherniak, R.; Valafar, H.; Morris, L. C.; Valafar, F. *Clin. Diagn. Lab. Immunol.* **1988**, *5*, 146–159.
6. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1987**, *168*, 67–77.
7. Stenutz, R.; Jansson, P.-E.; Widmalm, G. *Carbohydr. Res.* **1998**, *306*, 11–17.
8. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–181.
9. Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. *Carbohydr. Res.* **1988**, *175*, 59–75.
10. Toukach, F. V.; Shashkov, A. S. *Carbohydr. Res.* **2001**, *335*, 101–114.
11. Kochetkov, N. K.; Chizhov, O. S.; Shashkov, A. S. *Carbohydr. Res.* **1984**, *133*, 173–185.
12. Söderman, P.; Jansson, P.-E.; Widmalm, G. *J. Chem. Soc., Perkin Trans. 2* **1998**, 639–648.
13. van Kuik, J. A.; Hård, K.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1992**, *235*, 53–68.
14. Loß, A.; Bunsmann, P.; Bohne, A.; Loß, A.; Schwarzer, E.; Lang, E.; von der Lieth, C. W. *Nucleic Acids Res.* **2002**, *30*, 405–408.

15. Usui, T.; Yamaoka, N.; Matsuda, K.; Tuzimura, K.; Sugiyama, H.; Seto, S. *J. Chem. Soc., Perkin Trans. 1* **1973**, 2425–2432.
16. Breg, J.; Romijn, D.; Vligenthart, J. F. G.; Strecker, G.; Montreuil, J. *Carbohydr. Res.* **1988**, *183*, 19–34.
17. Bradbury, J. H.; Collins, J. G.; Jenkins, G. A.; Trifonoff, E.; Messer, M. *Carbohydr. Res.* **1983**, *122*, 327–331.
18. Bock, K.; Arnarp, J.; Lönngren, J. *Eur. J. Biochem.* **1982**, *129*, 171–178.
19. Allerhand, A.; Berman, E. *J. Am. Chem. Soc.* **1984**, *106*, 2400–2412.
20. Maaheimo, H.; Räbinä, J.; Renkonen, O. *Carbohydr. Res.* **1997**, *297*, 145–151.
21. European Carbohydrate DataBases (EuroCarbDB). URL http://www.eurocarbdb.org/, accessed December 12th 2005.
22. glycoSCIENCES.DE. URL http://www.glycosciences.de/sweetdb/, accessed December 12th 2005.