Pure & Appl. Chem., Vol. 61, No. 7, pp. 1181–1192 1989.
Printed in Great Britain.

Computer-assisted structural analysis of regular polysaccharides

Per-Erik Jansson, Lennart Kenne and Göran Widmalm

Department of Organic Chemistry, Arrhenius Laboratory, University of Stockholm, S-106 91 Stockholm, Sweden

Abstract - A computer program, CASPER, for structural analysis of oligo- and polysaccharides has been developed. Simple chemical analyses and NMR data (^{1}H and/or ^{13}C NMR chemical shifts and $^1J_{C,H}$ and $^3J_{H,H}$ coupling constants) are used for determination of the structure.

INTRODUCTION

Carbohydrates play an important role in Nature as *e.g.* energy source, building material and in recognition processes. Due to the many different monosaccharides existing, the large number of combinations in which they can be linked, and their ability to form hydrophilic and hydrophobic regions, carbohydrates have many different properties, and specific interactions with other molecules in biological systems are common. To increase the understanding of their function in different surroundings and of the relationship between structure and biological activity, more information on structural and physical properties of carbohydrates is of fundamental importance. To obtain relevant structural information on a carbohydrate molecule a complete structural analysis with the following steps should be performed:

Analysis of components
Analysis of linkages
Analysis of sequence
Analysis of preferred conformations
Analysis of the dynamics of the molecule

In Nature several different types of carbohydrates exist, *e. g.* bacterial polysaccharides, glycolipids, glycopeptides and different kinds of glycosides. Examples of these are given in Fig. 1. The bacterial polysaccharides are composed of oligosaccharide repeating units, in which a large number of different sugars could occur, many of them rare. Glycolipids and glycopeptides consist of a limited number of different sugars often occuring in similar oligosaccharide elements. These structural elements can form larger oligosaccharides with a different numbers of branches and chain length. Saponins, an example among the glycosides, often contain two different oligosaccharide chains linked to a triterpene or a sterol residue. In the structural elucidation of these compounds both the structures of the two oligosaccharides and the positions to which they are linked have to be determined.

In the structural studies of carbohydrates a variety of methods have to be considered. The most used techniques today are

SUGAR ANALYSIS
METHYLATION ANALYSIS
CHEMICAL AND ENZYMIC DEGRADATIONS
MASS SPECTROMETRY
NMR SPECTROSCOPY

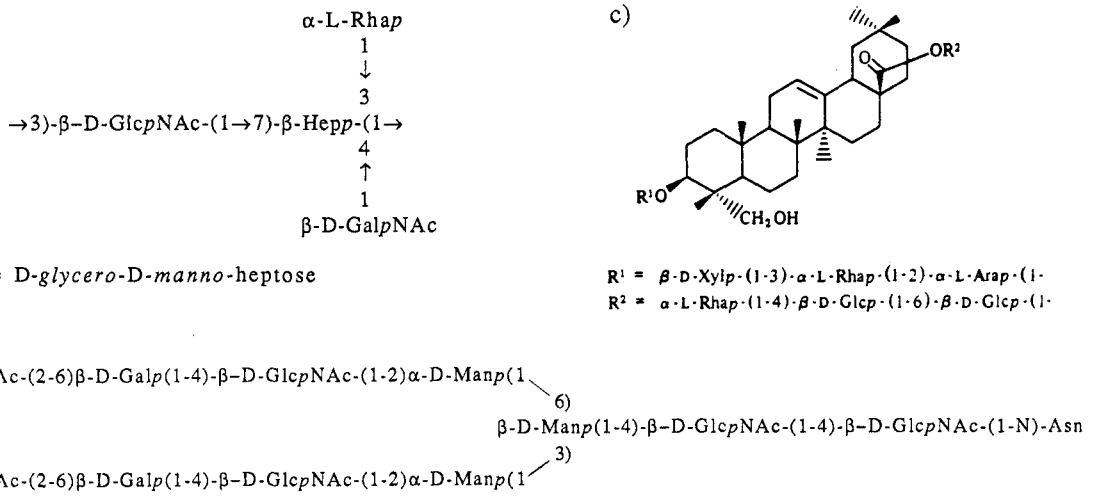


Fig. 1. Examples of different types of carbohydrates: a) O-Polysaccharide from *Vibrio cholerae* O:21; b) *N*-Acetyllactosamine type of carbohydrate chain from antithrombin III; c) Saponin from *Nigella sativa*.

The sugar and methylation analyses are still of fundamental importance for the determination of components and linkages. The anomeric configurations can be determined with both enzymic and chemical reactions, optical rotation and NMR spectroscopy. Determination of the sequence has been a difficult and time-consuming task and has mainly been performed by analysis of smaller oligosaccharides obtained by enzymic or specific chemical degradations (ref. 1). Still today it is a difficult problem and for each carbohydrate molecule different methods have to be used depending upon structure and components of the polysaccharide. Mass spectrometry is now one of the most important methods for sequence analysis of oligosaccharides. The advantage with this technique is the high sensitivity which makes analysis of small amounts of material possible but it is difficult to obtain information on anomeric configuration and linkage positions. Earlier only derivatized oligosaccharides could be analysed, inserted into the spectrometer either *via* a gas chromatograph or *via* a solid inlet interface. Recently soft-ionisation techniques, especially FAB-MS, has been introduced which enable sequence determination of underivatized oligosaccharides. FAB-MS has been used in several structural investigations of oligo- and polysaccharides (ref. 2).

During the last years also NMR spectroscopy has become of increasing importance as superconducting magnets, advanced computers, new multipulse experiments and two-dimensional techniques have become of general use. NMR spectroscopy is today used routinely for analysis of components, substituents, linkages and sequences and the structures of several oligo- and polysaccharides have been determined mainly or exclusively using one- or two-dimensional NMR techniques. Information from the chemical shift of structural reporter groups (ref. 3) and from nuclear Overhauser enhancements and scalar couplings over the glycosidic bond have been used to obtain sequence information. However, these techniques require expensive high-field NMR spectrometers and in most cases assignments of signals in the rather complex spectra, with several overlapping signals, are both difficult and time-consuming.

The efforts to develop new methods for future structural analysis of carbohydrates involve on one hand development of more sensitive techniques. These should enable analysis of a large number of glycolipids and glycopeptides that are involved in important biological processes. On the other hand, development of faster and simpler methods are desirable as those mentioned are time-consuming and sometimes also the experiments are difficult to perform.

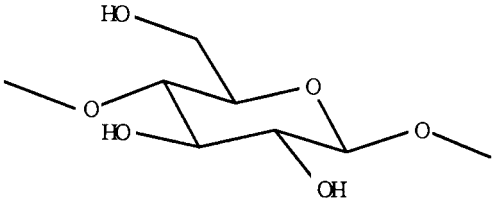
AIMS FOR FUTURE STRUCTURAL ANALYSIS: LESS MATERIAL USED SIMPLER ANALYSIS

In this paper a simpler and less time-consuming technique for sequence analysis of oligo- and polysaccharides, using computer evaluation of NMR data and simple chemical analysis, will be presented. This technique is based on the assumption that ^{1}H and ^{13}C NMR chemical shifts of the resonances from a monosaccharide residue within a larger saccharide depend mainly on the structure of the monosaccharide and on the nature of the flanking sugar residues. In the ^{1}H NMR spectra most of the ring protons have their resonances at similar chemical shifts forming a complex region (δ 3.5-4) with several overlapping signals. However, the resonances of some protons appear outside this region. These protons have been named "structural reporter groups" including, *inter alia*, anomeric protons, deoxy protons and some of the equatorially disposed ring protons. As the chemical shifts of the corresponding signals are dependent upon the components and linkages of the surrounding sugar residues they could be correlated to the structure. Computerised approaches to the structural analysis of oligo-saccharides from glycopeptides and glycolipids using the ^{1}H NMR chemical shifts of signals from the structural reporter groups have been reported (refs. 4-6). The chemical shifts are compared with those for the signals from corresponding residues in similar structures.

A computerised approach to the structural analysis of unbranched regular polysaccharides has been described (ref. 7). The method is based on evaluation of the ^{13}C NMR spectra for all possible structures made up of the constituent monosaccharides. It uses an additive scheme starting from the chemical shifts of the ^{13}C NMR resonances of the relevant monosaccharides to which the average values of the glycosylation shifts for signals from α- and β-carbons are added. This method was used in structural studies of some bacterial polysaccharides.

USE OF CASPER IN STRUCTURAL ANALYSIS OF OLIGO- AND POLYSACCHARIDES

We have described a computer program, CASPER, by which structural analysis of linear polysaccharides with repeating units could be performed (ref. 8). Sugar and methylation analysis data was used in combination with unassigned ^{13}C NMR chemical shifts. The program can handle linear oligo- and polysaccharides composed of repeating units for which ^{13}C NMR spectra for each possible permutation can be simulated and compared to the experimental spectrum. The simulation is based on an additivity approach, thus the chemical shift for each signal in the spectrum is assumed to be the sum of the chemical shift obtained from the monomeric residues and a number of induced chemical shift changes, glycosylation shifts, which depend upon the linkage positions and the stereochemistry around the glycosidic bonds (Fig. 2).



$\Delta\delta X + \delta 1\text{-}6 + \Delta\delta Y$

Fig. 2. The chemical shift values are the sum of the chemical shifts obtained from the monomer and the glycosylation shifts from each flanking sugar.

An improved and extended version of CASPER has recently been developed by which linear and branched structures can be analysed using one-dimensional ^{13}C NMR data, two-dimensional H,H- or C,H-COSY data. In addition coupling constants for anomeric protons and carbons can be used. The program consists of a database andprocedures for generation of 1D and 2D spectra, for fitting spectra, for comparison of *J*-values, for graphical display of spectra and an interface to the molecular modelling program CHEM-X.

```
*****************************************************************************
*    Welcome to                                                             *
*                CCCC   AAA    SSSS  PPPP   EEEEE  RRRR                    *
*               C      A   A  S     P   P  E      R   R                   *
*               C      A   A  S     P   P  E      R   R                   *
*               C      A   A   SSS  PPPP   EEEE   RRRR                    *
*               C      AAAAA      S P      E      R R                     *
*               C      A   A      S P      E      R  R                    *
*                CCCC  A   A  SSSS  P      EEEEE  R   R                   *
*                                                                          *
*      Computer Assisted SPectrum Evaluation of Regular polysaccharides    *
*   Created by G. Widmalm, P.-E. Jansson and L. Kenne. Version 2.0 July 1988 *
*****************************************************************************
   Top level menu

I : Information about CASPER
D : Determination of structure
R : Results of calculation
G : Graphical output
S : Spectrum Simulation of one structure
C : Chem-X 3D-file
E : End of CASPER run

Enter choice:
```

Fig. 3. The banner and the choices in the top level menu in CASPER.

The database contains ^{1}H and ^{13}C NMR chemical shifts for both anomeric forms of eleven monosaccharides of which both aminosugars and uronic acids are represented. The glycosylation shifts for all ^{1}H and ^{13}C NMR resonances for all types of disaccharide elements having pyranosidic rings, are included in the $\Delta\delta$-file of the database. These glycosylation shifts are mainly obtained from studies of 1,2-, 1,3-, 1,4- and 1,6-linked disaccharides with different stereochemistry around the glycosidic bond (refs. 9-13).

For some *Shigella* polysaccharides with branched O-polysaccharides it was shown that when the branched residue was not vicinally disubstituted additivity of glycosylation shifts holds, but for vicinally disubstituted residues deviations are found (ref. 14). To deal with such problems the extended version of CASPER contains a file with correction values for sugar residues in the branch point region. The correction values are obtained from studies on "branched trisaccharides" (ref. 13).

By the introduction of C,H-COSY spectra also the influence on the chemical shifts of the monosaccharide residue are considered. The basis for this is that the two-dimensional information present in a pair of chemical shifts is lost on going to one dimension, *i.e.*, the one-dimensional spectrum. An example of this is the chemical shifts for signals from H-3/C-3 and H-4/C-4 in β-Glc and in β-Gal, which pairwise appear at δ 3.50/76.6 and 3.42/70.7 for β-Glc and 3.59/73.8 and 3.89/69.7 for β-Gal. For the corresponding one-dimensional data the identification is less obvious.

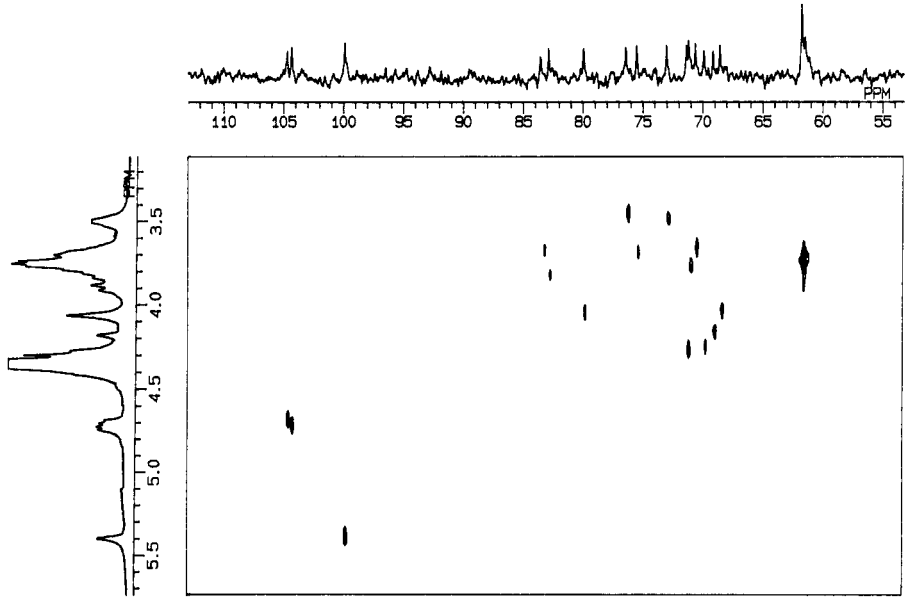
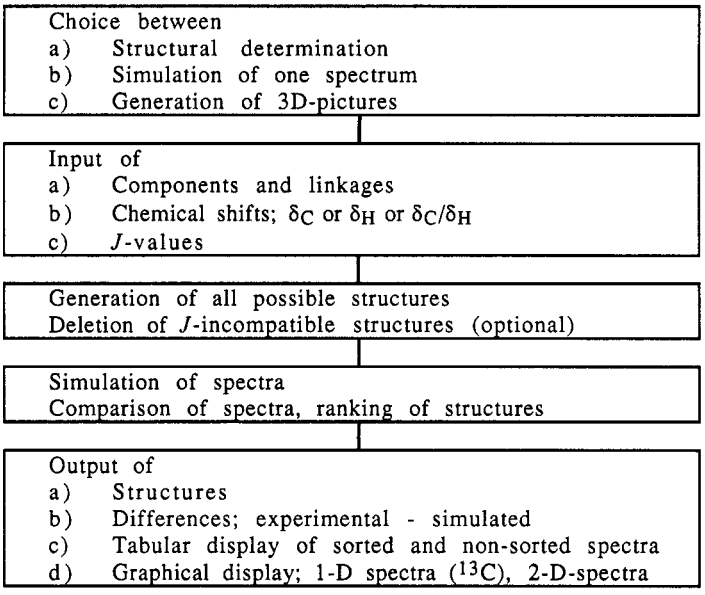


Fig. 4. The two-dimensional C/H-correlation spectrum of the Smith-degraded capsular polysaccharide from *Klebsiella* K8.

Another advantage with C,H-COSY spectra is the possibility to obtain all ^1H NMR chemical shifts also for polysaccharides which normally give spectra of low resolution and consequently less information. An example of this is given in Fig. 4. C/H-Correlation spectra has not been used extensively earlier because of the low sensitivity of the ^{13}C nucleus. Recently, however, a new technique, "inverse detection" HMQC, utilizing detection of the ^1H nucleus, has increased the sensitivity and made it possible to obtain C,H-COSY spectra of carbohydrates in considerably shorter time (ref. 15).
A flow diagram of the procedures for using CASPER in structural studies of carbohydrates is given below.



To use CASPER, data on components and linkage positions, normally obtained from sugar and methylation analysis, are given to the program. All possible structures are generated by permutation of the components and addition of all combinations of anomeric configurations. If the coupling constants for the anomeric protons, $^3J_{H,H}$, are available, these could be given to the program as the number of large (7-8 Hz), medium (3-4 Hz) and small (<2 Hz) coupling constants refering to the normal values for the β-*gluco/galacto*, α-*gluco/galacto* and α/β-*manno* configuration, respectively. Thus, the number 221 means that five sugar residues, two with β-*gluco*, two with α-*gluco* and one with α- or β-*manno* configuration, are the constituents of the repeating unit. The values for $^1J_{C,H}$ of the signals from the anomeric carbons, if available, are given as another number which refers to the size of the coupling constant, *ca* 170 Hz - α, or *ca* 160 Hz - β. A number of 32 thus means three α- and two β-configurations. These $^1J_{C,H}$ values also facilitate differentiation between residues having α- and β-*manno* configurations, as these have the same $^3J_{H,H}$ value, <2 Hz. A comparison of the given numbers, representing the anomeric configurations, with the numbers from all suggested structures can then be performed. This procedure drastically decreases the number of possible structures as all suggested structures with a non-corresponding set of anomeric configurations will be omitted.
In the next step the program will simulate the ^1H or ^{13}C NMR spectra or the C,H-correlation spectra from all possible structures, using chemical shifts, glycosylation shifts and correction values from the database, and compare these with the experimental spectrum. The fit, according to signal-to-signal comparison with the experimental spectrum, is calculated and the simulated spectra are ranked according to their fit. The results from CASPER include *inter alia* a chosen number of suggested structures ranked according to their fit with the experimental spectrum. In addition to these, the Δδ-sum, the deviation/signal, the values for the coupling constants and a check number are given for each suggested structure. The Δδ-sum is the total chemical shift difference between signals in the simulated spectrum and the

experimental spectrum. The structures are ranked according to this value. The check number describes the accuracy of the simulation, a low number is given when data from identical disaccharide elements or only minor approximations of these have been used. In addition the spectra can be shown and compared in the graphic mode as shown below.

APPLICATIONS OF CASPER IN STRUCTURAL STUDIES

The use of CASPER in structural studies of carbohydrates will be examplified by analysis of one oligosaccharide and three polysaccharides of known structure.

In the first example a branched oligosaccharide will be analysed using unassigned ^{13}C NMR chemical shifts only. The oligosaccharide, which represents the repeating unit in the O-polysaccharide from *Salmonella minnesota*, consists of four sugar residues of which a D-galactopyranosyl residue is branched. The oligosaccharide was synthesized and analysed with ^{13}C NMR spectroscopy by Kochetkov *et al.* (ref. 16) and the ^{13}C NMR chemical shifts were corrected for differences in temperature and reference before giving them to CASPER (Fig. 5). In this example only the ^{13}C NMR chemical shifts are used and the chemical shifts for signals from the α-form of the reducing residue were chosen. The results, obtained after permutation of the components, generation of the anomeric configurations, simulation of spectra and comparison of these with the experimental spectrum, are shown in Scheme 1, in which the four structures with the best fit are given. The correct structure has data with the best fit but the structure in which the rhamnopyranosyl residue has a β-configuration instead has only a somewhat larger Δδ-sum. If a comparison of the values for the coupling constants had been done structures 2 and 4 had been omitted due to the incompatible $^1J_{C,H}$-values.

The linear polysaccharide obtained by Smith-degradation of the capsular polysaccharide from *Klebsiella* K8 consists of a trisaccharide repeating unit. Also in this example only the unassigned ^{13}C NMR chemical shifts were used in the analysis. As all residues are 3-linked only the differences in the anomeric configurations and in the stereochemistry around the linkages will be the factors influencing the glycosylation shifts. After permutation of the components and generation of anomeric configurations the ^{13}C NMR spectra of the suggested structures were simulated.

The results of the analysis is shown in Scheme 2. The Δδ-sum showing the total deviation is much smaller for the first suggested structure than for the other structures. In the simulated spectrum of the second structure (no 4) it is possible to observe a large deviation for the signals from the anomeric carbons (Fig. 5). The third structure (no 6) has an incompatible number for the *J*-values and could also be omitted for that reason. The large Δδ-sums of the third and fourth structures show how different the glycosylation shifts are when the substitution position is adjacent to an axial substituent as *e.g.* in 3-substituted galactopyranosyl residues.

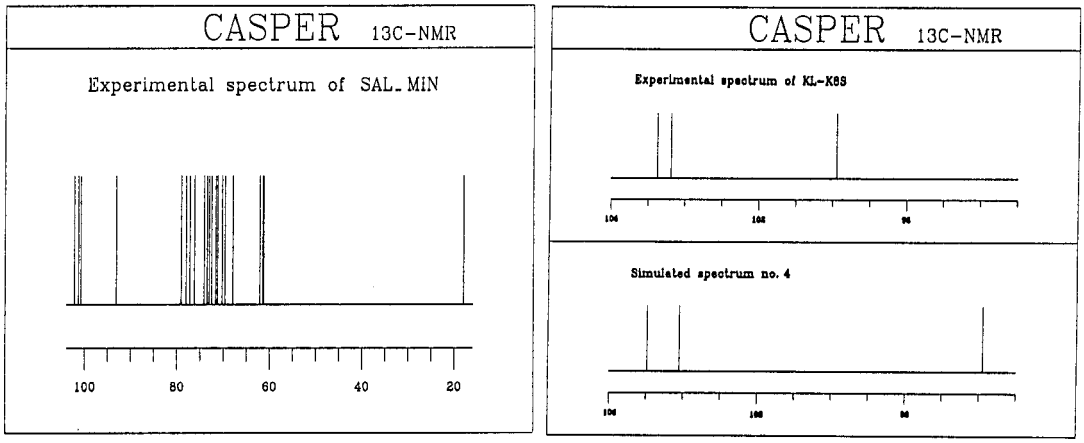


Fig. 5. Graphical output of the experimental ^{13}C NMR spectrum of the oligosaccharide representing the repeating unit in the O-antigen from *Salmonella minnesota* and the anomeric region of the experimental and a simulated spectrum from the Smith-degraded capsular polysaccharide from *Klebsiella* K8.

```
    SAL_MIN
No. Oligosaccharide.

  1  BDMAN -4ALRHA -3ADGAL
                      4
                    ADGLC

  2  BDMAN -4BLRHA -3ADGAL
                      4
                    ADGLC

  3  BDMAN -4ADGAL
              3
     ADGLC -4ALRHA

  4  BDMAN -3ADGAL
              4
     ADGLC -4BLRHA
```

No.	13C Deltasum	13C Sum /sig	J12	JCH	13C Check#	C-Branch Check#
1	9.0	0.38	22	31	0.21	0.01
2	9.4	0.39	22	22	0.21	0.01
3	10.8	0.45	22	31	0.21	0.01
4	11.3	0.47	22	22	0.30	0.01

```
13C Experimental spectrum.
 102.3 101.3 100.8  93.2  79.1  78.1  77.3  76.3  74.2  73.6
  73.4  73.1  72.6  71.7  71.3  71.3  70.3  69.7  67.9  67.9
  62.1  61.5  61.3  18.0

Spectrum number      1.
 103.1 101.7 100.8  93.3  81.6  79.0  77.1  76.9  74.0  73.5
  73.5  73.0  72.1  71.8  71.6  71.4  70.5  69.8  68.6  67.6
  61.9  61.6  61.5  17.7

Spectrum number      2.
 101.7 101.2 100.3  93.0  81.1  78.5  77.4  77.1  74.0  73.5
  73.4  73.1  72.9  71.9  71.8  71.8  70.7  70.6  68.4  67.6
  61.9  61.4  61.4  17.7
```

Scheme 1. Data from the four oligosaccharide structures with the best fit using ^{13}C NMR data in CASPER. Calculations were performed for the oligosaccharide, which represents the repeating unit of the O-antigen of *Salmonella minnesota*.

```
    KL-K8S
No. Polysaccharide.

  2 -3BDGLC -3BDGAL -3ADGAL -
  4 -3ADGLC -3BDGAL -3BDGAL -
  6 -3BDGLC -3BDGAL -3BDGAL -
  8 -3BDGLC -3ADGAL -3BDGAL -
```

No.	13C Deltasum	13C Sum /sig	J12	JCH	13C Check#
2	2.7	0.15	210	12	0.30
4	11.8	0.66	210	12	0.30
6	18.2	1.01	300	3	0.30
8	18.7	1.04	210	12	0.21

```
13C Experimental spectrum.
 104.8 104.4  99.9  83.4  82.9  80.0  76.4  75.6  73.1  71.4
  71.2  70.7  69.9  69.2  68.6  61.8  61.7  61.5

Spectrum number      2.
 104.9 104.5 100.3  83.8  83.5  80.3  76.4  75.6  73.1  71.7
  71.1  70.8  69.9  69.2  68.6  61.7  61.6  61.5

Spectrum number      4.
 105.0 104.1  95.9  83.5  83.4  78.8  75.8  75.7  72.6  71.6
  71.2  70.5  69.2  69.0  65.9  61.6  61.6  61.6
```

Scheme 2. Data from CASPER of the four structures for which the calculated ^{13}C NMR spectra have the best fit with the experimental spectrum of the Smith-degraded capsular poly-saccharide from *Klebsiella* K8.

```
    SH_FL_4A
No. Polysaccharide.

  2 -3BDGLCN-2ALRHA -2ALRHA -3ALRHA -
        6
      ADGLC

  4 -3BDGLCN-2ALRHA -3ALRHA -2ALRHA -
        6
      ADGLC

  6 -3BDGLCN-3ALRHA -2ALRHA -2ALRHA -
        6
      ADGLC

  8 -6BDGLCN-2ALRHA -2ALRHA -3ALRHA -
        3
      ADGLC


 No.    13C     13C Sum      13C
      Deltasum    /sig      Check#

  2      7.4     0.23       0.32
  4      7.5     0.23       0.32
  6      8.7     0.27       1.22
  8     11.1     0.35       0.32

 J12 = 113 and JCH = 41 used to eliminate structures

13C Experimental spectrum.
 174.9 103.0 101.8 101.8 101.6  99.1  82.2  79.6  79.1  78.3
  75.2  74.1  73.3  73.2  72.8  72.4  72.4  71.5  70.9  70.8
  70.7  70.0  69.9  69.9  69.4  67.0  61.7  56.4  23.2  17.6
  17.5  17.4

Spectrum number      2.
 175.5 103.9 102.5 102.0 101.8  98.8  82.7  80.3  78.8  78.8
  75.4  74.1  73.6  73.0  72.8  72.5  72.4  71.3  71.1  70.9
  70.7  70.0  70.0  69.9  69.5  66.9  61.7  57.5  23.1  17.6
  17.4  17.4

Spectrum number      4.
 175.5 103.9 102.8 102.1 101.4  98.8  82.7  79.8  79.7  78.7
  75.4  74.1  73.4  73.2  72.8  72.5  72.4  71.1  71.0  70.8
  70.7  70.1  70.0  69.8  69.5  66.9  61.7  57.5  23.1  17.5
  17.5  17.4

Spectrum number      2.
  98.8  72.4  74.1  70.7  72.8  61.7
 103.9  57.5  82.7  69.5  75.4  66.9  23.1 175.5
 102.0  80.3  71.1  73.6  69.9  17.4
 101.8  78.8  70.9  73.0  70.0  17.6
 102.5  71.3  78.8  72.5  70.0  17.4

Spectrum number      4.
  98.8  72.4  74.1  70.7  72.8  61.7
 103.9  57.5  82.7  69.5  75.4  66.9  23.1 175.5
 102.1  79.8  71.1  73.4  69.8  17.5
 102.8  70.8  78.7  72.5  70.1  17.4
 101.4  79.7  71.0  73.2  70.0  17.5
```

Scheme 3. Data from CASPER of the four structures for which the calculated ^{13}C NMR spectra have the best fit with the experimental spectrum of *Shigella flexneri* type 4a O-polysaccharide.

For the branched structure of the repeating unit of *Shigella flexneri* type 4a, 1536 possible structures were generated by permutation of the components and addition of all combinations of the anomeric configuration. If $^3J_{H,H}$ values are used to omit *J*-incompatible structures the number of possible structures decreases to 576. After the same procedure with also the $^1J_{C,H}$ values only 288 possible structures remained. Fitting the simulated spectra from these structures with the experimental ^{13}C NMR spectrum (Fig. 6) was then performed with the program.

For the calculation of spectra no correction factors were used as there is no vicinal disubstitution of the branching residue. In the program there are yet no special glycosylation shifts for the acetamido group yet and this is partly the reason for the rather high $\Delta\delta$-sums obtained for all structures. Only minor differences in the $\Delta\delta$-sum are found between the four structures with the best fit. The only differences in the structures are the sequence of the three α-L-rhamnopyranosyl residues (Scheme 3).

To obtain the correct structure specific chemical reactions or NMR experiments have to be performed.

When using C,H-correlation data it is recommendable to use *J*-values for omitting structures and simulation of the ^{13}C NMR spectra must first be performed and only the structures with the best fit should be used in the fitting of C,H-correlation spectra as this procedure is time-consuming. The chemical shift difference is not given in ppm as a weighting factor is given to the ^{1}H NMR chemical shifts for which there are smaller glycosylation shifts. An example of C,H-correlation simulations is described for the Smith-degraded capsular polysaccharide from *Klebsiella* K8. The simulated spectra and a comparison with the experimental spectrum can also be shown in the graphic mode (Fig. 7).

The structures with the best fit are the same as when using only ^{13}C data but with different order of the two last structures (Scheme 4). The $\Delta\delta$-sums are higher than in the ^{13}C NMR simulation described above and are caused by the weighting factor used for the ^{1}H NMR data.

In the last example only ^{1}H NMR data are used for calculation the structure of the repeating unit of the O-polysaccharide from *Shigella flexneri* type Y. The polysaccharide is a linear tetrasaccharide repeat and the chemical shifts for the signals have been determined using different H,H-COSY experiments (ref. 17). For the calculations only information on com-ponents, linkages and $^{3}J_{H,H}$ values together with the ^{1}H NMR chemical shifts for unassigned signals were given to the program. Structures with incompatible $^{3}J_{H,H}$ values were omitted before the calculations. The ^{1}H NMR chemical shifts for all possible structures were calculated and fitted to the experimental data. The results could be displayed in the graphic mode (Fig. 8) or in the tabular form (Scheme 5).

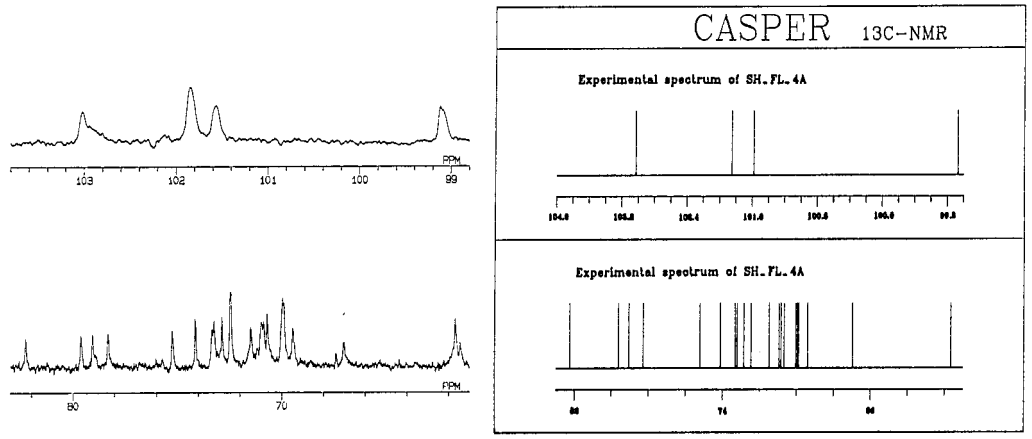


Fig. 6. Experimental spectrum and corresponding graphical output in CASPER of the *Shigella flexneri* type 4A polysaccharide.

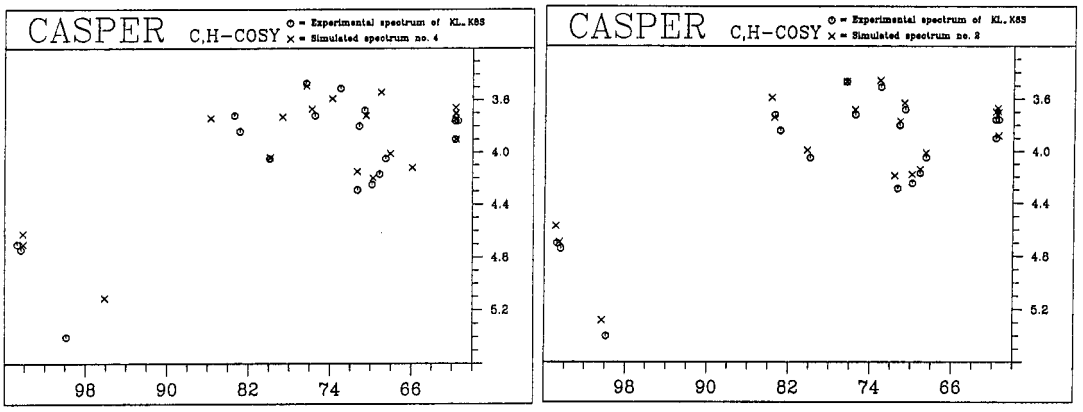


Fig. 7. Graphical output of two-dimensional C/H-correlation spectra. The experimental spectrum of the Smith-degraded capsular polysaccharide from *Klebsiella* K8 and the simulated spectra of the structures with best fit are shown.

```
    KL_K8S
No. Polysaccharide.

  2 -3BDGLC -3BDGAL -3ADGAL -
  4 -3BDGLC -3ADGAL -3BDGAL -
  6 -3ADGLC -3BDGAL -3BDGAL -

No.   H,C       13C       1H
    Deltasum   Check#    Check#

  2    13.6      0.30      0.30
  4    17.3      0.21      0.21
  6    18.4      0.30      0.30

J12 = 210 and JCH = 12 used to eliminate structures
H,C-Deltasum calculated with point of gravity algorithm

Experimental H,C-correlation spectrum.
  104.8 4.70  104.4 4.74   99.9 5.40   83.4 3.72   82.9 3.84
   80.0 4.05   76.4 3.47   75.6 3.72   73.1 3.51   71.4 4.29
   71.2 3.80   70.7 3.68   69.9 4.25   69.2 4.17   68.6 4.05
   61.8 3.76   61.8 3.76   61.7 3.76   61.7 3.90   61.5 3.76
   61.5 3.76

Spectrum number     2.
  104.9 4.57  104.5 4.69  100.3 5.28   83.8 3.59   83.5 3.74
   80.3 3.99   76.4 3.47   75.6 3.68   73.1 3.46   71.7 4.19
   71.1 3.77   70.8 3.63   69.9 4.18   69.2 4.14   68.6 4.01
   61.7 3.70   61.7 3.70   61.6 3.72   61.6 3.67   61.5 3.88
   61.5 3.70

Spectrum number     4.
  104.2 4.62  104.2 4.70   96.1 5.11   85.8 3.74   79.9 4.04
   78.7 3.73   76.4 3.49   75.9 3.67   73.9 3.59   71.4 4.15
   70.5 3.72   69.8 4.20   69.1 3.54   68.1 4.01   66.0 4.12
   61.7 3.75   61.7 3.66   61.7 3.71   61.7 3.70   61.6 3.90
   61.6 3.74
```

Scheme 4. Data from CASPER of the three structures for which the calculated C,H-correlation spectra have the best fit with the experimental spectrum.

The first suggested structure is the repeating unit of the O-polysaccharide. In the non-sorted data it is possible to see which of the simulated signals deviate from the experimental values. The largest deviations are found for the signals from protons close to the *N*-acetyl group. In a molecular model of the two disaccharide elements (Fig. 9) containing the *N*-acetyl-β-D-glucosamine residue interactions of the *N*-acetyl group with neighbouring protons can be observed. Some correction factors for the influence of these interactions on the glycosylation shifts have been introduced.
The ^{1}H NMR data can easily be obtained for the H-1 to H-4 signals of each residue in a polysaccharide, using different H,H-COSY experiments, and for many examples this is enough to find the correct structure using CASPER. This makes it possible to analyse polysaccharides of which there is only a small amount available.

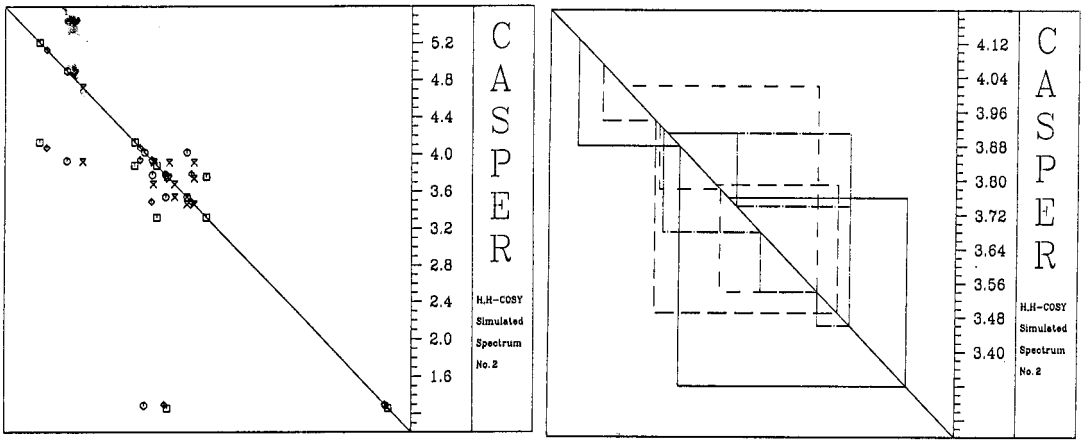


Fig. 8. Graphical output of the H,H-COSY spectrum from *Shigella flexneri* type Y polysaccharide. Each marker shows a certain spin-system (left) and spin-coupling connectivities for ring protons are shown by different traces (right).

```
      SHIG_FL_Y-H
  No. Polysaccharide.

    2 -2ALRHA -2ALRHA -3ALRHA -3BDGLCN-
    4 -2ALRHA -2ALRHA -3BLRHA -3BDGLCN-
    6 -2ALRHA -3ALRHA -2ALRHA -3BDGLCN-
    8 -2ALRHA -3ALRHA -2BLRHA -3BDGLCN-

  No.     1H       1H Sum     JCH      1H
        Deltasum    /sig              Check#

    2     0.48      0.02       31       0.22
    4     0.55      0.02       22       0.31
    6     0.88      0.03       31       0.22
    8     1.06      0.04       22       1.21

J12 = 103 used to eliminate structures

1H Experimental spectrum.
 5.16 5.15 4.88 4.75 4.14 4.06 4.01 3.93 3.91 3.87
 3.87 3.84 3.79 3.76 3.76 3.69 3.66 3.55 3.55 3.49
 3.46 3.34 2.06 1.31 1.26 1.26

Spectrum number      2.
 5.21 5.13 4.90 4.72 4.13 4.07 4.02 3.94 3.93 3.92
 3.91 3.88 3.79 3.78 3.76 3.74 3.68 3.54 3.54 3.49
 3.46 3.32 2.06 1.30 1.29 1.26

Spectrum number      4.
 5.21 5.13 4.88 4.76 4.13 4.11 4.07 3.94 3.91 3.88
 3.84 3.79 3.76 3.76 3.75 3.69 3.66 3.50 3.49 3.48
 3.46 3.32 2.06 1.34 1.30 1.26

Spectrum number      2.
 5.21 4.13 3.88 3.32 3.76 1.26
 5.13 4.07 3.94 3.49 3.79 1.30
 4.90 3.93 3.78 3.54 4.02 1.29
 4.72 3.92 3.68 3.54 3.46 3.74 3.91 2.06

Spectrum number      4.
 5.21 4.13 3.88 3.32 3.76 1.26
 5.13 4.07 3.94 3.49 3.79 1.30
 4.88 4.11 3.69 3.50 3.46 1.34
 4.76 3.84 3.75 3.66 3.48 3.76 3.91 2.06
```

Scheme 5. Data from CASPER on the four structures for which the calculated ^{1}H NMR spectra have the best fit with the experimental spectrum of *Shigella flexneri* type Y O-polysaccharide. The non-sorted data for the two structures with the best fit are also given.

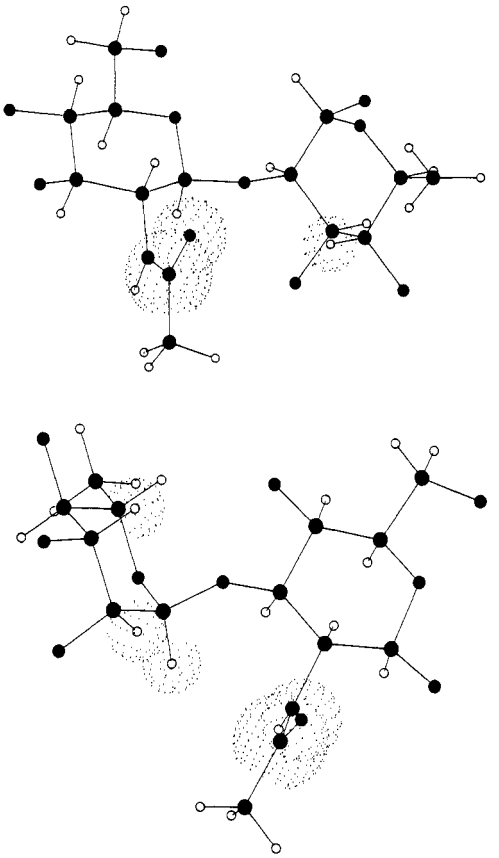


Fig. 9. Molecular models of disaccharide elements from the O-antigen from *Shigella flexneri* type Y. Inter-residue interactions from the *N*-acetyl group are depicted.

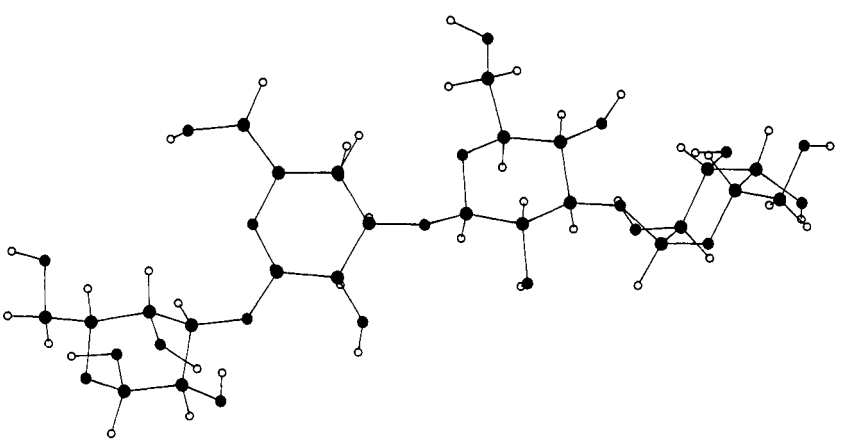


Fig. 10. Molecular model of the repeating unit of the Smith-degraded capsular polysaccharide from *Klebsiella* K8. The model is produced by the molecular modelling program CHEM-X, which is interfaced with CASPER, and the structures obtained in CASPER can be built and studied in CHEM-X.

In CASPER there is also an interface to the molecular modelling program CHEM-X (ref. 18). The suggested structure will automatically be built using standard parameters and the structure could be displayed as a 3D-picture (Fig. 10).

CONCLUSION AND FUTURE ASPECTS

The extension of CASPER to the use of more NMR parameters gives a larger possibility to select the correct structure. For compounds for which only ^1H NMR data can be obtained it is now possible to run a computer analysis. The inclusion of branched structures makes it possible to investigate a large number of carbohydrate structures. For some polysaccharides the database will not be accurate enough or $\Delta\delta$-sum differences too small to differentiate between some structures. Additional chemistry or NMR experiments then have to be engaged. These can normally be selective as most of the disaccharide elements in the structure are established through the computer analysis.

It is obvious that continuous investigations on model substances will enhance the quality of the database. There is also a large amount of information obtained from known polysaccharide structures which are examined by CASPER. The integration of such information automatically into the database is a challenging next step in the development.

Acknowledgements

This work was supported by grants from the Swedish Natural Science Research Council and the Swedish National Board for Technical Development. Herbert Baumann, Bertil Erbing, Irene Marklund, Håkan Ottosson, Elke Schweda, Marianne Walding, and Thomas Wehler are gratefully acknowledged for their contribution to the NMR studies of model substances for the database.

REFERENCES

1. G.O. Aspinall, in G.O. Aspinall (Ed.), *The Polysaccharides, Vol 1*, Academic Press, New York, 35-131 (1982).
2. A. Dell, *Adv. Carbohydr. Chem. Biochem.*, *45*, 19-72 (1987).
3. J.F.G. Vliegenthart, L. Dorland and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, *41*, 209-374 (1983).
4. E.F. Hounsell, D.J. Wright, A.S.R. Donald and J. Feeney, *Biochem. J. 223*, 129-143 (1984).
5. D.R. Anderson and W.J. Grimes, *Anal. Biochem. 146*, 13-22 (1985).
6. D.S.M. Bot, P. Cleij, H.A. van't Klooster, H. van Halbeek, G.A. Veldink and J.F.G. Vliegenthart, *J. Chemometr. 2*, 11-27 (1988).
7. G.M. Lipkind, A.S. Shashkov, Y.A. Knirel, E.V. Vinogradov and N.K. Kochetkov, *Carbohydr. Res. 175*, 59-75 (1988).
8. P.-E. Jansson, L. Kenne and G. Widmalm, *Carbohydr. Res. 168*, 67-77 (1987).
9. M. Forsgren, P.-E. Jansson and L. Kenne, *J. Chem. Soc., Perkin Trans. 1*, 2383-2388 (1985).
10. H. Baumann, P.-E. Jansson and L. Kenne, *J. Chem. Soc., Perkin Trans. 1*, 209-217 (1988).
11. I. Backman, P.-E. Jansson and L. Kenne, *J. Chem. Soc., Perkin Trans. 1*, 889-898 (1988).
12. P.-E. Jansson, L. Kenne and E. Schweda, *J. Chem. Soc., Perkin Trans. 1*, 2729-2736 (1988).
13. H. Baumann, B. Erbing, P.-E. Jansson and L. Kenne, *J. Chem. Soc., Perkin Trans. 1*, submitted.
14. P.-E. Jansson, L. Kenne and T. Wehler, *Carbohydr. Res. 179*, 359-368 (1988).
15. A. Byrd, W. Egan, M.F. Summers and A. Bax, *Carbohydr. Res. 166*, 47-58 (1987).
16. N.K. Kochetkov, V.I. Torgov, N.N. Malysheva, A.S. Shashkov and E.M. Klimov, *Tetrahedr. 36*, 1227-1230 (1980).
17. P.-E. Jansson, L. Kenne and T. Wehler, *Carbohydr. Res. 166*, 271-282 (1987).
18. CHEM-X, developed and distributed by Chemical Design Ltd, Oxford, England.