

Applications of Artificial Intelligence for Chemical Inference. Part XI.¹ Analysis of Carbon-13 Nuclear Magnetic Resonance Data for Structure Elucidation of Acyclic Amines

By Raymond E. Carhart and Carl Djerassi,* Departments of Computer Science and Chemistry, Stanford University, Stanford, California 94305, U.S.A.

This paper describes a computer program entitled AMINE, which uses a set of predictive rules to deduce the structures of acyclic amines from their empirical formulae and ¹³C n.m.r. spectra. The results of testing the program on 102 amines indicate that AMINE is quite accurate and selective, even for large amines with many millions of structural isomers, and demonstrate that the computerized analysis of ¹³C n.m.r. data can be a powerful analytical tool. The logical structure of the program is outlined here, including a section on the general problem of spectrum matching. Generalizations of the methods used by AMINE are suggested.

In recent years, there has been a substantial amount of research directed toward the computerized identification of molecular structure from mass,²⁻⁵ n.m.r.,^{4,6,7} and i.r. spectroscopic⁷ data. Our heuristic DENDRAL program,²⁻⁴ which relies primarily upon mass spectral data, has been shown to be quite accurate for certain classes of saturated, acyclic, monofunctional compounds, and more recently, the methods have been extended to the estrogenic steroids.³ There are limitations to the information content of mass spectral data, however, particularly when compounds are considered which have long, perhaps highly branched alkyl chains. An analysis of the mass spectrum of triheptylamine, for example, yields *ca.* 2×10^3 solution structures,⁴ and although this is only a small fraction of the roughly 4×10^7 (non-stereochemical) isomers of C₂₁H₄₅N, it is still an impractically large number. The problem is that alkyl groups do not give characteristic fragmentation patterns, and in fact, most spectroscopic methods are relatively insensitive to their structure.

However, recent studies indicate that ¹³C n.m.r. spectroscopy⁸ is an exception. For several classes of compounds,⁹ rules have been obtained which allow one to predict the ¹³C n.m.r. spectrum of a substance from its molecular structure, and in all cases, the rules indicate that the chemical shift of any carbon atoms, even one in a large alkyl chain-end, depends heavily upon branching at nearby centres. Thus, it appears that ¹³C n.m.r. spectroscopy, either alone or in combination with other methods, could be a powerful tool in

† Copies of the program, along with sample input decks, may be obtained from the authors.

‡ These isomer counts were computed using the enumeration algorithm of ref. 10.

¹ Part X, D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, *Tetrahedron*, in the press.

² B. G. Buchanan, A. M. Duffield, and A. V. Robertson in 'Mass Spectroscopy: Techniques and Applications,' ed. G. W. A. Milne, Wiley, New York, 1971, p. 121.

³ D. H. Smith, B. G. Buchanan, R. S. Engelmores, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, *J. Amer. Chem. Soc.*, 1972, **94**, 5962, and previous papers in the series.

⁴ A. Buchs, A. M. Duffield, G. Schroll, C. Djerassi, A. B. Delfino, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, *J. Amer. Chem. Soc.*, 1970, **92**, 6831.

⁵ (a) H. S. Herze, R. A. Hites, and K. B. Biemann, *Analyt. Chem.*, 1971, **43**, 681; (b) L. R. Crawford and D. J. Morrison, *ibid.*, p. 1790; (c) D. H. Smith, *ibid.*, 1972, **44**, 536; (d) P. C. Jurs, *ibid.*, 1971, **43**, 1812, and references cited therein.

the computerized analysis of molecular structure. This paper outlines the methods by which such an analysis may be carried out for the acyclic amines, and describes a FORTRAN IV computer program,† entitled AMINE, in which these methods are implemented.

This class of compounds was chosen for two reasons. First, the recent work of Eggert and Djerassi^{9a} has yielded a detailed set of predictive rules for the acyclic amines. Secondly, for a given number of carbon atoms, a saturated, acyclic amine has decidedly more structural possibilities than most other simple types of acyclic organic compounds (*e.g.*, stereochemistry aside, there are nearly 1.5×10^7 C₂₀ amines, but only about 6×10^6 C₂₀ alcohols),‡ and thus the structural analysis of amines represents a particularly challenging problem.

I. *Definition of the Problem.*—A fully proton-decoupled, natural-abundance ¹³C n.m.r. spectrum⁸ typically consists of a number of sharp peaks representing the resonance frequencies, in the applied magnetic field, of the various types of carbon atoms present in the sample. A standard compound, commonly tetramethylsilane (TMS), is usually included in the sample to provide a reference frequency, and the peak positions, or chemical shifts, are measured as fractional deviations from this reference, in p.p.m. Previous investigations have shown that the shift of a particular carbon atom is determined by its hybridization and local environment, and thus each shift contains some structural information.

⁶ (a) S.-I. Sasaki, Y. Kudo, S. Ochiai, and H. Abe, *Mikrochimica Acta*, 1971, 726; (b) S.-I. Sasaki, S. Ochiai, Y. Hirota, and Y. Kudo, *Japan Analyst*, 1972, **21**, 916.

⁷ H. Abe and S.-I. Sasaki, *The Science Reports of the Tohoku University, Series I*, 1972, **55**, 63.

⁸ For a general discussion of ¹³C n.m.r. spectroscopy, see G. C. Levy and G. L. Nelson, 'Carbon-13 Nuclear Magnetic Resonance for Organic Chemists,' Wiley-Interscience, New York, 1972.

⁹ (a) H. Eggert and C. Djerassi, *J. Amer. Chem. Soc.*, 1973, **95**, 3710; (b) J. D. Roberts, F. J. Weigert, J. I. Kroschwitz, and H. J. Reich, *ibid.*, 1970, **92**, 1338; (c) D. M. Grant and E. G. Paul, *ibid.*, 1964, **86**, 2984; (d) L. P. Lindeman and J. Q. Adams, *Analyt. Chem.*, 1971, **43**, 1245; (e) D. E. Dorman, M. Jautelat, and J. D. Roberts, *J. Org. Chem.*, 1971, **36**, 2757; (f) D. K. Dalling and D. M. Grant, *J. Amer. Chem. Soc.*, 1967, **89**, 6612; (g) M. Cristl, H. J. Reich, and J. D. Roberts, *ibid.*, 1971, **93**, 3463; (h) D. E. Dorman, S. J. Angyal, and J. D. Roberts, *ibid.*, 1970, **92**, 1351; (i) J. K. Crandall and S. A. Sojka, *ibid.*, 1972, **94**, 5084; (j) W. R. Woolfenden and D. M. Grant, *ibid.*, 1966, **88**, 1496; (k) F. J. Weigert and J. D. Roberts, *ibid.*, 1970, **92**, 1347, and references cited therein.

¹⁰ H. R. Henze and C. M. Blair, *J. Amer. Chem. Soc.*, 1931, **53**, 3042, 3077.

There are a few ranges of shifts which are characteristic of certain functional groups, such as C=O or C=C, but aliphatic carbon atoms in most molecules lie in a broad spectral region from which detailed structural information cannot be extracted readily.

For acyclic amines,^{9a} and a few other types of compounds,^{9b-k} there exist predictive rules which allow one to calculate the spectrum of a compound whose structure is known, with a typical accuracy of *ca.* 1–2 p.p.m. in a total range of roughly 100 p.p.m. For these classes, the structure-identification problem could in principle be solved *via* the generation of all possible structures of a particular type (say, acyclic amines with a particular number of carbon atoms), the prediction of their spectra, and the comparison of these predictions with the observed spectrum. In fact, Sasaki *et al.*^{6b} have used this procedure in the automated identification of a few small alkanes. For large molecules, though, the number of possible isomers can be overwhelming, and even a very efficient computer program could not carry out such an analysis in a reasonable length of time.

The program AMINE is designed to accomplish the same goal, but in a much more efficient manner. It takes, as its only input data, an observed ¹³C n.m.r. spectrum, the number of carbon atoms in the amine, and a goodness-of-fit criterion. The observed spectrum consists of a list of shifts, $o = (o_1, \dots, o_n)$, measured in p.p.m. relative to TMS. Each of these corresponds to one or more carbon atoms in the sample molecule. Under favourable circumstances,¹¹ it is possible to determine the number of carbon atoms corresponding to each observed shift (this will be called the *tally* of the shift) once the relative peak intensities and the empirical formula are known. If the tally of a shift is known to be at least 2, 3, *etc.*, then the shift is entered in duplicate, triplicate, *etc.*, in the observed shift list. These tallies are not necessary to the program's operation, but even if they are underestimated, they can add considerably to the speed and accuracy of the analysis. The number of carbon atoms, N_C , in the amine must be greater than, or equal to, the number of shifts in the observed spectrum. Generally, N_C cannot be determined from the ¹³C n.m.r. spectrum, but must be obtained from some other analytical method such as mass spectroscopy or elemental analysis. The goodness-of-fit criterion, δ , which is used in the comparison of o with the predicted spectra of molecules or molecular fragments, represents the maximum expected error in the predictive rules. The amine rules are derived, in part, from the alkane rules of Lindeman and Adams,^{9d} who note that 95% of the studied alkanes have predicted shifts within 1.5 p.p.m. of the observed values. A similar situation exists for the amines, so a value of $\delta = 1.5$ p.p.m. has been used in most of this work.

The goal of the program is to find all acyclic, N_C -carbon amines whose predicted spectra satisfy the following two criteria: (a) every predicted peak must lie within δ of one of the observed peaks; and (b) within this limit, the

predicted shifts must be assignable to the observed ones in such a way that all of the latter are accounted for.

II. *Overview of Program Operation.*—The operation of program AMINE can best be viewed in terms of four interconnected processes; structure generation, pruning, filtering, and spectrum matching. The structure generator builds a pool of increasingly large and complex alkyl chain-ends, and eventually uses these to construct amine molecules. It relies heavily upon the pruner to cull from the growing pool any chains which are inconsistent with the observed spectrum, and similarly upon the filter to test entire amine molecules. The filter also takes care of outputting the acceptable solution structures, and ranking them according to how well they fit the observations. Both the filter and the pruner use the spectrum matcher, which is responsible for the actual comparison of predicted and observed spectra. Each of these processes will be discussed in detail.

III. *Structure Generation.*—The structure generation scheme used in this study, which is related to the enumeration algorithm of Henze and Blair,¹⁰ is applicable only to saturated, acyclic, monofunctional compounds. It is an efficient approach from the standpoint of ¹³C n.m.r. structural analysis because it rapidly generates substructures which contain a relatively large number of 'predictable' carbon atoms (*i.e.* those near the ends of alkyl chains), and thus many of these substructures may be ruled out early in the analysis as being inconsistent with the observed data.

At any point in the generation, the structure generator contains a pool of monovalent alkyl radicals which, through pruning (see below) have been found to be consistent with the observed ¹³C n.m.r. spectrum. The pool initially contains only the CH₃ radical. By attaching one or more of these pool members (along with an appropriate number of hydrogen atoms) to a central carbon atom, it constructs new radicals, each of which is passed to the pruner for testing. Any that agree with the observed spectrum are included in the pool, and are subsequently used to construct larger chains. In the final step of the analysis, the structure generator similarly attaches alkyl groups to a central nitrogen atom, constructing amine molecules of the proper empirical formula. These it passes to the filter for final testing and ranking. At all stages of the generation, tests are made which insure that no radical or amine is considered twice.

As will be discussed below, a given alkyl radical actually undergoes several different tests during pruning, with each test corresponding to a distinct chemical environment in which the chain-end might exist. The structure generator keeps a record of these tests for each pool member, and constantly checks that it is using the radicals in a consistent fashion. If, for example, the pruner finds that the ethyl group is consistent with the observed spectrum only if it is attached to a nitrogen atom in a secondary amine, the structure generator will never construct a Prⁿ or Bu^s group, or any other radical which contains an ethyl group con-

¹¹ G. N. La Mar, *J. Amer. Chem. Soc.*, 1971, **93**, 1040.

nected to carbon. Neither will it generate primary or tertiary amines with *N*-ethyl groups.

IV. *Pruning*.—The pruner is the real heart of program AMINE. It is responsible for keeping the growing chain-end pool to a manageable size by weeding out alkyl radicals which are inconsistent with the observed spectrum. In testing a particular chain-end, R, shown schematically in Figure 1, the basic question considered

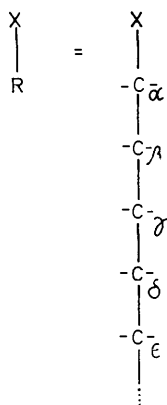


FIGURE 1 A schematic illustration of R, the alkyl chain-end to be tested by the pruner. The group X contains the nitrogen atom, along with any carbon and hydrogen atoms not included in R

by the pruner is, 'Of all possible sets of ^{13}C n.m.r. shifts which R could produce, is at least one consistent with the observed spectrum?' Actually, the question is somewhat more complex, but this provides a good starting point.

Now, according to the predictive rules,^{9a} a carbon shift is determined by the structure which surrounds it, up to four bonds away. Further, the effect of a first-row atom which is four bonds removed does not depend upon whether that atom is carbon or nitrogen. Thus, because X in Figure 1 must contain at least one such atom (namely nitrogen), the shifts of C_δ and any carbon atoms 'below' it are completely predictable and independent of the internal structure of X. The shifts of the remaining carbon atoms, C_γ , C_β , and C_α , depend to varying degrees upon the structure of X, with C_α being the most sensitive.

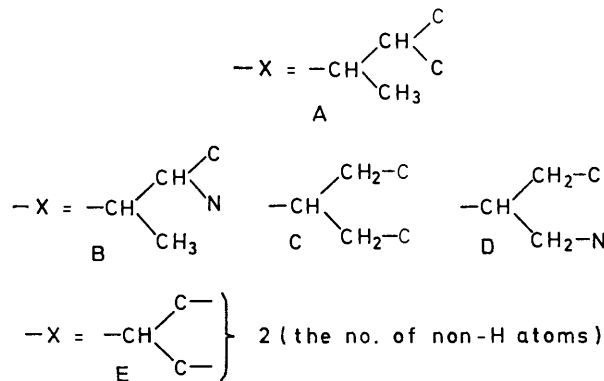
By investigating all possible X structures to a 'depth' of four atoms (measured from the R-X bond), the pruner could generate an exhaustive list of spectra that R might produce, testing each for inconsistency with the observed one. Usually, though, X contains enough atoms that there are several hundred of these 'depth-4' structures, and the above approach proves to be rather cumbersome. Instead, the pruner considers only 'depth-3' expansions of X, for which C_β and all carbon atoms below it are predictable. The shift of C_α is simply ignored, even when a reasonable estimate of its value might be made. This simplification cuts the number of unique X substructures to, at most, 94.

There are two factors which can reduce this number still further. First, some of the substructures may contain too many (or few) carbon atoms to be consistent

with the known atom-count of X. Secondly, there are many cases in which a single predicted spectrum for R may result from two or more related X structures. This situation arises because, according to the predictive rules,^{9a} the shifts of certain types of carbon atoms (specifically, those which are four or more bonds from nitrogen, or three if the degree of the amine is known) are not sensitive to the type or distribution of first-row atoms which are four bonds away, but only to their number. Thus, in the computation of the shift of C_β , the structure A is equivalent to three other structures B—D. All four may be considered as a single entity, which can be represented as E.

Once such a grouping of X substructures has been done, there remain, at most, 69 cases for the pruner to consider. These are summarized in Table 1, where they are further grouped into fifteen classes according to (a) the type of atom directly attached to R, (b) the degree of that atom and (c) if that atom is carbon, and is attached to nitrogen, the degree of the amine. The actual purpose of the pruner is to consider each of these classes, determining whether at least one class member gives R a predicted spectrum consistent with the observed one, and to return the results of the fifteen class-tests to the structure generator.

The efficiency of this class-by-class investigation can be greatly improved by the inclusion of a hierarchy of pre-tests, each of which is aimed at excluding one or more classes at once. For example, classes 1—12 in Table 1 all have one common feature; the atom to which



R is attached is carbon. Thus, as a pre-test for all twelve classes, the pruner treats X as a carbon atom whose neighbours are unknown (schematically, $\text{X} = \text{C}-?$) and predicts as much of the spectrum of R as possible. If these predictions do not match the observations, it bypasses all further consideration of classes 1—12 and proceeds with the $\text{X} = \text{N}-?$ pre-test for classes 13—15. Otherwise, it considers a number of more detailed pre-tests, each corresponding to a possible set of neighbours to the central carbon atom in $\text{X} = \text{C}-?$. The actual hierarchy is outlined in Figure 2. In each of the pre-tests, the local environment of either C_β or C is known to a depth of only three atoms, and hence the corresponding shift cannot be predicted precisely. In most of these cases, the pruner can derive upper and

lower limits for the shift from the predictive rules. These limits, which define an *estimated shift*, encompass a

TABLE I

Substructures X considered by the pruner

Class 'Depth 3' X structure(s)	Class 'Depth 3' X structure(s)
1 R-CH ₂ -C- } 1,2,3	12 R-C- N-C C- } 0,1,...,6
2 R-CH ₂ -NH ₂	13 R-NH ₂
3 R-CH ₂ -NH-C	14 R-NH-C- } 0,1,2,3
4 R-CH ₂ -N- C	
5 R-CH- C- C- } 1,2,...,6	15 { R-N- C- C- } 0,1,5,6 R-N- CH- C } R-N- CH ₂ - C } R-N- C- C- C- } 0,1,2,3 R-N- CH- C } R-N- CH ₂ - C } R-N- C- C- C- } R-N- CH- C } R-N- CH- C }
6 R-CH- NH ₂ C- } 0,1,2,3	
7 R-CH- NH-C C- } 0,1,2,3	
8 R-CH- N- C- } 0,1,2,3	
9 R-C- C- C- } 1,2,...,9	
10 R-C- NH ₂ C- } 0,1,...,6	
11 R-C- NH-C C- } 0,1,...,6	

relatively small spectral region (0–5 p.p.m.) because the shift of a carbon atom is usually not very sensitive to atoms which are four bonds away. Even though the estimated shifts are not exact, they convey useful information to the matcher, and thus increase the overall program efficiency.

V. *Filtering*.—In the final stages of the analysis, the structure generator constructs amine molecules with N_0 carbon atoms by attaching to a central nitrogen atom, one or more alkyl chains which have survived the pruning process. These amines are passed to the filter, which is responsible for calculating their total ¹³C n.m.r. spectra and, *via* the matcher, comparing these predictions with the observations. If an amine passes the test, the filter writes out the structure along with the predicted shifts. It then repeats the spectral comparison using progressively smaller values of δ until it finds the smallest value, δ_{\min} , for which a match still

exists. In the event that several solution amines result from a particular run of AMINE, these δ_{\min} values can be helpful in ranking the candidates according to how well they fit the observed spectrum.

VI. *Spectrum Matching*.—Eventually, the pruning and filtering processes reduce to problems in spectrum matching. Suppose the matcher receives for testing a list of m predicted shifts, some of which may be represented by small spectral regions rather than exact values. Now, the predictive rules are not precise, so each shift is actually associated with a *range* of acceptable values (given the generic symbol ν) whose size is controlled by the input parameter δ . This parameter measures the maximum tolerable disagreement between predicted and observed shifts, so the range for a shift, S , extends from $S + \delta$ to $S - \delta$, while that for an estimated shift, bounded above and below by S_u and S_l , extends from $S_u + \delta$ to $S_l - \delta$. It should be noted that, in the latter case, there are really two factors which contribute to the breadth of the range. One is the basic imprecision in the predictive rules, while the other arises because the pruner, in its pre-tests, sometimes calculates shifts for carbon atoms whose environment is not completely known. The spectrum-matching algorithm,

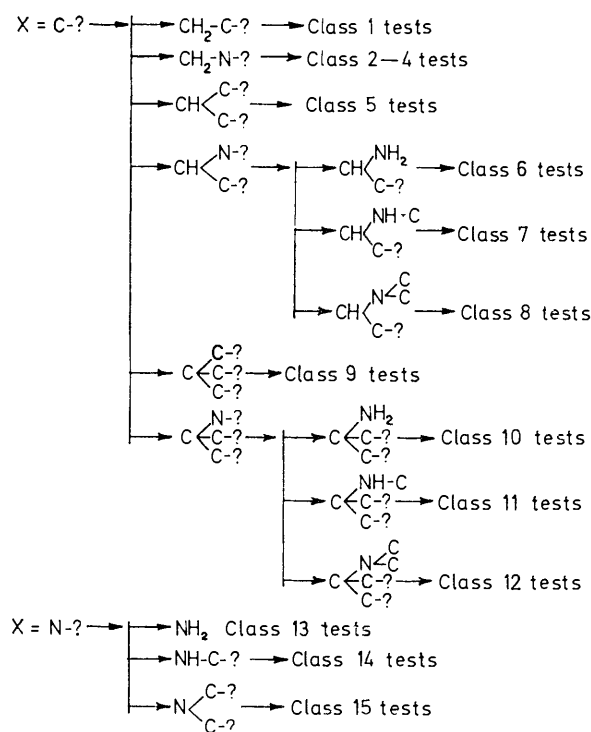


FIGURE 2 The hierarchy of pre-tests used by the pruner. A '?' attached to an atom indicates that the neighbours of that atom are unknown at testing time

though, makes no distinction between these; It only 'knows' that a predicted spectrum consists of a list of ranges. This list will be written as $\nu = (\nu_1, \nu_2, \dots, \nu_m)$, with u_i and l_i as, respectively, the upper and lower bounds for the range ν_i .

The nature of the observed spectrum, $o = (o_1, o_2, \dots, o_n)$,

has been discussed in Section I. The matcher takes these n shifts to be exact, because any estimated uncertainty (usually on the order of 0.1 p.p.m.) in their measurement may be included in the tolerance δ . It is the task of the matcher to ascertain whether r could be a subspectrum of o , or if $m = N_C$, whether r could be interpreted as o .

If, for a range r_i , there exists an observed shift o_j such that $u_i \geq o_j \geq l_i$, then it will be said that r_i can be assigned to o_j . The simplest test of agreement between r and o involves checking that each r_i in r can be assigned to at least one o_j in o . This test does not consider the important condition that, eventually, all shifts in o must be used, and therefore a stronger test can be defined.

If every carbon atom in the molecule gives a different observed shift, or if an analysis of peak intensity data gives the tally of each peak, then $n = N_C$. In this case, it is clear that no two predicted shifts can be assigned to the same o_j . Thus, referring to Figure 3, r does not

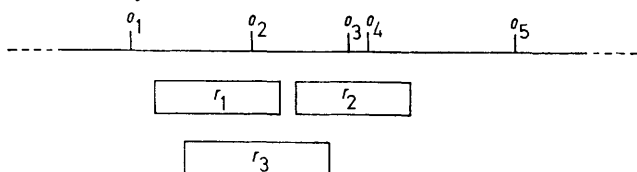


FIGURE 3 A case in which r and o do not match when $n = N_C$, even though the simple test is passed

match o even though the simple test is not violated, because r_1 and r_3 must both be assigned to o_2 . In more complicated cases, each of several r_i values may be assignable to two or more o_j values, and *vice versa*, so the application of this test in an efficient manner can present difficulties. Fortunately, simple matching theory,^{12a} an outgrowth of the mathematical field of graph theory, provides a general method (see matching algorithm, below) of finding the maximum number, M , of ranges which can be assigned to the elements of o without duplication. Clearly r cannot match o if this number is $< m$.

There may be cases, though, in which complete tally information is unavailable, which means that the number of observed shifts, n , is smaller than the number of carbon atoms, N_C . In such cases, there are $(N_C - n)$ 'extra' shifts which lie somewhere beneath the n observed ones, but there is no way of determining where they belong. It is still possible to strengthen the simple test, but here, the additional constraint is that the predicted spectrum, once assigned, can have no more than $(N_C - n)$ 'extra' peaks, either. If the simple test is passed, then every r_i can be assigned to at least one o_j . However, M is the maximum number of ranges which can be assigned to o without duplication, so $(m - M)$ must be the number of 'extra' ranges in r . The condition that strengthens the simple test here, then, is $(m - M) \leq (N_C - n)$. Because M cannot exceed m , this condition reduces to the previous one ($m = M$) when $n = N_C$.

* On the PDP-10, the program runs more slowly by a factor of ca. 4 (central processor time only).

The above spectrum-matching scheme is useful not only in the current study, but for general cases in which a set of predicted ¹³C n.m.r. shifts of variable uncertainty is to be compared with a observed spectrum with, perhaps, incomplete intensity information.

VII. *Matching Algorithm*.—The algorithm for determining M is related to the so-called Hungarian method^{12b} of simple matching, but takes advantage of certain special features of the spectrum-matching problem. It may be described briefly as follows. Begin with $M = 0$ and process the o_j values in algebraic order, beginning with the largest. For each o_j , scan the ranges r_i looking for those which satisfy $u_i \geq o_j \geq l_i$, but which have not yet been assigned. If there are none, proceed to the next o_j . If there is just one, assign it to o_j , increment M by 1 and proceed to the next o_j . If there are several, assign the one with the largest lower limit (l_i) to o_j , increment M by 1 and proceed to the next o_j .

It is possible to prove that this gives the maximum matching between r and o , but a presentation of our proof is beyond the scope of this paper.

VIII. *Results*.—The program AMINE has been implemented on IBM 360/67 and DEC PDP-10 computers. Any mention of timing in the following discussion refers to total execution time (central processor time plus 'wait time') on the former machine.* The program requires ca. 3.5×10^4 words of storage.

The only large set of amine ¹³C n.m.r. spectra available in the literature is that given by Eggert and Djerassi,^{9a} who used it in the derivation of predictive rules. The set consists of 102 amine spectra, including both shifts and tallies. Three of these spectra correspond to diastereomeric mixtures, and these are not suitable for testing AMINE, because the program assumes that the input spectrum corresponds to a pure compound. Neither is tridodecylamine suitable because it exceeds the maximum number of carbon atoms (currently 24) allowed by the program. The remaining 98 amines were used in the testing of the program.

Some experimentation indicated that $\delta = 1.5$ p.p.m. was small enough for efficient and selective program operation, yet large enough that ca. 95% of the test cases gave the correct solution among the output structures. Increasing δ by 50% to 2.25 p.p.m. slowed the program by a factor of 2–4, but AMINE always obtained the correct structure with this higher δ value. Generally, shift tallies were found to be unnecessary for amines containing 15 or fewer carbon atoms, but for larger molecules, the analyses proved to be excessively costly unless all the carbon atoms were identified in the observed spectrum.

With $\delta = 1.5$ p.p.m., and using tallies for the amines with 16 or more carbon atoms, the program obtained only one answer, the correct one, for the 88 amines listed in Table 2. The six cases summarized in Table 3 gave from two to seven solutions, with the correct structure ranked (see Section V) first or tied † for first. For three

† Two structures are considered to be tied when their δ_{min} values differ by 0.1 p.p.m. or less.

TABLE 2

Cases for which AMINE obtained only the correct structure using $\delta = 1.5$ p.p.m. and, except as noted, no tallies

Amine (prefix only)	Amine (prefix only)
Methyl	<i>NN</i> -Diethyl- <i>s</i> -butyl
Ethyl	<i>N</i> -Ethyl-di- <i>isopropyl</i>
Propyl	Nonyl
Isopropyl	<i>N</i> -Propylhexyl
Trimethyl	<i>N</i> - <i>s</i> -Butylpentyl
Butyl	<i>N</i> - <i>s</i> -Butyl-3-methylbutyl
<i>s</i> -Butyl	<i>N</i> - <i>t</i> -Butyl-3-methylbutyl
Isobutyl	<i>N</i> -Methyl-1,1,3,3-tetramethyl-butyl
<i>t</i> -Butyl	Tripropyl
Diethyl	Decyl
Pentyl	Dipentyl
1-Methylbutyl	<i>N</i> -Butylhexyl
2-Methylbutyl	<i>N</i> - <i>t</i> -Butylhexyl
3-Methylbutyl	<i>N</i> - <i>s</i> -Butyl-3,3-dimethylbutyl
2,2-Dimethylpropyl	Bis-(3-methylbutyl)
<i>N</i> -Methyl- <i>s</i> -butyl	<i>N</i> -Ethyl-dibutyl
<i>N</i> -Methyl- <i>t</i> -butyl	<i>NN</i> -Di- <i>isopropyl</i> butyl
<i>N</i> -Methyldiethyl	<i>N</i> -Pentylhexyl
Hexyl	<i>N</i> -Butyl-1-methylhexyl
1,3-Dimethylbutyl	<i>N</i> -Pentyl-1,3-dimethylbutyl
1,2,2-Trimethylpropyl	<i>N</i> -(3,3-Dimethylbutyl)pentyl
2,2-Dimethylbutyl	<i>N</i> -Butyl-1-ethylpentyl
Dipropyl	<i>N</i> -Methyl- <i>N</i> -butylhexyl
Di- <i>isopropyl</i>	<i>N</i> -Propyl-dibutyl
<i>N</i> -Ethyl- <i>s</i> -butyl	<i>N</i> - <i>isopropyl</i> -dibutyl
Triethyl	<i>N</i> -(1,3-Dimethylbutyl)hexyl
<i>NN</i> -Dimethyl- <i>s</i> -butyl	Tributyl
<i>NN</i> -dimethyl- <i>t</i> -butyl	<i>N</i> -Ethyl-dipentyl
Heptyl	<i>N</i> - <i>t</i> -Butyl-dibutyl
1-Methylhexyl	<i>NN</i> -Dibutyl-3-methylbutyl
1-Ethylpentyl	<i>NN</i> -Dibutylhexyl
1,3-Dimethylpentyl	<i>NN</i> -Dibutyl-3,3-dimethyl-butyl
<i>N</i> -Methylhexyl	<i>N</i> - <i>s</i> -Butyl-dipentyl
<i>N</i> -Ethylpentyl	<i>NN</i> -Dipentyl-1-methylpentyl
<i>N</i> - <i>isopropyl</i> butyl	Tripenyl
<i>N</i> - <i>isopropyl</i> - <i>s</i> -butyl	Tris-(3-methylbutyl)
Octyl	
1-Methylheptyl	Using tallies:
2-Ethylhexyl	Bis-(2-ethylhexyl)
1,5-Dimethylhexyl	<i>NN</i> -Dipentyl-1,3-dimethyl-butyl
1,1,3,3-Tetramethylbutyl	<i>NN</i> -Dibutyl-1,1,3,3-tetra-methylbutyl
Dibutyl	Trihexyl
Di- <i>isobutyl</i>	<i>N</i> -Butylbis-(2-ethylhexyl)
<i>N</i> -Ethylhexyl	
<i>NN</i> -Dimethylhexyl	
<i>NN</i> -Diethylbutyl	

of these six, the inclusion of tallies ruled out the incorrect answers. Four amines gave no solution with

$\delta = 1.5$ p.p.m. These were rerun using $\delta = 2.25$ p.p.m., and tallies were included to offset the longer running-times. As indicated in Table 4, three of these runs gave only the correct answer, while the fourth yielded two equally ranked solutions, including the correct one.

These analyses required from 0.02 to 100 s of computer time, with a typical C_{10} or C_{11} amine using *ca.* 1–2 s. In none of the runs was an incorrect solution obtained without the accompanying correct one, and in only four cases was it necessary to use the larger δ value. The results for the eight amines containing sixteen or more carbon atoms are especially encouraging. In a reasonable length of time, the program was able to select the correct structure, along with very few others, from an 'isomer space' containing from *ca.* 3×10^5 (for $N_C = 16$) to *ca.* 7×10^8 (for $N_C = 24$) members.¹⁰

The above results are biased to some extent because the amines used for testing the program are the same ones used by Eggert and Djerassi in the predictive-rule formation. As a test of the generality of the program, analyses have been run on the spectra of four 'unknown' amines which do not appear in the original list. The results of these test cases are summarized in Table 5. The spectra of the two C_{13} amines were analysed using $\delta = 1.5$ p.p.m., and no attempt was made to include tallies. Only the correct structure was obtained in these cases. For the two C_{20} amines, tallies were measured under special experimental conditions (see below). With $\delta = 1.5$ p.p.m., one of these gave two equally ranked structures, while the other gave none. A rerun of the second case with $\delta = 2.25$ p.p.m. yielded five solutions with the correct one ranked as tied for second. This is the only case in which the ranking procedure favoured an incorrect answer over the correct one, but here, as in most of the other multiple-result runs, the incorrect structures are sufficiently different from the correct one that they should be distinguishable by mass spectroscopic techniques.

IX. *Concluding Comments.*—Two major conclusions result from this study. First, the ^{13}C n.m.r. spectrum of an acyclic amine appears to be highly characteristic of

TABLE 3

Cases for which AMINE obtained two or more structures using $\delta = 1.5$ p.p.m. and, except as noted, no tallies

Amine (prefix only)	Solutions (prefix only)	Rank
Dihexyl	Dihexyl	Tied
	<i>N</i> -Pentylheptyl ^a	Tied
<i>N</i> -Pentyl-1,1,3,3-tetramethylbutyl	<i>N</i> -Pentyl-1,1,3,3-tetramethylbutyl	1
	<i>N</i> - <i>t</i> -Butyl-1,1-dimethylheptyl ^a	2
<i>N</i> -(1-Ethylpentyl)-1-propylbutyl	<i>N</i> -(1-Ethylpentyl)-1-propylbutyl	Tied
	<i>N</i> -(1-Ethylbutyl)-1-propylpentyl	Tied
<i>NN</i> -Dibutylheptyl	<i>NN</i> -Dibutylheptyl	1
	<i>N</i> -Butyl- <i>N</i> -pentylhexyl ^a	2
Diocetyl ^b	Diocetyl	Tied
	<i>N</i> -Heptylnonyl	Tied
	<i>N</i> -Hexyldecyl	Tied
Triocetyl ^b	Triocetyl	Tied
	<i>N</i> -Heptyl- <i>N</i> -octyl-nonyl	Tied
	<i>NN</i> -Diheptyldecyl	Tied
	<i>N</i> -Hexyl-dinonyl	Tied
	<i>N</i> -Hexyl- <i>N</i> -octyldecyl	Tied
	<i>N</i> -Hexyl- <i>N</i> -heptylundecyl	Tied
	<i>NN</i> -Dihexyldodecyl	Tied

^aThe use of tallies excludes these structures. ^bTallies were used in these runs.

the structure of the amine. For example, *only one* of the nearly 1.5×10^{10} structural isomers of $C_{20}H_{43}N$ gives a predicted spectrum which matches the observed

TABLE 4

Cases for which AMINE found no structures using $\delta = 1.5$ p.p.m. The correct solutions appeared when δ was increased to 2.25 p.p.m. and tallies were included

Amine
1-Isopropylhexylamine
<i>N</i> -Pentyl-1,2,2-trimethylpropylamine ^a
<i>N</i> -Butyl- <i>N</i> -(1,2,2-trimethylpropyl)pentylamine
<i>N</i> -Butyl- <i>N</i> -pentyl-(1,1,3,3-tetramethylbutyl)amine

^a A second structure, equally ranked, was found in this case: *N*-propyl-*N*-(1,2,2-trimethylpropyl)hexylamine.

spectrum of butylbis-(2-ethylhexyl)amine. Thus it can be concluded that ^{13}C n.m.r. data do indeed contain a tremendous amount of structural information. Secondly, it has been found that efficient methods for

In short, it appears that the computerized analysis of ^{13}C n.m.r. spectra holds great promise as an accurate and selective tool in the identification of unknown compounds.

EXPERIMENTAL

The four 'unknown' amines were prepared, and their proton-decoupled, Fourier-transform ^{13}C n.m.r. spectra obtained at 25.2 MHz, using previously described techniques.^{9a} The spectra of the two C_{20} amines were also run in the presence of chromium acetyl-acetonate, and the integrated intensities from these were used to determine the peak tallies.¹³ The observed shifts for the four amines are given, in p.p.m. downfield of internal TMS. The estimated uncertainty in these shifts is 0.1 p.p.m. For the two C_{20} amines, tallies are included in parentheses: 2-ethylhexyl-3-methylbutylamine, 53.6, 48.6, 39.8, 39.6, 31.7, 29.3, 26.3, 24.8, 23.2, 22.8, 14.1, and 11.0;

TABLE 5

Results obtained by AMINE for the four 'unknown' amines

Amine (prefix only)	Conditions		Solutions (prefix only)	Rank
	δ (p.p.m.)	Tallies used?		
<i>N</i> -(3-Methylbutyl)-1,5-dimethylhexyl	1.5	No	<i>N</i> -(3-Methylbutyl)-1,5-dimethylhexyl	
<i>N</i> -(3-Methylbutyl)-2-ethylhexyl	1.5	No	<i>N</i> -(3-Methylbutyl)-2-ethylhexyl	
<i>N</i> -Heptyl- <i>N</i> -(3-methylbutyl)-ethylhexyl	1.5	Yes	<i>N</i> -Heptyl- <i>N</i> -(3-methylbutyl)-ethylhexyl	1 (Tied)
			<i>N</i> -Pentyl- <i>N</i> -(3-methylbutyl)-ethylhexyl	1 (Tied)
<i>N</i> -Pentyl- <i>N</i> -(3,3-dimethylbutyl)-3,5,5-trimethylhexyl	2.25 ^a	Yes	2-Ethyl-1,5,5,7,7-pentamethyl-1-(2,2-dimethylpropyl)octyl	1
			<i>N</i> -Pentyl- <i>N</i> -(3,3-dimethylbutyl)-3,5,5-trimethylhexyl	2 (Tied)
			<i>NN</i> -Di- <i>t</i> -butyl-2-methyl-2-(2,2-dimethylpropyl)hexyl	2 (Tied)
			<i>N</i> - <i>t</i> -Butyl-1,1,3-trimethyl-3-(2,2-dimethylpropyl)octyl	2 (Tied)
			2-Ethyl-1,1,5,7,7-pentamethyl-5-(2,2-dimethylpropyl)octyl	2 (Tied)

^a With $\delta = 1.5$ p.p.m., no structure was found for this amine.

extracting this information exist, and can be implemented on the digital computer.

There is no reason to believe that these conclusions are peculiar to the acyclic amines. The computational techniques outlined in this paper can readily be generalized to other classes of saturated, acyclic, mono-functional compounds. To do so for a particular class, one needs to obtain an accurate set of predictive rules, and, perhaps, to modify the pruning process slightly to account for special features of those rules. Such rules already exist for alkanes,^{9c,d} alkenes,^{9e} and alcohols,^{9b} and as research in ^{13}C n.m.r. spectroscopy progresses, further sets should become available. Extensions to polyfunctional and/or cyclic classes would also require more sophisticated structure-generation methods, but these are available.^{2,3,6a,*}

* Exhaustive, irredundant methods for the generation of cyclic structures have recently been developed by L. M. Masinter and N. S. Sridharan as part of our heuristic DENDRAL project. A manuscript describing their work is in preparation.

1,5-dimethylhexylamine-3-methylbutyl, 53.5, 45.6, 39.9, 39.4, 37.8, 28.1, 26.4, 24.0, 22.7, and 20.6; 2-ethylhexyl-(3-methylbutyl)heptylamine, 59.6(1), 55.0(1), 53.1(1), 38.1(1), 36.8(1), 32.3(1), 31.8(1), 29.7(1), 29.4(1), 27.9(2), 26.5(1), 24.9(1), 23.6(2), 23.0(2), 14.3(2), and 11.0(1); and pentyl-3,3-dimethylbutyl-3,5,5-trimethylhexylamine, 54.5(1), 52.5(1), 51.9(1), 50.1(1), 40.9(1), 31.1(1), 30.2(3), 30.0(1), 29.8(1), 29.7(3), 27.7(2), 22.9(2), and 14.1(1).

Thanks are due to H. Eggert for preparing the 'unknown' amines and obtaining their spectra, and to L. Masinter together with Drs. N. S. Sridharan and B. Buchanan for helpful discussions and criticisms of this work. The financial support for this project provided by the National Institutes of Health is gratefully acknowledged.

[3/847 Received, 25th April, 1973]

¹² (a) 'The Theory of Graphs and its Applications,' C. Berge, Wiley, New York, 1964, p. 92; (b) p. 99.

¹³ S. Barcza and N. Engstrom, *J. Amer. Chem. Soc.*, 1972, **94**, 1762.