Computer-Assisted Structure Elucidation Using Automatically Acquired ^{13}C NMR Rules.
Gretchen M. Schwenzer, Tom M. Mitchell,
1977

5

Computer-Assisted Structure Elucidation Using Automatically Acquired ^{13}C NMR Rules

GRETCHEN M. SCHWENZER and TOM M. MITCHELL

Department of Computer Science, Stanford University, Stanford, CA 94305

Carbon-13 nuclear magnetic resonance (CMR) has developed into an important tool for the structural chemist. A CMR spectrum exhibits a wide range of shifts which have been shown to have a strong correlation with structure(1,2). A natural abundance CMR spectrum which is fully proton decoupled consists of a number of sharp peaks which correspond to the resonance frequencies in an applied magnetic field of the various types of carbon atoms present. A C-13 shift is the amount an observed peak is shifted from that of a reference peak, usually tetramethylsilane (TMS).

Molecular structure elucidation using CMR consists of establishing a set of rules which summarize the CMR behavior for a set of compounds and then using the rules to identify unknown compounds. In the traditional approach to structure elucidation using CMR the chemist forms a set of empirical rules by sorting through a large amount of data looking for correlations between structural arrangements in the molecules and the observed C-13 shift. The total shift is then given as a function of these structural parameters. The functional form is usually chosen to be a linear combination of independent parameters. The optimized value of the coefficient of each structural parameter is obtained by a curve fitting procedure. This approach leaves the decisions of the structural parameter selection and the selection of a reasonable functional form to the chemist. In both cases the correct decisions are easily overlooked. The best known example of this approach is that of Lindeman and Adams(3).

Although the rule form resulting from the parameter approach is useful for predicting spectra of a given structure it cannot be used efficiently for

Reprinted with permission from Schwenzer, Gretchen M. and Mitchell, Tom M. in "Computer-Assisted Structure Elucidation," Dennis H. Smith, Ed., in ACS SYMPOSIUM SERIES, No.54; American Chemical Society: Washington, D.C., 1977, pp.58-76.

structure elucidation. Observing a shift does not give any information about what substructure is present without constructing a set of candidate structures, calculating their shifts and then comparing the calculated shifts with the observed spectrum. This procedure was followed in the work of Carhart and Djerassi (4) which used the parameter set obtained by Eggert and Djerassi for the acyclic amines (5). However, this method does not reflect the actual thinking procedure a chemist would use when identifying an unknown and it suffers from its inability to be generalized. The selection of empirical rules and the algorithms for the generation of substructures are specific for each class of compounds. To the chemist it is more likely that the appearance of an observed shift or set of shifts would suggest a structural arrangement.

A second approach to structure elucidation is to establish a data base of C-13 NMR spectra and then use the data base to do structure elucidation by searching for peak patterns similar to those in the unknown spectrum(6,7,8). A data base which consists of the entire spectrum results in storage of a large amount of redundant information. Structure elucidation is accomplished by supplying the chemist with the molecules having portions of their spectra which are similar to that of the unknown. The chemist then uses these to suggest partial substructures that are present in the unknown. This method has suffered from the necessity of obtaining large data bases and methods of assembling the partial substructures.

We offer an alternative to these approaches with the following procedure. First rule formation is accomplished by a computer program which forms a set of empirical rules by associating an observed total shift with a substructural arrangement in the molecule. A second program uses these rules for structure elucidation by assembling substructural fragments suggested by the rules.

A sample rule constructed by the program from a set of paraffins and acyclic amines is

$$CH_3^{*}\text{-}CH_2\text{-}CH_2\text{-}CH_2\text{-} \quad \text{---> } 14.0\text{ppm} < \delta(*) < 14.7\text{ppm}.$$

The asterisk denotes the atom for which the shift is predicted. The prediction $\delta(*)$ is given in ppm downfield from TMS. The important substructural arrangements given in the rule are actually constructed by the program from a language of features supplied by the user. Molecular structure elucidation is accomplished

by observing a total shift and finding the rules which are possible explanations for the shift. The rules selected postulate partial substructures which might be in the molecule. These substructures are assembled to construct the final molecule. A description of the rule formation and structure elucidation programs applied to the paraffins and acyclic amines is given in the following sections. We believe the algorithms used are general enough to treat widely different classes of compounds. Rules generated for decalins, methyldecalins and hydroxy-steroids are shown in the third section.

Empirical Rule Formation

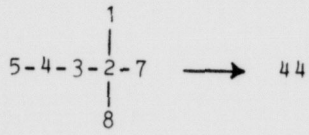
Rule Generation. The rule generation program(9) must be supplied a training set of known structures with their assigned spectra. A set of primitive terms which will form the language of atom features used to describe the atoms and bonds (atom type, number of non-hydrogen neighbors, orientation of substituents, etc.)must also be supplied. These terms are combined to construct structural fragments which imply a total shift. The chemist also sets two parameters which regulate the generality of the rules generated. MINIMUM-EXAMPLES is a parameter which specifies the minimum number of data points which a rule must explain within the training set. The other parameter, MAXIMUM-RANGE, specifies the maximum allowable shift range for a rule. If the chemist wants only the most general trends in the data he can require a larger number of examples with moderately sized shift ranges.

The format of the rules generated is

substructure $\xrightarrow{\text{implies}}$ ^{13}C shift range.

If the substructure to the left of the arrow is present within some molecule then there is a shift within the range given to the right of the arrow. The rule shown in Table I was generated on a combined set of acyclic amines and paraffins.

Table I Rule Form



Node	Atom Type	Number of non-hydrogen Neighbors
1,5,7,8	C	≥1
2	C	4
3	C	2
4	C	2

For the substructure shown in Table I with the corresponding atom features for atom type and number of non-hydrogen neighbors atom number 3 will have a C-13 shift in the range 44.7ppm to 44.9ppm downfield from TMS.

The rule search procedure is shown in Figure 1. The search begins with the general seed rule C->$-\infty \leq \delta_c \leq \infty$ (where C may be any carbon atom with any atom properties and δ_c is the observed shift) and proceeds to expand this rule by adding new atoms and atom features to the substructure which will narrow the predicted range of shifts. The seed rule in Figure 1 is expanded by considering all possible values of "number of neighbors" of the central carbon. Each resulting level 1 substructure is expanded in level 2 by adding either an "atom type" or "number of neighbors" specification to each atom one bond away from the central carbon. At each step only a single atom feature from the user supplied list is added. Each substructure generated is associated with a range of C-13 shifts. This range is determined by searching for occurrences of the substructure within the training set molecules. The shift range associated with the substructure is the range of all occurrences of the substructure in the training set.

Each substructure generated in the rule search is evaluated in terms of the associated shift range. If the shift range is narrower than the range of the parent rule then the added specification is considered to be useful and the search continues from the new substructure, otherwise the path is terminated. The

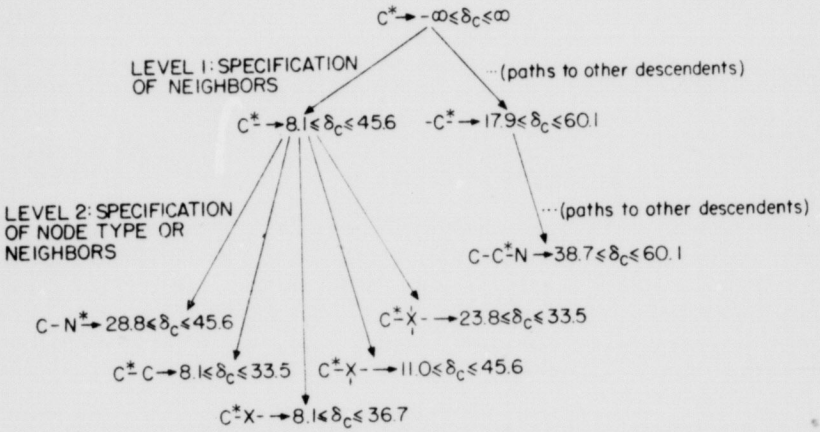


Figure 1. Partial schematic of the rule search. δ_C values are approximate and are given in ppm downfield from TMS. (*) identifies the carbon atom to which the shift is assigned; (X) indicates any non-hydrogen atom.

program runs until all branches of the search have been explored. For a substructure to be accepted as a final rule it must satisfy the conditions of MINIMUM-EXAMPLES and MAXIMUM-RANGE.

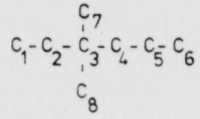
Since the rule search procedure can result in slightly different substructures being generated which cover the same data in the training set, the generated rule set may be redundant. A less redundant set of rules is chosen by assigning a score to each rule. The score is peaks/range^2 where peaks is the number of data peaks covered by the rule in the training data, and range is the width of the shift range for the rule. The rule with the highest score is selected and all the data points explained by that rule are removed. If unexplained data points remain, the score of the remaining rules are reevaluated and the procedure repeated. The intent is to select the strongest rule during each iteration and weaken the rules with evidence which overlap with it. The result is a subset of the strongest rules covering the same data as the original rule set

The algorithm used to generate the C-13 NMR rules is similar to the algorithm in the Meta-DENDRAL program which generates empirical rules of molecular fragmentation from mass spectral data(10,11).

For a set of 22 paraffins and 47 acyclic amines with a total of 435 data peaks a set of 138 rules were generated with the parameter settings MINIMUM-EXAMPLES=2 and MAXIMUM-RANGE=2.0 ppm.

Structure Selection. To test the information content of the rules and thus their value for use in structure elucidation, a structure selection test was designed. A program was written which uses the rules to predict spectra for a set of candidate molecules and compares the predicted spectra to that of an unknown. The candidates are then ranked according to a "best" match criteria to the unknown spectrum.

A spectrum is predicted by applying the rules to a molecule, searching for places the rule substructure fits into the molecule. This is done with a graph matching routine. When a match is found the shift range associated with the rule is predicted for the associated carbon atom. Figure 2 is a example of predicting a spectrum. Each rule shown has a substructure with corresponding atom features that maps into the molecule. Often more than one rule applies to a given atom to give different predictions. If the predicted ranges are consistent(i.e. one of the predicted ranges is contained in the others) the



OBSERVED SPECTRUM

8.7 15.4 17.9 27.1 33.4 34.9 44.9

PREDICTED SPECTRUM

(8.1 8.7) (14.0 15.4) (17.9 24.3) (27.1 29.7) (29.7 35.6) (31.2 34.9) (44.7 44.9)

CARBON ATOM	RULES WHICH APPLY TO CARBON ATOM				
	SUBSTRUCTURE	ATOM NODE	ATOM TYPE	NUMBER OF NON-H NEIGHBORS	PREDICTION
C_4	1 5-4-3-2-7 8	1,5,7,8	C	≥1	$44.7 \leq \delta(3) \leq 44.9$
		2	C	4	
		3	C	2	
		4	C	2	
	-4-3-2-	2	C	4	$41.1 \leq \delta(3) \leq 45.1$
		3	C	2	
		4	C	2	
	1-2-3	1,3	C	≥1	$17.9 \leq \delta(2) \leq 56.9$
		2	C	2	
C_3	3 8-2-1 7	1,3,7,8	C	≥1	$29.7 \leq \delta(2) \leq 35.6$
		2	C	4	

Figure 2. Partial spectrum predictions for 3,3-dimethylhexane. δ(n) is the shift for atom n in ppm downfield from TMS.

narrowest predicted range is used. This is illustrated by the rules which apply to C4 in Figure 2. The predicted range 44.7 to 44.9 is contained in the other two rule predictions thus it is selected as the prediction. This method can be rationalized since the actual shift should fall into all of the predicted ranges and thus into the narrowest. If the predicted ranges overlap incompletely or are disjoint then the ranges are merged to arrive at a final predicted range for the carbon atom.

The predicted spectrum is compared to an unknown spectrum by assigning each atom's predicted range to the closest observed shift in the unknown spectrum. In order to be a valid assignment the number of carbons in the structure less the number of observed shifts must be greater than or equal to the number of multiply assigned shifts. If the assignment does not satisfy this criterion the required number of multiply assigned observed shifts are reassigned. The atoms which are chosen to be reassigned are those whose reassignment will cause the smallest change in the comparison score assigned to the match of the predicted spectrum to the unknown spectrum.

Results. A set of rules has been generated using a subset of the paraffin data from Lindeman and Adams (3) and a subset of the acyclic amine data from Eggert and Djerassi (5). Molecules with the empirical formula C_9H_{20} and $C_6H_{15}N$ were excluded from the training set. The structure selection test was performed by generating all structural isomers with the empirical formulas C_9H_{20} (35 isomers) and $C_6H_{15}N$ (39 isomers). For each of the candidate isomers a spectrum was predicted. There were 24 C_9H_{20} spectra available from the work of Lindeman and Adams. The 35 predicted spectra were compared and ranked against each of these available spectra. The results of this ranking for C_9H_{20} as well as a similar test on $C_6H_{15}N$ are shown in Table II.

Table II. Results of Structure Selection

Empirical Formula	Number of Candidates	Number of Structures Ranking 1st	2nd	...6th	..9th
C_9H_{20}	35	20	3		1
		--	-		-
		24	24		24
$C_6H_{15}N$	39	8	2	1	
		-	-	-	
		11	11	11	

Peak intensity information which gives the number of carbon atoms corresponding to an observed peak was not used. The use of this information would have resulted in the correct assignment for those which were poorly ranked.

The form of the rule which is proposed has several advantages. Each rule has its own predicted error and the rule set consists of rules of varying detail. In the parameter set approach to rule formation the error is the standard deviation of the training set data from the fitted curve. When analyzing carbon atoms that exhibit magnetic nonequivalence (resulting in different chemical shifts for two identical groups in molecules having an asymmetric carbon atom) the parameter set approach which attempts to predict the arithmetic mean will always be in error. The advantage in predicting total shifts over hypothesizing partial contributions is in avoiding initial biases as to what contributes to the shift and how these contributions are to be combined. Our program bypasses the bias of assumed functional form and introduces only a weak bias concerning which structural features may be considered. The program can consider any structure which can be constructed from the atom feature language supplied by the chemist. Another advantage of the rule format is that it can be read backwards, that is, the appearance of a peak in an unknown spectrum implies a structural feature. This property is what distinguishes the efficiency of this rule form over other forms when used for structure elucidation.

The method of structure selection described above

is inefficient when there are a large number of structural isomers. Instead of applying this test to all structural isomers for a given empirical formula we wish to select a subset of likely candidates to be put through this final ranking procedure. The ability to read the rules backwards to do molecular structure elucidation will enable us to achieve this goal.

Structure Elucidation

Structure Search. The information the chemist must supply to the structure elucidation program includes the empirical formula of the unknown and the observed spectrum. Two parameters must also be set by the chemist. The first parameter is the number of plausible structures which should be found to be ranked by the structure selection procedure described earlier. The second parameter is the error range in ppm which should be assigned to the rules to account for deficiencies in the training set, experimental error, solvent effects, etc.

In order to make the generated rules more useful for structure elucidation additional information is added to the rules. The form of the rule used for structure elucidation is shown in Table III. The rule now says the observation of a shift implies a partial substructure. This is the same rule as shown in Table I with the addition of two new properties, support prediction and secondary prediction.

Table III Rule Form for Structure Elucidation

$$44.7\text{ppm} \leq \delta(3) \leq 44.9\text{ppm} \text{ ----> } 5\text{-}4\text{-}3\text{-}\overset{1}{\underset{8}{2}}\text{-}7$$

Node	Atom Type	Number of non-H Neighbors	Support Prediction	Secondary Prediction
1	C	≥ 1		$27.1 \leq \delta(1) \leq 34.9$
2	C	4	$29.7 \leq \delta(2) \leq 35.6$	$30.7 \leq \delta(2) \leq 33.4$
3	C	2		
4	C	2	$17.9 \leq \delta(4) \leq 56.9$	$17.9 \leq \delta(4) \leq 27.6$
5	C	≥ 1		$15.4 \leq \delta(5) \leq 24.3$
7	C	≥ 1		$27.1 \leq \delta(7) \leq 34.9$
8	C	≥ 1		$27.1 \leq \delta(8) \leq 34.9$

Once a set of rules has been obtained from the rule generation program, support predictions can be made. For each rule an attempt is made to find support predictions for all atoms in the rule substructure except the atom for which the rule is defined. For each rule's substructure the rules are applied to it to obtain predictions. Predictions made by the rule set for any atom in the substructure are tabulated. The final support prediction for a particular atom is obtained by merging the tabulated predictions for that atom. Support predictions are really main predictions merged into the rule. This is done to introduce additional constraints upon the selection of the rules early in the search procedure. In the rule shown in Table III there were other rules which gave predictions for nodes 2 and 4.

Although support predictions cannot be obtained for all atoms in the rule substructure, predicted ranges are associated with all atoms using the following procedure. Secondary predictions are made by finding all places the rule substructure applies in the training set data and tabulating the observed shifts for the atoms in the substructure. For each atom in the substructure the observed shifts found in the training set are merged to form the secondary predictions. The reliability of a secondary prediction is highly dependent on the variety of molecular structures in the training set. Secondary predictions for atoms which are fewer bonds away from the rule's predicted atom will have a smaller expected error than those more bonds away. The shift ranges of the secondary predictions are broadened by a predefined constant to account for the deficiencies in the training set. The addition of support and secondary predictions to the original rules will form the new rule set.

The first step in the rule search is to select a subset of the rule set which will act as possible explanations for the observed spectrum. The rules which have shift ranges consistent with the shifts in the observed spectrum are selected. Each rule selected is checked for agreement with the empirical formula of the unknown molecule and for the presence of observed peaks in the spectrum for all support predictions in the rule. The constraint that the number of carbons in the structure less the number of observed shifts must be greater than or equal to the number of multiply assigned shifts must also be satisfied. For the rule to be a valid possible explanation there must be an assignment of the main and support predictions which

does not violate this constraint. The subset of rules selected by this procedure from the set of partial substructure hypotheses for the observed spectrum.

The structure search procedure is shown in terms of a specific example in Figure 3. The observed spectrum is for a molecule with the empirical formula C_8H_{18}. The number of rules which are possible explanations for the observed shifts are shown. The observed spectrum in Figure 3 corresponding to 3,3-dimethylhexane was also shown in Figure 2. In Figure 2 only the rules which were correct explanations for the molecule are shown. For example there was one correct explanation for C3 explaining the observed shift 33.4. In Figure 3 there are eight possible explanations for 33.4 of which only one is correct for this particular spectrum.

An approximate lower bound to the number of possible ways the partial substructures may be assembled is the product of the number of explanations given for the observed shifts. Although in this example this lower bound is greater than 1000, the constraints imposed by the observed spectrum, empirical formula and rule set directed the search to the correct structure after considering only 12 paths through the search tree.

The search strategy shown in Figure 3 is that of a depth first tree search. Solutions exist at unknown locations in the tree. A set of heuristics or judgmental rules are chosen to guide the search to the final structure in the most efficient manner.

The subset of rules which are possible explanations for the observed spectrum are ordered to select the rule which has the greatest chance of being in the molecule and which will lead most rapidly to the final molecule. The heuristics which order the rules include the quality of the main and support predictions. The quality of a prediction means the width of the prediction range and the closenesss of the observed peak to the rule prediction range. The number of peaks explained in the training set by the rule is also a factor in ordering the rules. An analogy can be drawn between this heuristic and the selection by the chemist of a substructure which has frequently been observed to be present when a particular shift occurs in a spectrum. Another heuristic is the number of other explanations a particular rule has. If there is only one explanation for an observed shift it is very likely that the rule's substructure will be in the unknown molecule. These conditions could be considered

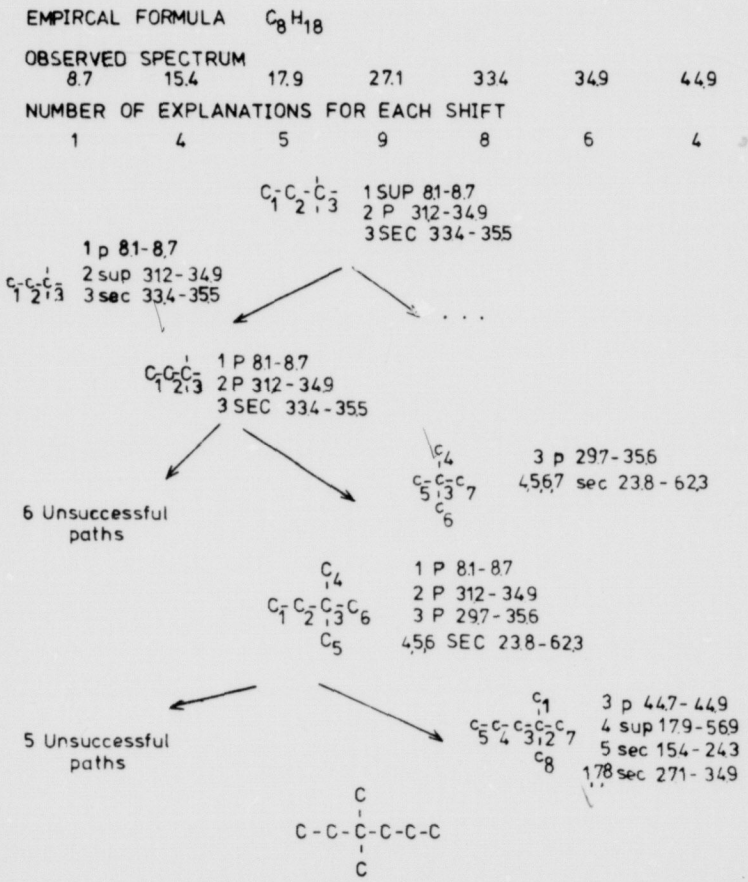


Figure 3. Structure search. Main (P), support (SUP), and secondary (SEC) predictions are given in ppm downfield from TMS. The candidate structures are shown in capitals; the rules selected to build on the candidate structure are shown in lower case.

as contributions to the certainty that the rule substructure is in the final molecule.

The final heuristic which orders the subset of rule explanations will depend on the width of the secondary and support predictions. Instead of being a certainty measure of the rule substructure's presence it will be a measure of efficiency of building with this particular substructure. The neighbors of the rule's predicted atom are examined as potential building sites. The subset of neighbors which have more than one non-hydrogen neighbor will be the atoms on which construction will take place. The support and secondary predictions of these atoms suggest a subset of rules from those selected as explanations for the observed spectrum. This subset will form explanations for the shifts of these atoms. The narrowness of the support or secondary prediction determines the number of rules which are possible explanations. The rule which has the narrowest prediction for a neighboring atom will lead most rapidly to the final structure or the elimination of the rule as a possible explanation.

The combination of these heuristics will order the set of rule explanations, hopefully placing those which are the correct explanations with the most promising structures to build on early in the list. The rule chosen which satisfies these conditions for the example is shown in Figure 3. The substructure C1-C2-C3- has a support prediction for C1, a main prediction for C2 and a secondary prediction for C3.

Once a rule has been selected it is necessary to choose an atom in the rule to build on. Neighbors of atoms with main predictions and which do not have main predictions themselves will be the atoms which are possible candidates for building. First the support predictions for these atoms are considered. If any of the support predictions are sufficiently narrow that atom will be chosen as the construction site otherwise the secondary predictions are examined. The atom with the narrowest secondary prediction range is chosen. For the example in Figure 3 atom C1 was chosen for the site of construction. Rules are selected from the subset of rule explanations which have shift ranges consistent with the secondary or support prediction of the atom chosen. The set of rules selected is ordered with the most likely rule first using the criterion stated previously. The example in Figure 3 chose C1 to build on and the rule explanation for C1 is shown.

Two rules have now been chosen. The process of selection of the second rule results in an intial mapping between the two rule substructures. The atom

from the first rule which was selected as the construction site maps to the atom with the main prediction in the second rule. It is now necessary to find all substructures consistent with the initial mapping that can result from the overlap of the rules substructures. A graph matching routine is used to find the overlaped substructures. The properties checked in overlapping are those which make up the language of atom features. For the paraffins and acyclic amines properties checked are atom type and number of non-hydrogen neighbors. If no possible overlap exists a new rule explanation is tried, if there are no possible explanations for the selected atom the original rule chosen is discarded and the next choice is tried. Any candidate structures which result from the overlap must be consistent with the empirical formula of the unknown.

The overlap process must also merge the predictions for the two rules. An atom from one rule which has been mapped into an atom in the other rule must have a resultant prediction consistent with both predictions from the original rules. For instance the merging of two support predictions results in a prediction range which is the intersection of the two predictions.For the candidate substructure to be valid there must be an observed shift which is consistent with the new main and support predictions. In addition there must be an assignment of observed shifts to main and support predictions which satisfies the constraint of the number of multiply assigned shifts. In Figure 3 the overlapping of the first two rules does not result in any change in the substructure but the predictions of the two rules are merged.

The list of candidate structures which survive these tests have the same form as the original rules. The decision as to the likelihood of these candidate structures being in the molecule can be made on the basis of the candidate structure itself without combining the likelihood of the parent rules. The candidate structures are ranked on the basis of the quality of the main and support predictions.

The new subproblem is now identical to the original problem and the steps of selecting an atom to build on, selecting rule explanations for that atom, overlapping the two substructures, merging predictions, and the ranking of the resultant candidates are repeated. The procedure terminates when the required number of plausible molecules are found. Figure 3 shows the paths examined until the first plausible structure was found which was the correct solution for

the observed spectrum. An unsuccessful path in Figure 3 means that the overlapping of a rule with the candidate structure failed on structural grounds or the candidate structures generated were not consistent with the observed spectrum. If a rule fails another rule is tried until a candidate structure is generated which is consistent with the constraints of the observed spectrum and empirical formula of the unknown molecule.

Results. One possible search procedure has been explained whose goal is to obtain a few plausible structures which can then be ranked by the structure selection procedure described earlier. Other possible search strategies are being considered whose goal would be to obtain substantial sized substructures which would then be used as starting points in a structure generating program such as CONGEN(12).

Handling Stereochemistry

The work on the paraffins and acyclic amines requires only topological descriptors in the language of atom features. Because of the dependence of C-13 shifts on stereochemical features (13,14) it is necessary to have the facility to include stereochemical terms when they are required. Substituents placed on systems which have static conformations such as trans decalin and androstane with trans ring fusions can be described in discrete terms. The terms we selected describe the orientation on the ring of the substituent as either axial or equatorial, and either alpha or beta. A substituent is beta in 10-methyl-trans-decalin if it is on the same side of the ring as the methyl group and alpha if on the opposite side of the ring from the methyl group. The rule generation program with the extension of the language to include these atom features was run on a combined set of trans decalins, 10-methyl-trans-decalols and monohydroxylated androstanes with trans ring fusions selected from the works of Grover and Stothers (13) and Eggert et. al.(14). Sixty rules were generated to cover the 249 data peaks of 17 compounds. Samples of the rules generated are shown in Figure 4. The examination of these rules will show that they are useful for the chemist who wants to study contributions to the total shift as well as for the structure elucidation procedure we have outlined.

In Figure 4 are shown two pairs of rules for alpha carbons. Within each pair of alpha carbon rules there is little shift difference between the axial or

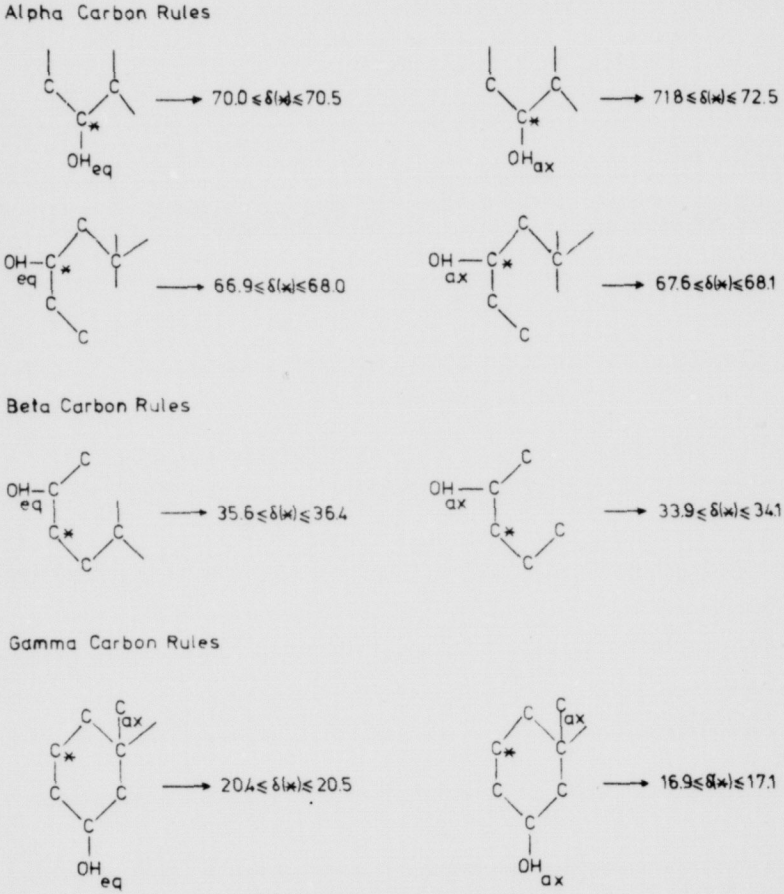


Figure 4. Sample rules constructed from decalins and hydroxy steroids with trans ring fusions. (*) identifies the carbon atom to which the shift is assigned; δ(*) is in ppm downfield from TMS.

equatorial orientation of the hydroxyl substituent. However, there is a difference between the pairs which could reflect the importance of the number of gamma carbons. The pair of beta carbon rules show the equatorial substitution results in a shift at lower fields than the axial substitution. The gamma rules shown illustrate an axial substituted hydroxy group will result in a gamma carbon shift at higher fields than that of an equatorial substituted hydroxy group.

Conclusion

Computer programs have been written to generate empirical C-13 NMR rules and to do structure elucidation using these rules. Rules generated on a set of paraffins and acyclic amines have successfully identified the C-13 NMR spectra of molecules not in the training set data. The form of the rule is suited for efficient structure elucidation using an algorithm which assembles substructures suggested by the rules as explanations of the observed shifts. The introduction of a limited set of stereochemical terms to the rule generation procedure demonstrated the feasibility of extending the method to more complicated systems.

Acknowledgements. This work was supported by the National Institutes of Health under grants 5R24 00612-07 and AM-17896-02 and by the Advanced Research Projects Agency under grant DAHC 15-73-C-0435.

Computer resources were provided by the SUMEX facility at Stanford University under National Institutes of Health grant RR-00785.

We are grateful to Jim McDonald, Bruce Buchanan, and Carl Djerassi for helpful discussions and William C. White for providing parts of the Meta-DENDRAL program code for use in this work.

Literature Cited

1. Stothers, J.B., "Carbon-13 NMR Spectroscopy," Academic Press, New York, N.Y. 1972.
2. Levy, G.C. and G.L. Nelson, "Carbon-13 Nuclear Magnetic Resonance for Organic Chemists," Wiley-Interscience, New York, N.Y. 1972.
3. Lindeman, L.P. and J.Q. Adams, Anal. Chem., (1971), 43,p. 1245.
4. Carhart,R. and C. Djerassi, J.Chem. Soc., Perkin Trans., (1973),2,p. 1753.
5. Eggert, H. and C. Djerassi, J. Amer. Chem. Soc. (1973),95,p. 3710.

6. Bremser, W., M. Kiler, and E. Meyer, Org. Magn. Resonance, (1975),7,p. 97.
7. Jezl, B.A. and D.L. Dalrymple, Anal. Chem. (1975), 47,p. 203.
8. Schwarzenbach, J., J. Meili, H. Konitzer, J.T. Clerc, Org. Magn. Resonance,(1976),8,p. 11.
9. Mitchell, T.M. and G.M. Schwenzer, to be published
10. Buchanan, B.G., D.H. Smith, W.C. White, R.J. Gritter, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, J. Amer. Chem. Soc., (1976),98, p. 6168.
11. Buchanan, B.G., T.M. Mitchell, Proceedings of the Workshop on Pattern Directed Inference, Honolulu, Hawaii, (1977).
12. Carhart, R., D. Smith, H. Brown, and C. Djerassi, J. Amer. Chem. Soc., (1975),97,p. 5755.
13. Grover, S.H. and J.B. Stothers, Can. J. Chem. (1974),52,p. 870.
14. Eggert, H., C. VanAntwerp, N. Bhacca, and C. Djerassi, J. Org. Chem.,(1976),41,p. 71.