186. Computer-Assisted Structural Interpretation of ^{1}H-NMR. Spectral Data[1])

by **Huldrych Egli**[2]), **Dennis H. Smith** and **Carl Djerassi**

Department of Chemistry, Stanford University, Stanford, California 94305, U.S.A.

(8.IV.81)

Summary

Interactive computer programs for the establishment and maintenance of a ^{1}H-NMR. data base, the prediction of ^{1}H-NMR. shifts and the rank-ordering of structural candidates based on comparison between observed and predicted spectra are presented. The programs take into account configuration, at stereocenters and double bonds, as well as diastereotopy. We demonstrate how, for purposes of structure elucidation, these new programs can be linked to the **GENOA** and **STEREO** programs.

1. Introduction. – Earlier work in the area of computer aids to structure elucidation using ^{1}H-shift data has, to our knowledge, never explicitly included configuration. *Skolnik* [2] described structures with a linear notation system denoting C-atoms in terms of bonds and attached H-atoms. He demonstrated that this system could be used to correlate proton groups in organic molecules with chemical shifts. *Mlynarik et al.*, [3] used a slightly modified linear notation system for the coding of structures allowing also substructure search. A file search system for mutual assignment of subspectra and substructures was established handling ^{1}H- and ^{13}C-spectral data. *Erni & Clerc* [4] used an efficient search system for spectral similarities of an unknown compound and references in a data base. This structure search is based on weighted 'signatures' considering IR., ^{1}H-NMR. and MS. data. Other spectral retrieval systems were designed to search for compound names rather than structures [5] [6] or are limited in their scope for application to open-chain structures with essentially first-order spectra [7].

As a part of the structure elucidation programs in the **DENDRAL** project, algorithms were developed to deal with the analysis of MS. [8] [9] and ^{13}C-NMR. data [10–12]. The encouraging results from the analysis of ^{13}C-NMR. spectral data [1] led us to examine the applicability of these methods to the analysis of ^{1}H-NMR. data. In this report we describe how ^{1}H-shift data, collected in a suitable

[1]) Part XLI. of the series 'Application of Artificial Intelligence for Chemical Inference'. For part XL. see [1].

[2]) Present address: *Spectrospin AG*, Industriestrasse 26, CH-8117 Faellanden.

0018-019X/82/6/1898-23$01.00/0

data base, can be applied to eliminate incompatible candidates from the list of structures produced by exhaustive generation of constitutional isomers with **GENOA** [13] and stereoisomers with **STEREO** [14] [15].

2. Methods. – Our approach to use of ^{1}H-NMR. data in computer-assisted structure elucidation involves obtaining a set of candidate structures for an unknown compound using chemical and spectroscopic information in conjunction with a program for structure generation, such as **GENOA** [13]. These structural candidates can then be evaluated by prediction of the ^{1}H-NMR. spectrum for each candidate, comparison of the predicted and observed spectra, and rank-ordering of the candidates based on such comparisons. Our procedure for spectrum prediction is strictly empirical. We seek only to derive a set of expected chemical shifts for the protons in each candidate. These predictions are based on what is in effect a very large and detailed correlation chart, a data base, relating structural features to chemical shifts. Our approach makes use of several interactive computer programs designed to assist the chemist at each step in the method.

Our programs for the analysis of ^{1}H-NMR. spectra consist of:

1) Programs to build and maintain a ^{1}H-NMR. data base that correlates substructural environments with observed proton resonances (chemical shifts) for known compounds, including: a) a program, HCODE, for data base construction, including automatic generation of substructure codes; b) programs, HDCODE and HDBCHK, to help validate substructure/shift assignments in the data base.

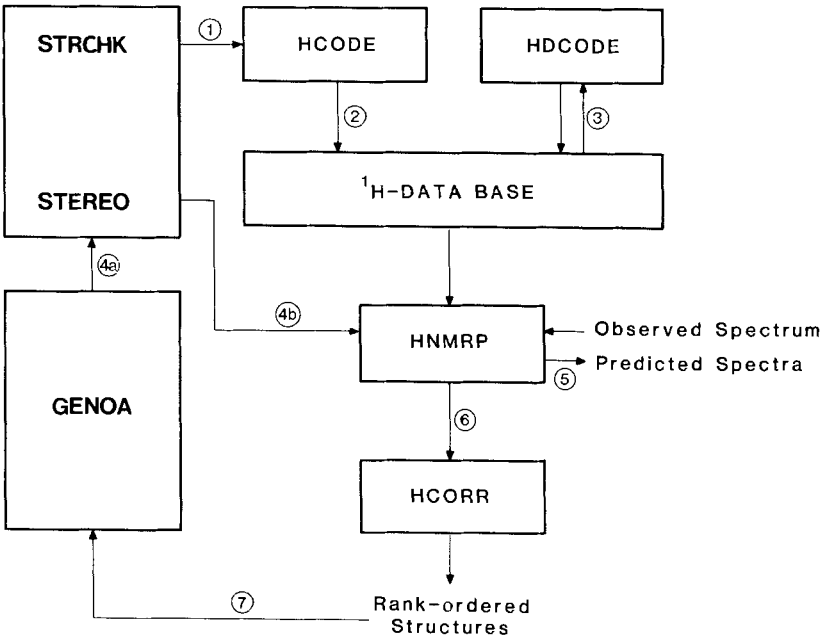
2) Programs to analyze the ^{1}H-NMR. spectrum of an unknown compound, including: a) a program, HNMRP, to predict the spectrum of one or more candidate structures for an unknown compound; b) a program, HCORR, to compare the predicted spectra of candidate structures to the observed spectrum of the unknown structure and rank-order the candidates on the basis of agreement between predicted and observed.

We present in *Scheme 1* how these programs are embedded within other **DENDRAL** programs for structure elucidation.

2.1. *Building and maintaining the data base.* The organization of a data base relating chemical structures and ^{1}H-NMR. shifts could take several forms depending on the application. For our purposes of structure elucidation outlined at the beginning of this section we need to utilize the data base for structural analysis of compounds that are unlikely to be represented in the data base. However, it *is* likely that portions, or substructures, of novel compounds will be represented in the data base. Therefore, the organization we have chosen is similar to that chosen for ^{13}C-NMR. analysis [11]. This organization includes descriptions of the local substructural environment of resonating protons together with chemical shifts. Further, a shell structure is imposed on the substructural representations that corresponds to consideration of α-, β-, γ-, and δ-substituent effects. These representations are discussed in more detail in subsequent sections.

2.2. *Standard atoms and extended atom types.* In describing chemical structures to our programs we must include a sufficient number of different types of atoms to cover most structural types that might be analyzed by ^{1}H-NMR. spectroscopy. Further, we differentiate several of the standard chemical atoms by number of

Scheme 1. *Organization of the ^{1}H-NMR. analysis programs.* 1) Definition of known structures; 2) additions to the data base; 3) maintenance and checking of the ^{1}H-data base; 4a) generation of constitutional isomers; 4b) generation of stereoisomers; 5) spectrum prediction; 6) rank-ordering of predicted spectra compared to observed shifts; and 7) use of results as further constraints.



H-substituents and hybridization. Our program currently recognizes 18 standard atoms further broken down into 55 extended atom types or functional groups, as summarized in *Table 1.* The term *standard atom* is merely a symbol for an atom or functionality whose real description is perceived automatically by the programs as given in *Table 1*, column 4.

Table 1. *Standard atoms and extended atom types* (functional groups)

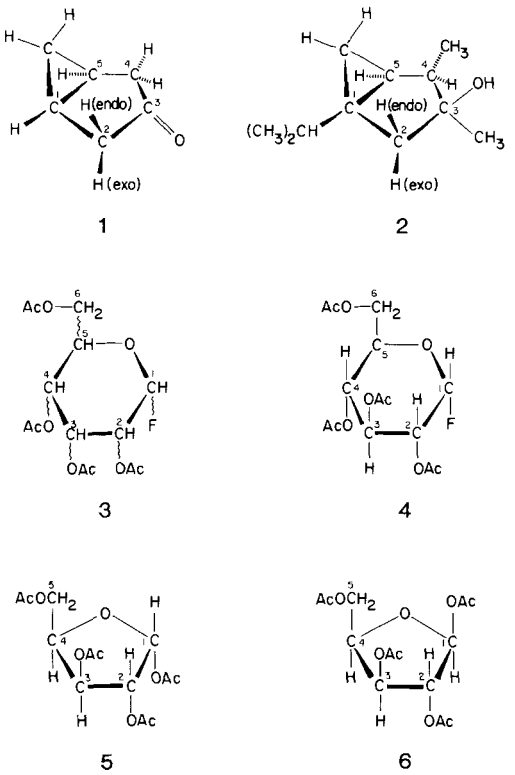
Code number	Standard atom	Valence	Extended atom type or functional group[a])	Atom code
1	C	4	$-CH_3$	1
2			$-CH_2-$	2
3			>CH–	3
4			>C<	4
5			$CH_2=$	5
6			–CH=	6
7			>C=	7
8			–C*H=(arom.)	8
9			>C*=(arom.)	9
10			HC≡	A
11			–C≡	B
12			=C=	C

Table continued

Code number	Standard atom	Valence	Extended atom type or functional group[a])	Atom code
13	N	3	$-NH_2$	D
14			$-NH-$	E
15			$>N-$	F
16			$=NH$	G
17			$=N-$	H
18			$\equiv N$	I
19			$-N*H-$ (arom.)	J
20			$>N*-$ (arom.)	K
21			$=N-$(arom.)	L
22	O	2	$-OH$	M
23			$-O-$	N
24			$=O$	O
25			$-O*-$ (arom.)	P
26	S	2	$-SH$	Q
27			$-S-$	R
28			$=S$	S
29			$-S*-$ (arom.)	T
30	F	1	$-F$	U
31	Cl	1	$-CL$	V
32	Br	1	$-BR$	W
33	I	1	$-I$	X
34	SO_2	2	$-SO_2-$	Y
35	NO	3	$>NO-$	Z
36	PO	3	$-POH_2$	a
37			$>POH$	b
38			$>PO-$	c
39	SO	2	$-SO-$	d
40	NO_2	1	$-NO_2$	e
41	P	3	$-PH_2$	f
42			$>PH$	g
43			$>P-$	h
44	D	1	$-D$	i
45	O^-	1	$-O^-$	j
46	N^+	4	$-N^+H_3$	k
47			$>N^+H_2$	l
48			$>N^+H-$	m
49			$>N^+<$	n
50			$=N^+H_2$	o
51			$=N^+H-$	p
52			$=N^+<$	q
53			$\equiv N^+H$	r
54			$*N^+H$	s
55			$*N^+$	t
56			?	u

2.3. *Definition of structures.* In building the data base our programs require first the definition of the structures of the compounds whose spectral data are to be added to the data base. The constitution and configuration of each structure is defined using a structure-editing module in the **STRCHK** program *(Scheme 1)*,

Scheme 2



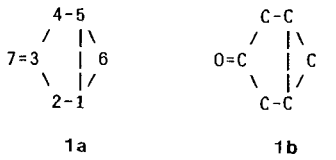
using as atom types the standard atoms summarized above. Thus, from the very beginning, our structural representation includes configuration; subsequently, other programs will automatically analyze the assigned configuration and include it in the data base. As an example, we present in *Scheme 2* the molecular structure of bicyclo[3.1.0]hexan-3-one (**1**), using both atom numbers and names, together with the corresponding connection table as presented by our programs. The configurational designations given in the connection table are assigned by the program based on the convention described previously [14] [15].

2.4. *The* HCODE *program.* The HCODE program, adapted from programs used for analysis of ^{13}C-NMR. data [10] [11], is used to build the data base. HCODE has an interactive segment that allows a chemist to assign resonance values (chemical shifts) to H-atoms in known structures previously defined in **STRCHK** ((1) in *Scheme 1*). It also has a segment that automatically generates *codes* which represent the substructural environments of the H-atoms for subsequent addition to the data base ((2) in *Scheme 1*).

It is the function of HCODE, subsequent to interaction (an example of which is given below), to take a known structure and its partially or fully assigned spectrum and to generate an atom-centered, canonical code [11] that defines the substructural

Scheme 3. *Structure representations of bicyclo[3.1.0]hexan-3-one* (**1**)

a) *Molecular structure*



1a **1b**

b) *Connection table including stereochemical information*

ATOM#	TYPE	NEIGHBORS				CONFIG
1	C	5	6	2		'1'
2	C	3	1			
3	C	7	7	4	2	
4	C	5	3			
5	C	1	6	4		'1'
6	C	1	5			
7	O	3	3			

environment of each resonating proton. By canonical we mean unique and unambiguous, because it is a critical requirement of our method that substructures identical in constitution and configuration *always* be given the same name, independent of the identity of the structures in which they occur. The coding scheme is also independent of the structure numbering, so that the same code is generated whether a structure is numbered according to chemical convention or numbered arbitrarily by a chemist (or a program).

2.5. *Constitutional code of a substructure.* The first step in generating canonical codes for proton environments is to describe the constitution of the substructure. In the constitutional code the atoms are symbolized by the atom codes given previously *(Table 1)*. The atom codes in this substructure representation are ordered according to their 'distances' from the resonating atom expressed in *bond-radii* [10] [11], or *shell-level*. Thus, for a proton code the H-atom is at shell-level zero and the atom attached to it at the distance of one bond-radius belongs to shell-level one. The view of the substructure while encoding includes atoms out to shell-level five.

Beginning with shell-level two, more than one atom-code per shell-level has to be expected. For such a case the various atoms follow one another reflecting the priorities given by the procedure that derives the canonical codes. The codes obtained from this procedure are merely strings of characters. Included with the strings are special symbols that indicate the separation of shell-levels and ring closures in cyclic systems [11]. These character strings are easily compared in the computer, as complete strings or as partial strings up to selected shell-levels.

2.6. *Encoding of configuration.* We have described the incorporation of configuration as part of the initial structure definition *(Scheme 2)*. The codes generated by HCODE also include a canonical description of the stereochemical environment of each proton [11]. In the code describing the constitution and configuration of a substructure the shell-oriented stereo codes are preceded by a '+'-sign.

2.7. *Encoding of diastereotopic protons.* Diastereotopic methylene protons have to be distinguished. This is achieved by temporarily converting the atom with the geminal H-atoms into a stereocenter. Thus a canonical representation of such diastereotopic protons similar to the *'re/si'* designation is produced and the protons are coded uniquely. Diastereotopic vinyl protons are handled in a similar way. A configuration at a double bond is defined by appropriate stereo tags which are either one or zero at each of the two atoms connected by the double bond. The characterization of geminal diastereotopic vinyl protons is performed by distinguishing the protons and expressing their *cis*- or *trans*-relationship to other substituents on the double bond.

2.8. *Spectrum assignment in* HCODE. As examples to show interaction with HCODE we use the structure of **1** *(Scheme 2).* Protons are not explicitly defined in the connection tables set up by the **CONGEN, GENOA** or **STRCHK** programs. But every standard atom is defined in the program together with its valence *(Table 1).* The number of H-atoms at any particular atom is evaluated as a difference between the valence and the number of bonds originating at this atom. For each atom bearing H-atoms, HCODE begins a dialog with the chemist to assign the appropriate chemical shift. There are essentially three cases to be distinguished.

2.8.1. *Atoms bearing one H-atom.* The proton implicitly receives the same number as the atom to which it is bonded. The dialog with HCODE for C(1) of **1** is demonstrated in the following example (the responses provided by the chemist are given in italics to help differentiate the computer's output from chemist's input).

>CH – 1 Assigned Shift? *Y*
Shift value: *1.54*

The program prints out the atom type, a (>CH)-group, and the number of the atom, in this instance 1. Shift values are entered in ppm with respect to TMS. If shifts are not assigned, or not available, the chemist's response is *NO*, or *N*.

Atoms C(1) and C(5) of structure **1** are equivalent due to symmetry. Thus, the codes obtained for their H-substituents must be the same. The codes produced for the isochronous protons at the atoms C(1) and C(5) are shown here to indicate that they are indeed given the same code, although in actual use of the programs the chemist is never confronted with such codes without accompanying drawings of the substructures they represent:

Atom 1: 154 0 0/3/223)2+1/72)5+1/O+1$
Atom 5: 154 0 0/3/223)2+1/72)5+1/O+1$

2.8.2. *Atoms bearing two protons.* – a) sp^3*-centers.* HCODE displays a *Fischer*-type geometric representation of the prochiral center. In *Scheme 4* we present the dialog with HCODE for the geminal protons at C(2) of **1**. Referring to the numbering established at the time of the structure definition the chemist may choose the cases ONE or ZERO which results in the processing of the selected protons. If the geminal protons are not diastereotopic the choice would be BOTH. The option NONE is

Scheme 4. *Display of diastereotopic methylene protons at sp^3-centers and assignment of shifts*

```
-CH2- 2

   H*          H           H*
   |           |           |
3>>C<<1     3>>C<<1     3>>C<<1
   |           |           |
   H           H*          H*

  ONE         ZERO      BOTH OR NONE

WHICH *H(s) TO PROCESS? ONE
Assigned Shift? YES
Shift value : 2.12
2nd *H SHIFT ASSIGNED? (Y or CR)
2nd shift value : 2.58
```

meant for unassigned shifts and results in skipping of the coding procedure for this particular substructure.

The entries to the data base created for the four protons at C(2) and C(4) of **1** are shown below. The shift-structure code relations are identical for the two *exo*-protons and the two *endo*-protons as they should be, because these two pairs of diastereotopic protons are chemically equivalent and isochronous.

Atom 2, choice ONE, (*endo*):	212 0 0/2/37/23)402)5 + 1$
Atom 2, 2nd proton, (*exo*):	258 0 0/2/37/23)402)5$
Atom 4, choice ONE, (*endo*):	212 0 0/2/37/23)402)5 + 1$
Atom 4, 2nd proton, (*exo*):	258 0 0/2/37/23)402)5$

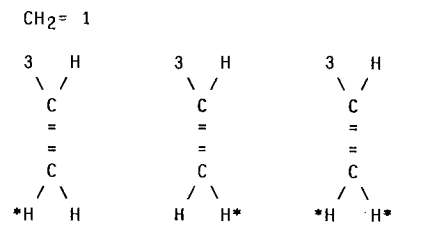
b) *Diastereotopic protons at* sp^2*-atoms.* A geometric representation displayed to the chemist allows the distinction of the two diastereotopic protons at a double bond. The procedure is similar to the encoding at sp^3-centers. The dialog with HCODE is demonstrated in *Scheme 5.*

The treatment of diastereotopic protons on double bonds includes proper analysis of terminal (=CH_2)-groups of allenes.

2.8.3. *Atoms bearing three protons.* For atoms bearing three protons, such as CH_3-groups, the dialog with HCODE is similar to that shown above for (>CH)-groups. The atom type and number is given and a request is made for entry of the observed chemical shift.

2.9. *The* HDCODE *and* HDBCHK *programs.* Once known structures and spectra are processed by HCODE and the resulting substructural codes added to the data base ((2) in *Scheme 1*), the HDCODE and HDBCHK programs can be used to check the data base for erroneous assignments ((3) in *Scheme 1*). These programs are modified versions of the checking programs DCODE and DBCHECK for the ^{13}C-NMR. data base [1]. These programs perform consistency checks on the substructure/shift combinations in the data base and report combinations that have an anomalously high shift range. In this way the data base can be monitored continously as it is

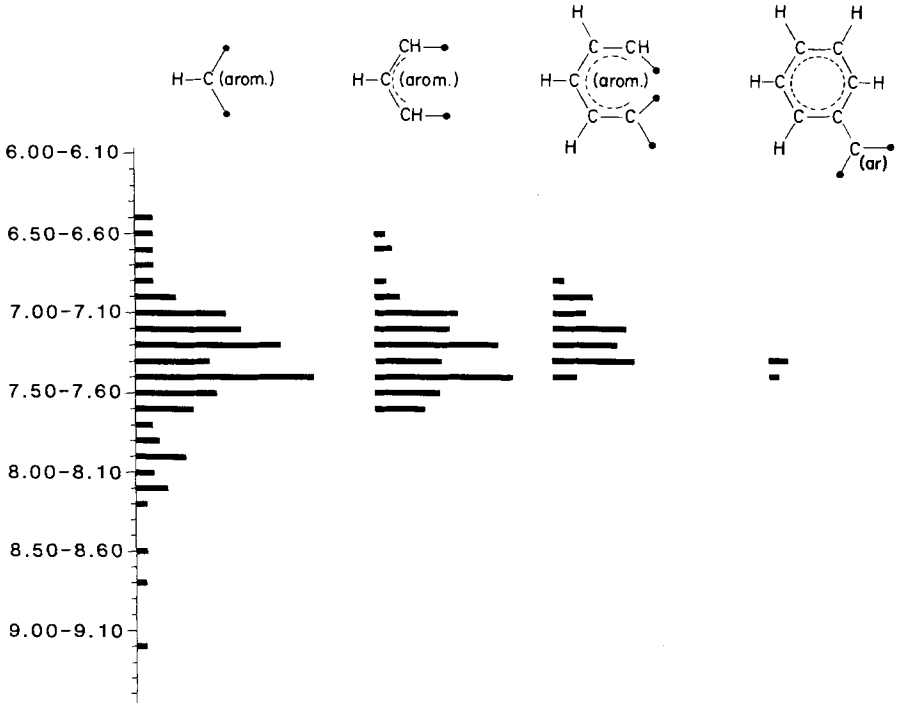
Scheme 5. *Display of diastereotopic protons at a double bond by* HCODE




```
WHICH *H(s) TO PROCESS? ZERO
Assigned Shift? Y
Shift value : 5.12
2nd *H SHIFT ASSIGNED? Y
2nd shift value : 5.26

-CH= 2 Assigned Shift? Y
Shift value : 5.98
```

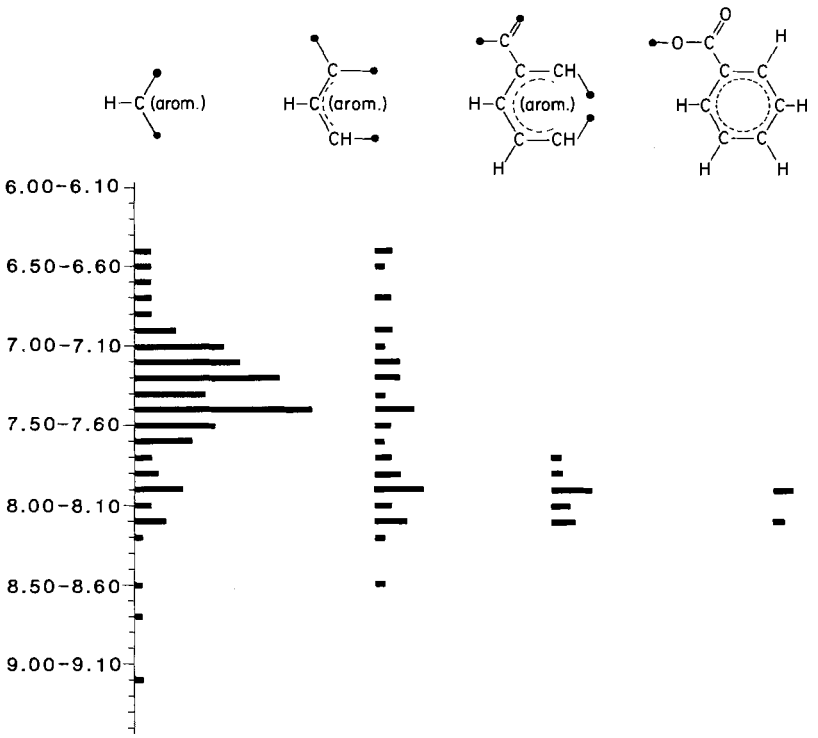
Scheme 6. *Extended correlation chart for aromatic protons* (Successively, more detailed substructural environments lead to successively narrower shift ranges)



built up to prevent errors. Auxiliary programs allow removal or correction of data found to be in error.

The HDCODE program can also be used to query the data base and retrieve information on specific substructures and chemical shifts, actually shift ranges,

Scheme 7. *Extended correlation chart for aromatic protons* (Incomplete specification of *o*-substituents leads to broad shift distribution at shell 2. As the substituent is characterized in more detail at shell-levels 3 and 4, the shift distributions become narrow)



associated with the substructures. These queries are useful as illustrations of the shell-level structure of the codes in the data base, which has important implications when the data base is used for spectrum prediction (next section). Histograms representing the distribution of shifts can be obtained, as illustrated in *Scheme 6* and *7* for aromatic protons in about 50 monosubstituted benzenes [16]. These histograms can be regarded as extended correlation charts, where the shift distributions become more narrow as more of the substructural environment (larger shell-levels) is included in the codes.

In *Scheme 6*, the histogram in column 1 represents the distribution of chemical shifts observed in the data base for shell-1 aromatic protons, *i.e.*, a proton connected to an aromatic C-atom. The second histogram obtained from HDBCHK displays the shell-2 results, *i.e.*, an aromatic proton attached to a C-atom that is in turn attached to two aromatic (>CH)-groups. With this more specific environment, the shift distribution is narrowed significantly. The third and fourth histograms represent successively more detailed specifications of environments at shell 3 and shell 4, respectively (the substructural environments are given in the column headings of *Scheme 6*). Each more detailed specification results in narrowing of the shift ranges found in the data base for substructures described at the given shell-level.

In *Scheme 7* a similar presentation is made, this time considering substructure representing *ortho*-substitution to the proton of interest. Here the first histogram is the same as in *Scheme 6;* the substructural environment is simply an aromatic proton attached to C-atom. At shell 2, however, there is only a slight decrease in the range of shifts observed. Although the substructural environment at shell 2 implies *ortho*-substitution, the environment is not yet specific enough to indicate the details of the substituent, and different substituents yield markedly different chemical shifts. This situation is resolved immediately at shell 3, where even the incomplete specification of a doubly-bonded C-atom is sufficient to narrow the shift ranges. At shell 4, the most complete environment coded in the data base, selection of carboxyl group yields a histogram with a quite narrow shift range.

In summary, the more specific the substructural description, the narrower the shift range found in the data base. Here configuration is an important element of the substructure descriptions. Protons that differ only in their configurational environment often display markedly different chemical shifts. Unless configuration is included, shift distributions tend to broaden excessively.

2.10. *Spectrum prediction and structure ranking.* – 2.10.1. *The* HNMRP *program.* Given a data base of substructures and associated chemical shifts, we can use this data base to *predict* the ^{1}H-NMR. spectra of new compounds using the HNMRP program ((5) in *Scheme 1*). In our scheme, these compounds are usually a set of structural candidates for an unknown compound obtained from another program *(Scheme 1),* or, alternatively, supplied by a chemist.

Our method finds in the data base substructural prototypes representative of substructures in a new structure, and retrieves the associated *shift range* observed for the substructure [10]. *Schemes 6* and *7* represent a good visual introduction to the procedure used. Each new structure is passed through the same coding scheme used in HCODE. This results in a set of codes for the substructural environment of every proton. The data base is searched ((5) in *Scheme 1*) for occurrences of the most complete description of an environment, *i.e.,* out to shell 4. If no entry is found in the data base, the shell structure imposed on the codes is used to try matching at successively smaller shell-levels until a match is found and the associated shift range retrieved. As indicated in the *Schemes 6* and *7*, if a match can be found at shell 3 or shell 4, a prediction of a narrow shift range results. At lower shell-levels, the shift ranges predicted are expected to broaden, but in any case, an expected shift range will be retrieved from the data base. This procedure is repeated by HNMRP for every structure given to the program. Each predicted spectrum consists of a set of expected *shift ranges* for the protons in the corresponding structure.

2.10.2. *The* HCORR *program.* It is the function of the correlation and ranking program HCORR to compare the predicted spectra with an observed spectrum ((6) in *Scheme 1*) in order to rank the structural candidates on the basis of agreement between the predicted spectra, sets of expected shift ranges, and the observed spectrum, a set of experimentally determined chemical shifts.

In a first step the number of given experimental shifts is compared to the number of shifts expected for a particular structure. This allows the elimination of structures that do not possess the minimum number of protons determined experimentally. For the remaining structures a correlation of observed and

predicted spectra is performed leading to the most probable matching of resonances for each predicted spectrum. A score reflecting the degree of matching with the experimental data is then calculated. Ranking of the structural candidates is then performed by simply ordering the calculated scores. The rankings can be used to eliminate implausible structures in the **STRCHK** program ((7) in *Scheme 1*). These procedures have been described in detail for applications to ^{13}C-NMR. data [10] [12].

An illustrative example. We illustrate the methods described above in the context of a simple example. We used the HCODE program to build a class-specific data base which included a series of differently substituted bicyclo[3.1.0]hexanes [17], bicyclo[3.2.1]octanes [18], and norcamphor [19]. A thujane derivative, 4β-*H*-3-methylthujan-3α-ol (**2**), *not included in the data base*, was used as an 'unknown' for the spectrum prediction and rank-ordering programs.

We begin this example by using the structure generator **GENOA** [13] to obtain a set of structural isomers that will represent our candidates for the 'unknown' **2**. Let us assume that a mass spectrum, eventually combined with an elemental analysis, yielded the molecular formula $C_{11}H_{20}O$. From a ^{1}H-NMR. spectrum we might learn that the molecule possesses an isopropyl group and two methyl groups. The typical proton shifts in cyclopropanes and the absence of olefinic protons lend support to the idea that a bicyclo[3.1.0]hexane skeleton is present. In addition to this fragmentary structural information an IR. spectrum could assure the presence of a COH-group. The constraints for a typical structure generation for this molecular formula with **GENOA** would then be as given in *Table 2.*

Table 2. *Constraints for structure generation of bicyclo[3.1.0]hexanes*

Isopropyl groups	exactly 1
Methyl groups	exactly 4
Hydroxyl groups	exactly 1
Bicyclo[3.1.0]hexane skeleton	exactly 1

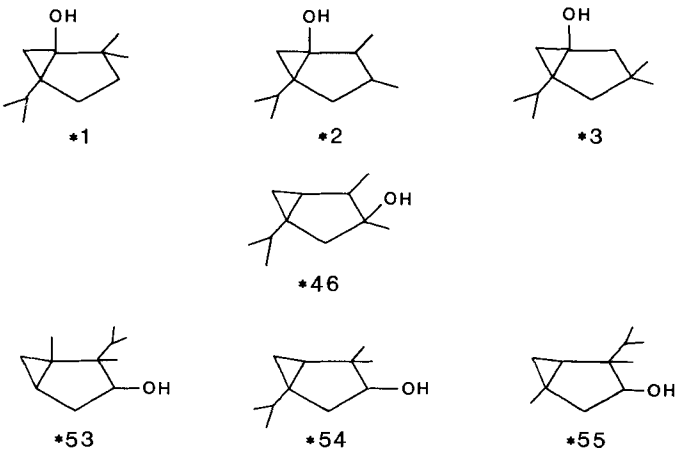
This structure generation resulted in 55 constitutional isomers. Some of them are shown in *Scheme 8.* The constitution of **2** in this test corresponds to isomer *46 (see *Scheme 8*). These structures are, of course, very similar, given the common skeleton, differing only in the location of substituents on the skeleton.

The program HNMRP was used to predict the proton chemical shifts for all these 55 structures, initially using just molecular constitution as a pre-screening of the candidates prior to prediction using the stereoisomers of the most plausible remaining candidates.

Following the prediction a correlation of the predicted and observed shifts was performed using HCORR. We present in *Table 3* a portion of the output of HCORR, showing the scores for the four scoring functions *SUMSQ, SBEL, SDIS2, SMBEL.* These functions have previously been described in detail [12].

SHELL gives the average shell-level on which the prediction is based and thus reflects the degree of matching with the data base. *SUMSQ* is proportional to the minimized sum of the squares of the differences between the observed and the

Scheme 8. *Selected constitutional isomers of 55 structures generated by* **GENOA** *with the constraints in* Table 2



corresponding predicted mean resonances. This scoring function does not reflect the quality of the prediction with respect to the size of the matched substructure. *SMBEL*, however, also takes into account the number of shell-levels matched for the prediction and is therefore a measure of belief that an observed spectrum matches the predicted resonances. *SDIS2* is a measure of disbelief that a predicted spectrum and an observed spectrum correspond to the same structure. A small *SDIS2* value means that the prediction is either based on poor models available in the data base, or that the errors between matched observed and predicted resonances are small. The last function *SMBEL* is the score when a slightly modified version of the *Mitchell-Schwenzer* [20] scoring function is employed.

As mentioned above, there are two steps to the scoring procedure. In the first step the best correspondence between predicted and observed resonances is sought. For this step *SDIS2* was used since this function has been shown to yield good results when applied to matching of ^{13}C-NMR. spectra [12]. The second step is computation of scores reflecting the degree of match. We have not performed the detailed evaluation of alternative scoring functions, applied to ^{1}H-NMR. spectra,

Table 3. *Results of scoring functions*

COMP	*SHELL*	*SUMSQ*	*SBEL*	*SDIS2*	*SMBEL*
*1	2.00	1.26	32.16	0.38	0.35
*2	2.18	2.89	44.82	0.65	0.31
*3	2.00	2.59	30.41	0.71	0.35
:	:	:	:	:	:
*46	4.54	0.38	345.87	0.22	0.92
:	:	:	:	:	:
*53	2.18	2.77	48.68	0.69	0.39
*54	2.54	11.38	91.10	1.82	0.37
*55	2.18	11.38	47.17	1.82	0.35

that was carried out for ^{13}C-NMR. spectra [12]. For that reason HCORR presents the results for each of the four functions *(Table 3)*.

Although the constitution of structure *46 is not included explicitly in the data base enough substructural elements seem to be present to allow a prediction on an extremely high average shell-level of 4.54. This, together with a very smooth correspondence of predicted and observed shifts results in a very high belief rate which outnumbers *SBEL* of all other 54 structures by a factor of 3 to 10 *(Table 3)*. For this reason structure *46 has to be considered, among other reasonably well-ranked structures, as a candidate structure which deserves further attention.

The final output of HCORR is the combined results of the scoring functions, a portion of which is presented in *Table 4*. It is the chemist's responsibility, based on these results, to select among the initial candidate structures those which merit further analysis. Structure *46 is only ranked fifth by this procedure *(Table 4)*, so that it and other highly ranked structures must still be considered as candidates.

Table 4. *Rank-ordering based on scores obtained (Table 3)*

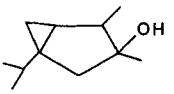
Stereoisomer number	Overall ranking	Based on rank-order in			
		SUMSQ	*SBEL*	*SDIS2*	*SMBEL*
*1	21=	16	38	16	31
*2	37=	33	28	23	52
*3	41=	23	47	31	41
:	:	:	:	:	:
*46	5=	13	1	15	1
:	:	:	:	:	:
*53	19=	29	23	26	17
*54	33=	48	3	48	26
*55	47=	48	25	48	43

With a reduced set of candidate structures a user can perform a spectrum prediction and correlation including configuration with the **STEREO** program in **STRCHK** ((4b) in *Fig. 1*). Thus, the next step would be to generate stereoisomers for the remaining candidates. Structure *46 is used to demonstrate this general procedure. The *cis*-oriented bicyclo[3.1.0]hexane of *46 was defined as a substructure in **STRCHK** and used as a constraint in the generation step of the stereoisomers for this structure because *trans*-ring junctures in such systems are energetically very unfavorable. The configurations at the five-membered ring for the four possible stereoisomers are shown in *Scheme 9*.

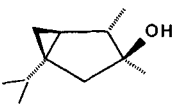
Table 5 shows the predicted resonances for the four stereoisomers where *RES* is the predicted chemical shift, followed by the atom number and shell-level of prediction in parentheses. In *Table 6* the results of the scoring functions and their rank-ordering is given. Based on these results stereoisomer *2 is ranked highest. This in fact is the correct stereochemistry for 4β-*H*-3-methylthujan-3α-ol (**2**).

3. Examples. – In this section we present two examples which we selected for the following reasons. Because it is not possible, with limited resources, to attempt to build a general data base of ^{1}H-NMR. spectra, we have focussed on applications where a significant number of structures and assigned spectra are available. Such

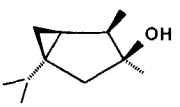
Scheme 9. *Configurational stereoisomers for the* cis-*oriented bicyclo[3.1.0]hexane of ring system of structure * 46* (The numbering of * 46 is that obtained from the programs)



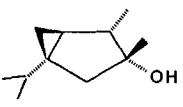
*46



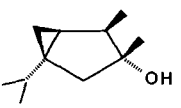
#1



#2



#3



#4

sets of data tend to be in restricted classes of compounds, and one of the advantages of our approach is that class-specific data bases can be built using our programs. For illustration of our methods we used a data base containing ^{1}H-NMR. spectra for about 110 sugars, about 50 mono-substituted benzenes, and about 50 substituted methanes [16].

3.1. *Stereoisomers of tetra-*O*-acetyl-1-fluoro-hexopyranose.* The first example was chosen to illustrate the capabilities of our programs to deal with configuration. Here we consider only the problem of stereoisomers of tetra-*O*-acetyl-1-fluoro-hexopyranoses (**3**) with the molecular formula $C_{14}H_{19}O_9F$.

Table 5. *Comparison of observed and predicted spectra, RES* (atom, shell), *for the 3-methyl-thujan-3-ol stereoisomers* 1–4

Line	Obs.	#1	#2	#3	#4
1	2.08	1.93 (7,2)	1.93 (7,2)	1.93 (7,2)	1.92 (12,2)
2	1.88	1.93 (7,2)	1.93 (7,2)	1.93 (7,2)	1.83 (7,4)
3	1.79	1.92 (12.2)	1.92 (12.2)	1.92 (12,2)	1.81 (10,5)
4	1.30	1.35 (5,3)	1.34 (5,4)	1.35 (5,3)	1.59 (7,4)
5	1.17	1.10 (3,3)	1.14 (6,4)	1.10 (3,3)	1.30 (3,5)
6	1.00	1.07 (9,2)	1.10 (3,3)	1.07 (9,2)	1.27 (5,5)
7	1.00	0.93 (1,4)	1.01 (9,3)	0.93 (1,4)	1.09 (6,5)
8	0.93	0.92 (4,3)	0.92 (4,3)	0.92 (4,3)	0.92 (4,5)
9	0.91	0.90 (2,4)	0.92 (1,5)	0.90 (2,4)	0.92 (1,5)
10	0.82	0.49 (6,3)	0.87 (2,5)	0.49 (6,3)	0.87 (2,5)
11	0.18	0.45 (6,3)	0.55 (6,4)	0.45 (6,3)	0.37 (6,5)
12		2.25 (10,2)	2.55 (10,2)	2.25 (10,2)	0.82 (9,4)

Table 6. *Results of scoring functions for stereoisomers * 1–* 4*

COMP	*SHELL*	*SUMSQ*	*SBEL*	*SDIS2*	*SMBEL*
Results of scoring functions					
*1	2.81	0.24	105.98	0.08	0.35
*2	3.36	0.19	210.69	0.08	0.52
*3	2.81	0.24	105.98	0.08	0.35
*4	4.54	0.24	369.72	0.13	0.93
Rank-ordering of scoring functions					
*1	2=	2	3	2	3
*2	1=	1	2	1	2
*3	2=	2	3	2	3
*4	2=	4	1	4	1

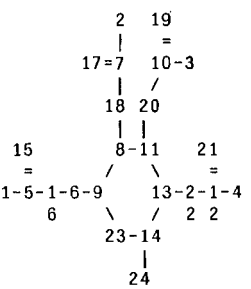
The **STEREO** program *(Fig. 1)* was used to generate the 32 possible stereoisomers according to the five stereocenters at the sugar ring. We chose as the 'unknown' spectrum that of tetra-*O*-acetyl-1-fluoro-α-D-glucopyranose (**4**) for which the proton shifts were partly assigned [21].

In *Scheme 10* we present a numbered drawing of **3**, as drawn by **GENOA**. The numbering assigned by the program is not related to the conventional numbering according to the IUPAC nomenclature, but this same numbering is used subsequently for the drawings *(Scheme 11)* of some of the stereoisomers.

We present in *Table 7* some of the results of spectrum prediction using the HNMRP program. Each row comprising a set of predicted spectra is preceded by the observed spectrum so that the matching of predicted to observed resonances by HCORR can be seen. Each predicted resonance, *RES*, is accompanied by the atom number to which the proton is attached (see *Scheme 10*) and the shell-level at which a matching substructure was found in the data base. In *Table 7*, pairs of enantiomers are grouped together. Thus, for example, stereoisomers *1 and *2, which are enantiomers, display the same predicted spectrum. Obviously, it is difficult to tell from this presentation which predicted spectrum matches most closely the observed

Scheme 10. *Constitutional formula of tetra-*O*-acetyl-1-fluoro-hexopyranose generated by* **GENOA**

```
NON C ATOMS ARE: 15->O; 16->O; 17->O; 18->O; 19->O; 20->O; 21->O; 22->O;
                 23->O; 24->F;
NO STEREOISOMER CONFIGURATIONS SPECIFIED
```



Scheme 11. *Display of some of the 32 stereoisomers of tetra-O-acetyl-1-fluoro-hexapyranose* (**3**) *generated by* **STEREO** (The numbering is the same as given in *Scheme 10*)

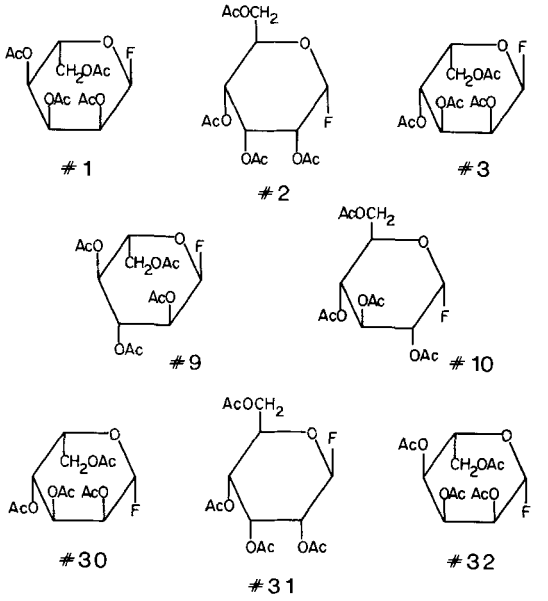


Table 7. *Comparison of observed and predicted spectra, RES* (atom, shell), *for the 12 of the 32 stereoisomers of tetra-O-acetyl-1-fluoro-hexopyranose* (**3**)

Line	Obs.	#1	#2	#3	#4	#5	#6
1	5.74	5.69 (14,3)	5.69 (14,3)	5.69 (14,3)	5.69 (14,3)	5.69 (14,3)	5.69 (14,3)
2	5.48	5.40 (8,3)	5.40 (8,3)	5.46 (8,3)	5.46 (8,3)	5.53 (8,3)	5.53 (8,3)
3	5.14	5.36 (13,3)	5.36 (13,3)	5.43 (11,3)	5.43 (11,3)	5.36 (13,3)	5.36 (13,3)
4	4.96	5.35 (11,3)	5.35 (11,3)	5.36 (13,3)	5.36 (13,3)	5.35 (11,3)	5.35 (11,3)
5	4.28	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)
6	4.18	4.17 (6,4)	4.17 (6,4)	4.41 (6,3)	4.41 (6,3)	4.41 (6,3)	4.41 (6,3)
7	4.12	4.07 (9,3)	4.07 (9,3)	4.33 (9,3)	4.33 (9,3)	4.33 (9,3)	4.33 (9,3)
8	2.10	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)
9	2.10	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)
10	2.02	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)
11	2.02	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)
Lines	Obs.	#7	#8	#9	#10	#31	#32
1	5.74	5.69 (14,3)	5.69 (14,3)	5.73 (14,5)	5.73 (14,5)	5.40 (8,3)	5.40 (8,3)
2	5.48	5.43 (11,3)	5.43 (11,3)	5.46 (11,3)	5.46 (11,3)	5.35 (11,3)	5.35 (11,3)
3	5.14	5.40 (8,3)	5.40 (8,3)	5.40 (8,3)	5.40 (8,3)	5.34 (13,3)	5.34 (13,3)
4	4.96	5.36 (13,3)	5.36 (13,3)	5.20 (13,3)	5.20 (13,3)	5.30 (14,3)	5.30 (14,3)
5	4.28	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)	4.42 (6,2)
6	4.18	4.17 (6,4)	4.17 (6,4)	4.38 (9,4)	4.38 (9,4)	4.17 (6,4)	4.17 (6,4)
7	4.12	4.06 (9,4)	4.06 (9,4)	4.17 (6,4)	4.17 (6,4)	4.07 (9,3)	4.07 (9,3)
8	2.10	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)	1.83 (1,3)
9	2.10	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)	1.80 (2,5)
10	2.02	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)	1.80 (3,5)
11	2.02	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)	1.80 (4,5)

Table 8. *Results of scoring functions for the predicted ^{1}H-NMR. spectra of the 32 stereoisomers of tetra-O-acetyl-1-fluoro-hexopyranose* (**3**)

COMP	*SHELL*	*SUMSQ*	*SBEL*	*SDIS2*	*SMBEL*
* 1	3.54	0.49	127.87	0.22	0.50
* 2	3.54	0.49	127.87	0.22	0.50
* 3	3.45	0.62	102.13	0.27	0.47
* 4	3.45	0.62	102.13	0.27	0.47
* 5	3.45	0.58	101.92	0.26	0.47
* 6	3.45	0.58	101.92	0.26	0.47
* 7	3.63	0.51	139.83	0.23	0.53
* 8	3.63	0.51	139.83	0.23	0.53
* 9	3.81	0.44	185.33	0.21	0.61
*10	3.81	0.44	185.33	0.21	0.61
*11	3.54	0.52	114.15	0.24	0.50
*12	3.54	0.52	114.15	0.24	0.50
*13	3.45	0.53	103.35	0.24	0.47
*14	3.45	0.53	103.35	0.24	0.47
*15	3.54	0.41	129.71	0.19	0.50
*16	3.54	0.41	129.71	0.19	0.50
*17	3.54	0.56	123.24	0.25	0.50
*18	3.54	0.56	123.24	0.25	0.50
*19	3.45	0.63	98.70	0.28	0.47
*20	3.45	0.63	98.70	0.28	0.47
*21	3.45	0.60	98.45	0.26	0.47
*22	3.45	0.60	98.45	0.26	0.47
*23	3.81	0.46	147.18	0.23	0.59
*24	3.81	0.46	147.18	0.23	0.59
*25	3.63	0.57	119.44	0.26	0.53
*26	3.63	0.57	119.44	0.26	0.53
*27	3.45	0.59	97.02	0.26	0.47
*28	3.45	0.59	97.02	0.26	0.47
*29	3.45	0.61	98.25	0.27	0.47
*30	3.45	0.61	98.25	0.27	0.47
*31	3.54	0.57	122.44	0.25	0.50
*32	3.54	0.57	122.44	0.25	0.50

one. The agreement between predicted and observed spectra is generally quite good, with only seemingly minor discrepancies in going from one diastereomer to the next.

In *Table 8* we present the results obtained by HCORR for the stereoisomers in the form of scores for each of the four key-scoring functions. Again, enantiomers are grouped together and, of course, receive the same score. Every stereoisomer possessed substructures that were well-represented in the data base, as evidenced by the high *SHELL* values (column two in *Table 8*).

The final output of HCORR is the rank-ordering based on analysis of the results of the scoring functions, shown in *Table 9*. The enantiomers *9 and *10 are clearly favored by the overall-ranking, and in fact correspond to the structure and configuration of the 'unknown' structure **4**.

3.2. *Epimeric pentofuranoses.* Our programs for ^{1}H-NMR. analysis can be used in a variety of ways in addition to the general scheme outlined in *Scheme 1*. For example, we have already discussed methods for accessing the data base to retrieve

Table 9. *Rank-ordering of scoring functions for the 32 stereoisomers of tetra-O-acetyl-1-fluoro-hexopyranose* (**3**)

Stereoisomer number	Overall ranking	Based on rank-order in			
		SUMSQ	*SBEL*	*SDIS2*	*SMBEL*
* 1	9=	7	9	5	11
* 2	9=	7	9	5	11
* 3	23=	29	21	29	23
* 4	23=	29	21	29	23
* 5	21=	21	23	19	21
* 6	21=	21	23	19	21
* 7	7=	9	5	9	5
* 8	7=	9	5	9	5
* 9	1=	3	1	3	1
*10	1=	3	1	3	1
*11	13=	11	17	11	17
*12	13=	11	17	11	17
*13	17=	13	19	13	19
*14	17=	13	19	13	19
*15	3=	1	7	1	9
*16	3=	1	7	1	9
*17	11=	15	11	15	13
*18	11=	15	11	15	13
*19	31=	31	25	31	31
*20	31=	31	25	31	31
*21	23=	25	27	25	25
*22	23=	25	27	25	25
*23	3=	5	3	7	3
*24	3=	5	3	7	3
*25	17=	19	15	23	7
*26	17=	19	15	23	7
*27	27=	23	31	21	29
*28	27=	23	31	21	29
*29	29=	27	29	27	27
*30	29=	27	29	27	27
*31	15=	17	13	17	15
*32	15=	17	13	17	15

shift distributions for various substructures. In some structural problems a chemist may already have deduced the constitution of an 'unknown', and perhaps also several details of configuration. The remaining problem may be differentiation among a small set of remaining stereoisomers. In this case, there is no need to invoke the structure generating programs **GENOA** and **STEREO**. The chemist merely defines for the program the molecular constitution and configuration of the remaining structures and analyzes them with HNMRP and HCORR.

For this example we will focus on differentiating furanose epimers. For hexopyranoses and pentopyranoses lone-pair effects relating to the orientation of the free electron pairs at the ring O-atom support in many cases [22] [23] the determination of the ring conformation. For furanoses such lone-pair effects do not seem to possess a comparable discriminatory influence on neighboring protons. The

Table 10. *Comparison of predicted ^1H-chemical shifts for tetra-O-acetyl-α-D-xylofuranose* (**5**, = *1) *and tetra-O-acetyl-β-D-xylofuranose* (**6**, = *2) *to the observed spectrum of tetra-O-acetyl-α-D-xylofuranose* (**5**) [24]

Line	Obs.	*1	*2
1	6.44	6.43 (2,3)	6.28 (2,4)
2	5.54	5.65 (4,3)	5.65 (4,3)
3	5.32	5.27 (3,3)	5.46 (3,3)
4	4.64	4.93 (5,3)	4.84 (5,3)
5	4.24	4.42 (6,2)	4.42 (6,2)
6	4.13	4.41 (6,3)	4.41 (6,3)
7		2.16 (9,3)	2.16 (9,3)
8		2.24 (15,4)	2.40 (13,5)
9		2.40 (18,5)	2.40 (17,5)
10		2.40 (21,5)	2.24 (21,4)

determination of the configuration at the anomeric center in furanoses represents a substantial challenge to our programs, and also points out some limitations for applications involving subtle configurational differences.

We have selected tetra-*O*-acetyl-α-D-xylofuranose (**5**) and tetra-*O*-acetyl-β-D-xylofuranose (**6**) for this example. Because the structures were defined manually, the substituents received different numberings, as shown in structures **5** and **6**. The respective, partially assigned spectra were obtained from [24].

In *Tables 10* and *12* the predicted ^{1}H-NMR. shifts are listed together with the observed shifts for the two compounds. Subsequent rank-orderings are given in *Tables 11* and *13*.

For ranking based on the observed spectrum of structure **5**, the results in *Table 11* must be regarded as inconclusive. Although the β-epimer is slightly favored, the differences in scoring functions do not allow a definite conclusion. For ranking based on the observed spectrum of structure **6**, however, one can conclude that the predicted spectrum of stereoisomer *2, corresponding to the configuration of **6**, is a better match to that of **6**, the correct result.

The inability of our programs to perform such subtle configurational distinctions without occasional ambiguity is related to limitations of our method, as summarized in the conclusions.

Table 11. *Results of scoring functions and rank-ordering for comparison of predicted ^1H-spectra for tetra-O-acetyl-α-D-xylofuranose* (**5**, = *1) *and tetra-O-acetyl-β-D-xylofuranose* (**6**, = *2) *with the observed assigned shifts of tetra-O-acetyl-α-D-xylofuranose* (**5**) [24]

COMP	*SHELL*	*SUMSQ*	*SBEL*	*SDIS2*	*SMBEL*
Results of scoring functions					
*1	2.83	0.20	44.02	0.07	0.13
*2	3.00	0.20	45.32	0.07	0.16
Rank ordering of scoring functions					
*1	2=	2	2	1	2
*2	1=	1	1	2	1

Table 12. *Comparison of predicted ^1H-chemical shifts for tetra-O-acetyl-α-D-xylofuranose* (**5**, = *1) *and tetra-O-acetyl-β-D-xylofuranose* (**6**, = *2) *to the observed spectrum of tetra-O-acetyl-β-D-xylofuranose* (**6**) [24]

Line	Obs.	*1	*2
1	6.12	6.43 (2,3)	6.28 (2,4)
2	5.39	5.65 (4,3)	5.65 (4,3)
3	5.22	5.27 (3,3)	5.46 (3,3)
4	4.67	4.93 (5,3)	4.84 (5,3)
5	4.27	4.42 (6,2)	4.42 (6,2)
6	4.27	4.41 (6,3)	4.41 (6,3)
7		2.16 (9,3)	2.16 (9,3)
8		2.24 (15,4)	2.40 (13,5)
9		2.40 (18,5)	2.40 (17,5)
10		2.40 (21,5)	2.24 (21,4)

4. Conclusions. – The described programs can predict accurately individual shifts according to the information collected in the data base. The examples demonstrate that despite overlapping shift ranges our encoding system is generally adequate to the problem and that the programs, considering substructures together with configuration, are well-suited for spectrum prediction. In many structure elucidation problems involving structural types covered by the data base our methods can result in a substantial reduction of the number of candidate structures originally considered.

At the current state of development of these programs, we cannot expect that the correct structure will always be top-ranked. We do expect that many structural candidates that provide poor explanations of an observed proton spectrum can be eliminated.

Our methods have significant limitations. The lack of a large, general-purpose data base is one limitation. This will restrict the application of programs such as our own to more specific classes of compounds for which extensive proton data can be brought together from existing compilations. A second limitation is in the representation of structures and the substructures derived therefrom. Although the programs represent configurational aspects in a general way, there is no treatment of conformational aspects included in the programs described in this contribution.

Table 13. *Results of scoring functions and rank-ordering for comparison of predicted ^1H-spectra for tetra-O-acetyl-α-D-xylofuranose* (**5**, = *1) *and tetra-O-acetyl-β-D-xylofuranose* (**6**, = *2) *with the observed assigned shifts of tetra-O-acetyl-β-D-xylofuranose* (**6**) [24]

COMP	*SHELL*	*SUMSQ*	*SBEL*	*SDIS2*	*SMBEL*
Results of scoring functions					
*1	2.83	0.27	37.89	0.10	0.13
*2	3.00	0.22	43.95	0.08	0.16
Rank ordering of scoring functions					
*1	2=	2	2	2	2
*2	1=	1	1	1	1

This limitation can be overcome to an extent by restricting applications to classes of compounds where configurations restrict accessible conformations, *e.g.*, at ring junctures of edge-fused and bridged ring systems. But until conformational information is made part of the coding scheme, valuable information on the influence of conformation on chemical shifts and coupling constants cannot be captured and used for spectrum prediction.

We are currently developing a revised coding scheme that takes into account conformational aspects, a scheme derived from previous work on computer representation and manipulation of conformations [25]. There are two problems here. The first is that such conformational information is seldom available in sufficient detail in the literature. Where it is available, it seems to be obvious that the introduction of conformational descriptions will make the predictions even more accurate. However, it must be pointed out that a more precise description of substructures goes hand in hand with a lower probability of finding a matching substructural environment in the data base.

We have attempted to reduce the effect of these limitations by designing our programs so that structural problems can be analyzed in a stepwise fashion. First, prediction and ranking can be made on the basis of molecular constitution, leading to a smaller number of plausible candidates. Those that remain are reanalyzed, now using configurational aspects for finer discriminations. The addition of conformational aspects will allow a third step that should provide a better distinction between the correct structure and the remaining candidates.

We wish to thank the *National Institutes of Health* for their generous financial support. This work was supported in part by a grant from the *Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung* (to *H.E.*). Computer resources were provided by the SUMEX facility at Stanford University under *National Institutes of Health*. *H.E.* also wishes to express his gratitude to the members of the DENDRAL group for the friendly atmosphere and support. Special thanks to *Neil A.B. Gray* and *Chris W. Crandell* for providing programs on which this work is based.

Experimental Part

These programs are implemented in the ALGOL-like BCPL programming language on a *Digital Equipment Corporation KI-10* computer at the SUMEX-AIM facility at Stanford University. The programs are available to an outside community of users, by request to the authors, *via* an international computer network. The experimental shift data were collected from sources using only CCl_4, $CDCl_3$, (D_6)acetone, (D_6)benzene as solvents and TMS as reference standard. Spectra recorded using as solvents D_2O, (D_5)pyridine, (D_6)DMSO should be treated individually in order to prevent inaccurate shift predictions caused by broadened shift ranges. The actual data base is also limited to examples which were measured at a temperature range of 20 to 40° and were not subject to experiments with shift reagents.

REFERENCES

[1] *M. R. Lindley, N. A. B. Gray, D. H. Smith & C. Djerassi,* J. Org. Chem. (in press), 1982.
[2] *H. Skolnik,* J. Chem. Soc. *10,* 216 (1970).
[3] *V. Mlynarik, M. Vida & V. Kello,* Anal. Chim. Acta, Vol. *122,* 47 (1980).
[4] *F. Erni & J. T. Clerc,* Helv. Chim. Acta *55,* 489 (1972).
[5] *M. H. Jacobs & L. van Derslice,* Appl. Spectrosc. *26,* 218 (1972).
[6] *S. R. Heller & R. J. Feldman,* J. Chem. Educ. *49,* 291 (1972).
[7] *G. Beech, R. T. Jones & K. Miller,* Anal. Chem. *46,* 714 (1974).
[8] *N. A. B. Gray, R. E. Carhart, A. Lavanchy, D. H. Smith, T. Varkony, B. G. Buchanan, W. C. White & L. Creary,* Anal. Chem. *52,* 1095 (1980).
[9] *N. A. B. Gray, A. Buchs, D. H. Smith & C. Djerassi,* Helv. Chim. Acta *64,* 458 (1981).
[10] *N. A. B. Gray, C. W. Crandell, J. G. Nourse, D. H. Smith, M. L. Dadgeford & C. Djerassi,* J. Org. Chem. *46,* 703 (1981).
[11] *N. A. B. Gray, J. G. Nourse, C. W. Crandell, D. H. Smith & C. Djerassi,* Org. Magn. Reson. *15,* 375 (1981).
[12] *C. W. Crandell, N. A. B. Gray & D. H. Smith,* J. Chem. Comp. Sci., in press.
[13] *R. E. Carhart, D. H. Smith, N. A. B. Gray, J. G. Nourse & C. Djerassi,* J. Org. Chem. *46,* 1708 (1981).
[14] *J. G. Nourse,* J. Am. Chem. Soc. *101,* 1210 (1979).
[15] *J. G. Nourse, R. E. Carhart, D. H. Smith & C. Djerassi,* J. Am. Chem. Soc. *101,* 1216 (1979).
[16] *W. Bruegel,* 'Handbook of NMR. Spectral Parameters', Heyden & Son Ltd., GmbH, 4440 Rheine, West Germany, Vol. 1–3, 1973.
[17] *J. C. Rees & D. Whittaker,* Org. Magn. Reson. *15,* 363 (1981).
[18] *B. V. Crist,* PhD dissertation, University of Nevada, Reno 1981.
[19] *J. L. Marshall & S. R. Walter,* J. Am. Chem. Soc. *96,* 6358 (1974).
[20] *T. M. Mitchell & G. M. Schwenzer,* Org. Magn. Reson. *11,* 378 (1978).
[21] *L. D. Hall, J. F. Manville & N. S. Bhacca,* Can. J. Chem. *47,* 1 (1969).
[22] *K. Bock & C. Pedersen,* J. Chem. Soc., Perkin II, Vol. 1974, 293 (1974).
[23] *K. Bock & C. Pedersen,* Acta Chem. Scand. B *29,* 258 (1975).
[24] *J.-P. Utille & Ph. Vottero,* Bull. Soc. Chim. Fr. *1976,* 1101 (1976).
[25] *J. G. Nourse,* J. Chem. Inf. Comput. Sci. *21,* 168 (1981).