(6) Hendrickson, J. B.; Toczko, A. G. "Unique Numbering and Cataloguing of Molecular Structures". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171-177.
(7) Mackay, A. L. "A Metaphor for Molecular Evolution". *J. Theor. Biol.* **1975**, *54*, 399-401.
(8) Wipke, W. T.; Dyott, T. M. "Stereochemically Unique Naming Algorithm". *J. Am. Chem. Soc.* **1974**, *96*, 4834-4842.
(9) Morgan, H. L. "The Generation of a Unique Machine Description for Chemical Structures". *J. Chem. Doc.* **1965**, *5*, 107-113.
(10) Jochum, C.; Gasteiger, J. "Canonical Numbering and Constitutional Symmetry". *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 113-117.
(11) Wagner, R. A.; Fischer, M. J. "The String-to-String Correction Problem". *J. Assoc. Comput. Mach.* **1974**, *21*, 168-178.
(12) Angell, R. C.; Freund, G. E.; Willett, P. "Automatic Spelling Correction Using a Trigram Similarity Measure". *Inf. Process Manage.* **1983**, *19*, 255-261.
(13) Owolabi, O.; McGregor, D. R. "Fast Approximate String Matching". *Software Pract. Exper.* **1988**, *18*(4), 387-393.

^{13}C NMR Assignments of the Bases in Oligodeoxynucleotides: An Automated Procedure Using Bayesian Statistics†

TIMOTHY J. HYMAN, EILIS A. BOUDREAU, GILLES G. MARTIN,‡ BÉAT M. JUCKER, PHILIP N. BORER, and GEORGE C. LEVY*

NIH Research Resource for NMR and Data Processing and CASE Center, Syracuse University, Syracuse, New York 13244-1200

Received May 20, 1988

A statistical method, Bayes Maximum Likelihood, has been applied to the classification of base ^{13}C NMR resonances in DNA oligomers. An accuracy of 100% for carbon class discrimination was achieved for a preliminary training set of four oligomers using the following four parameters: (1) the chemical shift; (2) the temperature at which the spectrum was obtained; (3) the difference in chemical shift from the C5 resonances; and (4) a sequence factor representing the neighboring nucleotides. Classification of a fifth oligomer, previously assigned and not contained in the original training set, gave reasonable carbon class assignments.

INTRODUCTION

^{13}C NMR is a valuable tool for studying the conformation of nucleic acids and is particularly useful for studying internal dynamics in DNA oligomers. The most significant feature of ^{13}C NMR in contrast to ^{1}H NMR is that the carbon signals appear over a chemical shift range almost 20 times larger than for proton resonances. Since structural effects on the chemical shifts are on the order of 1-2 ppm for both nuclei, ^{13}C assignment is much simpler. The preferred methods for assigning the carbon signals are (1) selective isotopic substitution or (2) coherence transfer from the protons, where the proton spectrum can be unequivocally assigned by COSY and NOESY techniques.[1] Recently, the coherence-transfer method has been refined to be very sensitive by use of the two-dimensional proton-detected double quantum ^{1}H-^{13}C correlation experiment.[2-8] Both assignment methods require a large investment in time and effort by experts in chemical synthesis and/or NMR spectroscopy and further require facilities that are not routinely available.

A third method involves the comparison of monomer and oligomer reference spectra. This method has several drawbacks: (1) Peak assignment is time consuming. (2) A large amount of documentation is needed for comparison. (3) NMR experts are required to perform peak assignments. (4) Errors can occur when only comparative assignment is performed. An automated statistical approach to these assignments would be attractive.

An ideal computer program to perform peak assignment would possess a database of spectra for which all the peaks are known. Using expertise built into the software, the computer would compare unknown peaks to its database and automatically assign these unknown peaks to the correct carbon class. This could be performed in seconds by a laboratory technician vs hours for an NMR expert using current comparative assignment methods. Even if isotopic substitutions or 2-D methods are available, the classification scheme outlined here would provide independent confirmation of the assignments and help to resolve ambiguities that are inevitable when current methods are used.

There are several statistical methods to assign an unknown sample to one of several populations represented in a training set. The field of pattern recognition provides abundant literature on the optimal method depending on the training set used. Among these commonly used methods are clustering, Fisher discriminant analysis, principal component analysis, and nearest-neighbor classification. Bayes Maximum Likelihood (BML), which will be implemented in this study, is one of the most powerful and most general multivariate pattern recognition methods.

METHODS

(A) Collection and Assignment of ^{13}C Data. ^{13}C NMR data obtained on the molecules (1) $[d(TAGCGCTA)]_2$, (2) $[d(GGTATACC)]_2$, (3) $[d(CGCGCG)]_2$, and (4) $[d(GGCCT)]_2$ constitute the training set in this study. The synthesis and purification of the oligonucleotides and collection of the one-dimensional ^{13}C NMR spectra have been described previously.[9] The conditions for these experiments have been reviewed elsewhere.[7,8,10] A fifth molecule, $[d(TCGCG)]_2$, was implemented in testing the program's ability to assign carbon classes to the chemical shifts of an unassigned spectrum. Therefore, this molecule was not included in the training set. The test oligomer was classified by the program for spectra taken at 10, 30, and 75 °C.

The term "carbon class" refers to all resonances that correspond to a specific carbon type in a molecule. For example, there are two resonances belonging to a guanine carbon 2 (GC2) in $[d(TAGCGCTA)]_2$, shown in Figure 1. Individual carbon positions are specified as, for example, G3,2, which refers to guanine carbon 2 in chain position 3 from the 5′-end of the oligomer. This paper addresses only the base carbon

† This work was supported by grants from the National Institutes of Health (RR01317, GM32691, and GM35069).
* To whom correspondence should be addressed.
‡ Present address: Ecole Centrale de Paris, Department de Mathematiques Appliquees, 92230 Chatenay Malabry, France.

0095-2338/88/1628-0226$01.50/0 © 1988 American Chemical Society

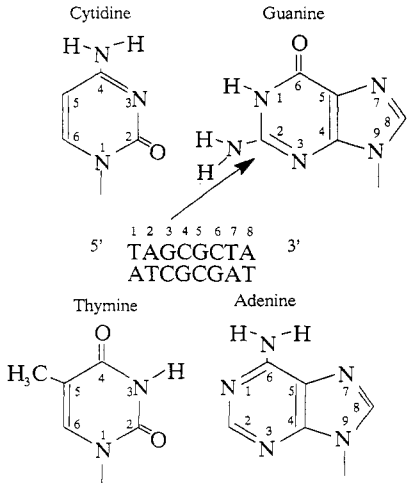


Figure 1. Segment of the $[d(TAGCGCTA)]_2$ strand showing the numbering of the carbons within the bases and the numbering of the bases within the strand.

Table I. Numerical Representations for the Nearest-Neighbor Sequence

sequence no.	base sequence[a]	sequence no.	base sequence[a]
01	AA/TT	07	CG/CG
02	AC/GT	08	GA/TC
03	AG/CT	09	GC/GC
04	AT/AT	10	TA/TA
05	CA/TG	11	end base
06	CC/GG		

[a] From 5′ to 3′.

classes and leaves consideration of the deoxyribose carbons for a later study.

Original assignment of carbon classes to chemical shifts for all five molecules was performed by using the comparative method.[11] In this method, the carbon classes of a molecule are assigned by comparing monomer, dimer, and trimer reference spectra and by using chemical shift vs temperature profiles obtained for each molecule. It has been shown by our group that the comparative method assignments of the base carbons are confirmed in coherence-transfer experiments.[11]

(B) Assignment of Parameters. Four parameters were determined for each carbon observation in the training set. These parameters were chemical shift (ppm or δ), the temperature at which the spectrum was obtained (T), the difference in chemical shift from the C5 resonances (DC5), and a sequence factor that represents the nucleotide neighbors of the given base (RL).

A sequence factor (RL) was devised to account for differences in the nearest-neighbor arrangement of bases. A particular number was assigned to each of the possible combinations of bases in a sequence as shown in Table I. Each base was assigned two numbers: one describing the neighboring base on its left side (ls) or 5′-end and another describing the neighboring base on its right side (rs) or 3′-end. These values were then used to obtain the sequence factor, RL = rs × 100 + ls. Using $[d(TAGCGCTA)]_2$ as an example, the values for ls and rs for G5,j are 07 and 09, respectively, to give an RL of 0709 for each carbon j in guanine 5.

The difference from the C5 chemical shift, DC5, was determined by subtracting the value of the C5 chemical shift in a particular base from all the other chemical shifts for that base. For example, for G5 in $[d(TAGCGCTA)]_2$ the value of DC5 would be δ(G5,2) − δ(G5,5), for G5,2, δ(G5,4) − δ(G5,5), for G5,4, etc. Of course, DC5 can be simply computed for the training set where the assignments are all known. For the test set, all possibilities must be evaluated as described later.

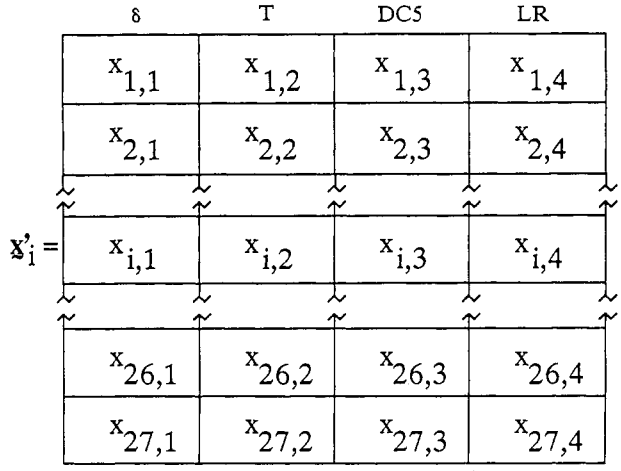


Figure 2. Example of one of the carbon class matrices (**X**) with vector $\mathbf{x}'_i$ shown.

(C) Classification Using Bayes Maximum Likelihood (BML). Mathematically, the training set can be represented as a series of matrices. Each matrix (**X**) represents a particular carbon class, where each column represents a parameter (δ, T, DC5, RL) and each row represents an observation from a spectrum. For instance, row 1 of the GC6 matrix has the GC6 chemical shift at the first temperature for the first G in the first duplex as x_{11}, the lowest temperature as x_{12}, the DC5 value as x_{13}, and RL as x_{14}; row 2 has the corresponding entries for the second temperature, etc. (see Figure 2). The 19 **X**-matrices in this training set have 27 rows for each of the five G-classes and four C-classes (nine separate sequence environments, each at three temperatures), 15 rows for the five T-classes (five sequence environments), and 12 rows for the five A-classes (four environments).

There are 22 separate carbons in the test spectrum of $[d(TCGCG)]_2$ to be classified by the method described here. Each of the observations can be described by a row vector $\mathbf{y}' = (y_1,y_2,y_3,y_4)$, where y_1 is δ and y_2 is the corresponding temperature. At the first stage of analysis, these are the only parameters defined; later, y_3 = DC5 and y_4 = RL will be defined to consider all $\mathbf{y}'$ vectors consistent with preliminary classification based on δ and T. At the various stages of analysis the $\mathbf{y}'$ are classified by BML into the carbon class for which $f(\mathbf{y}')$, the n-dimensional multivariate normal density function, is the highest. For each class, $f(\mathbf{y}')$ is given by

$$f(\mathbf{y}') = \{|\mathbf{S}^{-1}|/[(2\pi)^n]\}^{1/2} \exp[(-\tfrac{1}{2})Q(\mathbf{y}')] \quad (1)$$

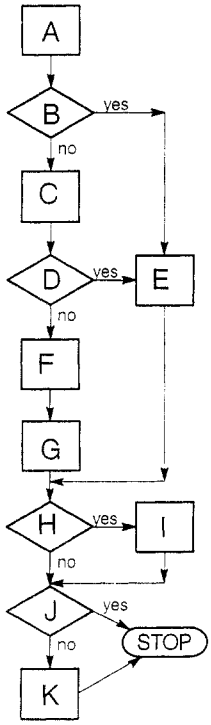
where

$$Q(\mathbf{y}') = (\mathbf{y}' - \mathbf{m}')^t\mathbf{S}^{-1}(\mathbf{y}' - \mathbf{m}') \quad (2)$$

$\mathbf{m}'$ being the mean vector of the class, $\mathbf{S}^{-1}$ the inverse of the covariance matrix of the class, and $|\mathbf{S}^{-1}|$ its determinant.

Bayes Maximum Likelihood classification yields optimal results when the distribution of original data is multivariate normal. A class is considered multivariate normal if all the observations fall within a multivariate normal distribution for all variables. This is questionable in our case since the temperature variable only takes a few discrete values. Therefore, an evaluation of the quality of the classification ability of our training set was performed by cross-validation.

Cross-validation computes a classification model as explained above (**S** and $\mathbf{m}'$ for each class of the training set are determined). This model is then used to reclassify every peak of the training set. A chemical shift is considered correctly classified if the BML classifier selects the carbon class the peak originates from and is considered incorrectly classified if any other carbon class is selected. The fraction of points in the training set that were correctly classified is then computed.



A. Calculate probability and proximity for all peaks using chemical shift and temperature.

B. For each chemical shift, can the signal be classified using the probability criterion?

C. For each peak not classified, calculate probability and proximity using chemical shift, temperature and distance from C5 for all possibilities found in (A).

D. For each chemical shift, can the signal be classified using the proximity ratio criterion?

E. Chemical shift is classified as that carbon class.

F. For each peak not classified, calculate probability and proximity using chemical shift, temperature, distance from C5 and sequence factor for all possibilities remaining in (D).

G. For each chemical shift, assign to the carbon class with the highest measure of proximity.

H. Does any carbon class have more assignments than is possible for the oligomer?

I. Classify the chemical shift as the carbon class that has the highest probability, measure of proximity criteria for that chemical shift in all previous steps.

J. Have all carbon classes in the oligomer been assigned to a particular chemical shift?

K. Classify remaining carbon classes to the chemical shift with the highest criteria for that class.

Figure 3. Flow chart outlining the classification scheme used for an unassigned spectrum.

This fraction is an indicator for the expected classification quality of the training set. A high percentage of correctly classified observations (over 90%) indicates that the data are well discriminated and can serve to classify other peaks whose assignments are unknown.

The Bayes Maximum Likelihood classifier used in this study is contained in ESS (Expert Statistical System), developed by New Methods Research, Inc. An introductory explanation of this system has been published.[12]

(D) Classification Ability of the Training Set. The quality of the training set was further tested by using the results of the cross-validation to classify the chemical shifts of the fifth oligomer, not contained in the training set. The procedure followed for the assignment of the chemical shifts to carbon classes for the test molecule is presented in Figure 3. At each classification step, the Bayes Maximum Likelihood classifier computes the probability of assignment and a measure of proximity for all possible carbon classes for each chemical shift. The probability of assignment is the probability that the chemical shift belongs to a given carbon class. The numerical representation is $f(\mathbf{y}')$ for the carbon class divided by the sum of all $f(\mathbf{y}')$ values for that observation. The measure of proximity is defined as 100(1 − contour level), where the contour level is the percentile value of the bivariate-normal equal probability ellipse on which the observation falls for a given class. Figure 4 represents the ellipses for the GC4 class.

RESULTS AND DISCUSSION

(A) Training Set. The cross-validation of the training set was performed by using the variables individually and in combinations. The overall percent correct classification was found to be 100% when all four of the variables (chemical shift, temperature, DC5, and RL) were used and when chemical shift and DC5 were used in combination as seen in Table II. This was significantly better than the use of chemical shift and temperature alone, which are the parameters normally used in the visual comparative method.

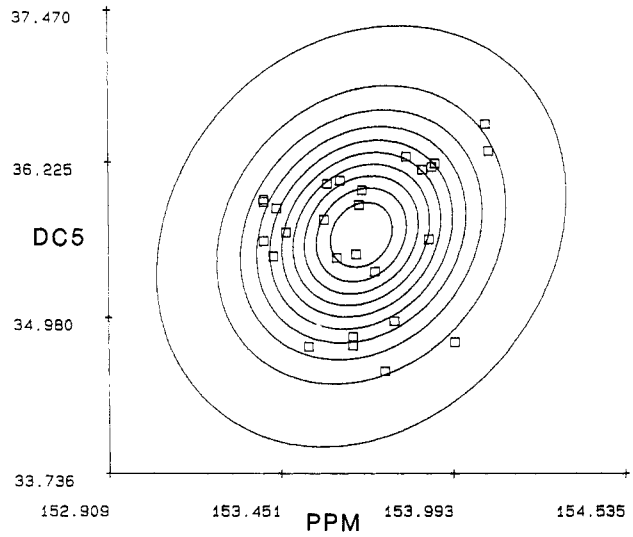


Figure 4. Graphical representation of the bivariate-normal equal probability ellipses for GC4 used in determining the measure of proximity during classification of unknowns. Contours are at 10% intervals, with the outermost contour being 99.5%.

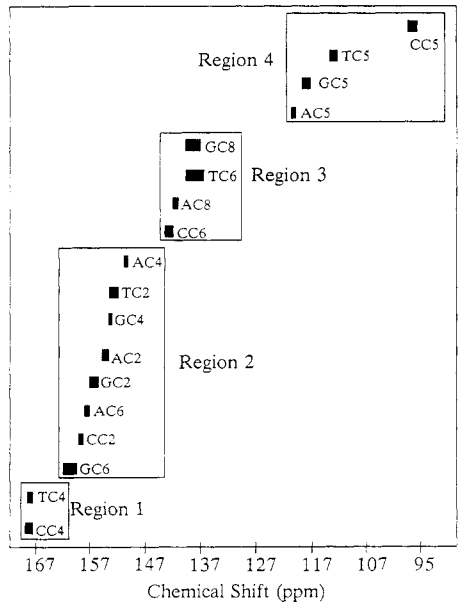


Figure 5. Plot of the chemical shift ranges for each carbon class showing the carbon class overlaps. The boxed regions are the problem areas in the comparative assignment.

Table II. Accuracy of Classification for Training Set Using Different Combinations of Parameters

parameters	accuracy, %
ppm	88.2
ppm and temp	89.0
ppm, temp, and RL	97.3
ppm, temp, RL, and DC5 or ppm and DC5	100

From the four duplexes in the training set, information was obtained for the chemical shift ranges for each of the base carbon classes. These ranges are shown in Figure 5. The boxes outline the problem regions in the visual comparative method where chemical shift ranges infringe on each other. There are four of these major infringing ranges, the pairs CC4,TC4; GC4,TC2; AC8,CC6; and GC8,TC6. Four other sets, GC6,CC2; CC2,AC6; AC6,GC2; and AC2,GC4, did not infringe on each other but had ranges separated by less than 1 ppm.

The difference from the respective C5 was a useful parameter for separating classes with infringing chemical shifts.

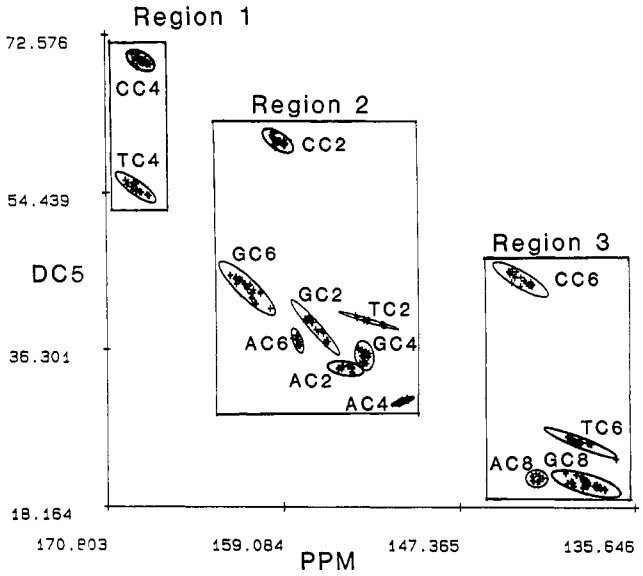


Figure 6. Plot of chemical shift vs the distance from the C5 carbon chemical shift of the base showing the separation between the carbon classes. The boxed regions correspond to the same problem regions as in Figure 5. Ellipses are at 99.5% contour levels.

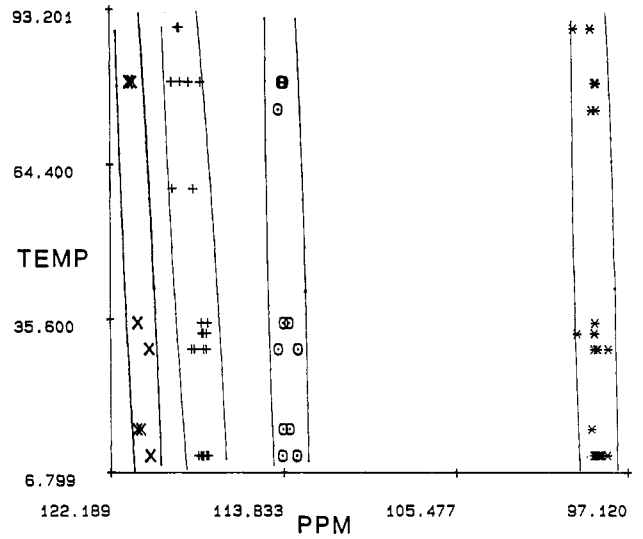


Figure 7. Plot of chemical shift vs temperature showing the definite separation between the different C5 carbon classes: (+) GC5; (*) CC5; (O) TC5; (X) AC5. Lines are at 99%.

Figure 6 shows the separations for the carbon classes based on chemical shift and DC5. In region 2 of Figure 5 there is extensive overlap of the chemical shift ranges; however, in Figure 6, the addition of DC5 allows unique distinction of the carbon classes. It seems the difference between chemical shifts was more constant than the chemical shifts themselves, making DC5 highly useful. As the structure of the oligomer is disrupted when the temperature is increased, there is either an increase or decrease in chemical shifts depending on the carbon class. Within a particular base of the oligomer, the chemical shifts change in a concerted fashion. When a change in structure alters a C5 frequency, the other carbon resonances on the same base will also change, and DC5 monitors the correlated changes.

The difference from the C5 chemical shift was implemented instead of any other difference because the C5 chemical shifts are well separated with respect to each other (see Figure 7) and easily classified on the basis of chemical shift and temperature alone. Their clustering away from the rest of the resonances in the spectra (see Figure 5) also made them the logical choice.

Table III. Percent Correct Classification Using Different Combinations of rs and ls

mathematical combination	accuracy,[a] %
rs + ls	90.6%
ls × 10 + rs	94.3%
rs × 10 + ls	97.1%
rs × 100 + ls	97.3%

[a] Chemical shift and temperature were also used in the analysis.

Table IV. Program Classification for Three Chemical Shifts of $[d(TCGCG)]_2$ at 10 °C

peak	δ	possible class	step A probability, %	step A proximity, %	step C proximity, %	step F[a] proximity, %
a	138.55	TC6	39.9	47.0	0.57	
		GC8	60.1	58.6	69.8	
b	139.27	TC6	67.5	53.6	9.3	0
		GC8	32.5	21.3	37.7	52.9, 35.2
c	139.80	TC6	87.6	32.8	21.4	0
		GC8	12.4	3.8	7.7	14.4, 11.6

[a] The two numbers in this column represent analyses for the two possible sequence factors for that base in the molecule.

A mathematical combination of the 5′-sequence (ls) and 3′-sequence (rs) factors had to be devised since they are correlated and cannot be included separately; Bayes Maximum Likelihood is designed for linearly independent variables. Table III shows the prediction accuracies for mathematical combinations of rs and ls. The mathematical combination RL = rs × 100 + ls gives a unique factor that accounts for sequence effects on chemical shift for every nearest-neighbor arrangement. The first two digits of the four-digit number are rs, and the last two digits are ls. This combination of rs and ls was chosen over the other possibilities on the basis of the high accuracy of discrimination that it produced. The other combinations in Table III produce degenerate sequence factors for certain nearest neighbors, and this probably gives rise to their lower accuracies in the discrimination. The significance of RL can be illustrated by an example: the chemical shifts in a guanine will be different if a neighboring adenine is on its 5′- or 3′-side. Therefore, it is essential to keep the uniqueness of the sequence numbers and in turn the placement of the neighboring bases.

As mentioned earlier, an accuracy of 100% was obtained when the analysis of the training set was performed by using chemical shift and DC5. It will be shown below that these two parameters alone were insufficient to completely assign an unknown spectrum. The temperature and RL had to be added for additional confidence in the assignments.

(B) Classification of an Unknown Spectrum. The spectrum of $[d(TCGCG)]_2$ was assigned according to carbon class by using the automated scheme outlined in Figure 3, which currently depends on user interaction and is the basis for a fully automated algorithm under development.

(1) Classification Using δ and *T*. The initial classification was accomplished by using steps A and B of the flow chart in Figure 3 and assigned 38 of the 66 observations with greater than 98% probability (the criterion used in step B), including all of the C5 peaks. At 10 °C the two peaks at 99.39 ppm were classified as CC5, the peak at 114.03 ppm was classified as TC5, and the peaks at 117.50 and 118.10 ppm were classified as GC5 with a probability of 100%.

(2) Classification Using δ, *T*, and DC5. The remainder of the analysis is illustrated with the three peaks in Table IV, which have possible assignments of either TC6 or GC8. At step B (Figure 3) none of their class probabilities exceeded 98%, so step C was performed. First, all of the possible DC5 values were computed. For peak a (Table IV), the value of DC5 for the TC6 possibility was 138.55 − 114.03 (the chemical shift assigned to TC5) and for the GC8 possibility was 138.55

– 117.8 (the average chemical shift for GC5 in the duplex). At step D, the ratio of the lower measure of proximity to the higher measure of proximity is determined. If this ratio is less than 0.1, the peak is assigned to the carbon class with the higher measure of proximity. At step D, eight more peaks are assigned. The proximity ratios for the peaks in Table IV are 0.008 (=0.57/69.8), 0.2 (=9.3/37.7), and 0.4 (=7.7/21.4) for peaks a, b, and c, respectively. The ratio for peak a is the only ratio less than 0.1, and peak a is classified as GC8. Peaks b and c must continue on to step F.

(3) Classification Using δ, T, DC5, and RL. At step F (Figure 3), RL is added to the classification for peaks b and c. The possible RL for each of the carbon class possibilities of step A are utilized. Only three pairs of infringing carbon classes are carried to this step: (1) GC4,TC2; (2) GC8,TC6; (3) CC4,TC4. In our example, there are two guanines in the oligomer: G3,8 has an RL of 0709, and G5,8 has an RL of 0711. For GC8, two classification analyses are performed for each of these sequence factors, along with chemical shift, temperature, and the DC5, for both peaks. This produces the two measures of proximity for step F (Table IV) for the GC8 possibility of peaks b and c. A classification analysis is also performed by using RL for T1,6 (1108) along with chemical shift, temperature, and the DC5 for TC6. Both peaks have a measure of proximity of zero for the TC6 classification. Therefore, both peaks are classified as GC8 at step G (Figure 3) since the GC8 classifications have a higher measure of proximity.

(4) Final Classification. We have now reached step H in Figure 3. The number of possible GC8 peaks in the oligomer is two, one for G3,8 and one for G5,8. By step G, all three peaks are classified as GC8. One of these peaks must be the TC6 since that is the only other possibility from the analysis and the possibility of another peak being TC6 did not occur. We have already determined that peak a is a GC8 peak at step E. Therefore, either peak b or peak c is the TC6. Peak b has a higher measure of proximity for GC8 than peak c for steps C and F and, peak c has the higher probability for TC6 at step A. Therefore, the assignment of GC8 to peak b and TC6 to peak c is the optimal conclusion.

A total of 58% of all base carbons were assigned to a unique class by step B in Figure 3. The additional discrimination of the later steps, while rather complex, allows logical class assignment for the remaining carbons.

The statistical classification of the test molecule not only yielded excellent results in assigning all 22 carbons but also allowed identification of chemical shifts in the areas where the visual comparison technique can be ambiguous. A major component of visual comparative assignment is the comparison of δ vs T profiles, whereas the BML assignments are made at three temperatures, independent of any correlations between observations at different temperatures. The visual and BML assignments for these carbons are in good agreement. The ultimate test of the method would be an independent assignment based on coherence transfer or isotopic labeling.

CONCLUSION

This computer technique used for the assignment of a chemical shift to a particular carbon class was shown to be useful in the identification of resonances in complex overlapping areas of spectra of DNA oligomers. This technique does not depend on the comparison of δ vs T profiles, which is part of the visual method and provides an independent check on the assignments. Furthermore, the BML method can be employed without much training in the interpretation of the ^{13}C spectra of oligonucleotides. Addition of other oligonucleotide data should enhance the accuracy of assignments. The long-term goal of this study is to create an automated method for assigning ^{13}C chemical shifts to the appropriate carbon class for both base and sugar carbons of oligonucleotides. This preliminary study has allowed us to determine a useful method for the classification of ^{13}C chemical shift data for the bases.

ACKNOWLEDGMENT

We thank Steven LaPlante for providing data on $[d(TAGCGCTA)]_2$, Nilo Zanatta for data on $[d(GGTATACC)]_2$ and $[d(CGCGCG)]_2$, and Anna Hakinnen for data on $[d(GGCCT)]_2$ and $[d(TCGCG)]_2$; the latter three molecules were kindly provided by Prof. Andy H.-J. Wang (MIT). All discrimination and classification analyses were performed by using ESS, an expert statistical system developed by New Methods Research, Inc., Syracuse, NY.

REFERENCES

(1) Wemmer, D. E.; Chou, S. H.; Hare, D. R.; Reid, B. R. "Sequence-Specific Recognition of DNA: Assignment of Nonexchangeable Proton Resonances in the Consensus Pribnow Promoter DNA Sequence by Two-Dimensional NMR ". *Biochemistry* **1984**, *23*, 2262–2268.

(2) Minoretti, A.; Aue, W. P.; Reinhold, M.; Ernst, R. R. "Coherence Transfer by Radiofrequency Pulses for Heteronuclear Detection of Multiple-Quantum Transitions". *J. Magn. Reson.* **1979**, *40*, 175.

(3) Griffey, R. H.; Poulter, C. D.; Bax, A.; Hawkins, B. L.; Yamaizumi, Z.; Nishimura, S. "Multiple quantum two-dimensional ^1H–^{15}N nuclear magnetic resonance spectroscopy: Chemical shift correlation maps for exchangeable imino protons of *Escherichia coli* $tRNA_f^{met}$ in water". *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 5895–5897.

(4) Bax, A.; Griffey, R. H.; Hawkins, B. L. "Sensitivity-Enhanced Correlation of ^{15}N and ^1H Chemical Shifts in Natural-Abundance Samples via Multiple Quantum Coherence". *J. Am. Chem. Soc.* **1983**, *105*, 7188–7190.

(5) Live, D. H.; Davis, D. G.; Agosta, W. C.; Cowburn, D. C. "Observation of 1000-fold Enhancement of ^{15}N NMR via Proton-Detected Multiquantum Coherences: Studies of Large Peptides". *J. Am. Chem. Soc.* **1984**, *106*, 6104–6105.

(6) Bax, A.; Aszalos, A.; Denya, Z.; Sudo, K. "Structure Elucidation of the Antibiotic Desertomycin through the Use of New Two-Dimensional NMR Techniques". *J. Am. Chem. Soc.* **1986**, *108*, 8056–8063.

(7) Leupin, W.; Wagner, G.; Denny, W. A.; Wüthrich, K. "Assignment of the ^{13}C nuclear magnetic resonance spectrum of a short DNA-duplex with ^1H-detected two-dimensional heteronuclear correlation spectroscopy". *Nucleic Acids Res.* **1987**, *15*, 267.

(8) LaPlante, S. R.; Ashcroft, J.; Cowburn, D.; Levy, G. C.; Borer, P. N. "^{13}C-NMR Assignments of the Protonated Carbons of $[d(TAGCGCTA)]_2$ by Two dimensional Proton Detected Heteronuclear Correlation". *J. Biomol. Struct. Dyn.* (in press).

(9) Borer, P. N.; LaPlante, S. R.; Zanatta, N.; Levy, G. C. "Hydrogen bonding Effects and ^{13}C-NMR of the DNA Double Helix". *Nucleic Acids Res.* **1988**, 16, 2323.

(10) Zanatta, N.; Borer, P. N.; Levy, G. C. In *Recent Advances in Organic NMR Spectroscopy*; Lambert, J. B., and Rittner, R., Eds.; Norell: Landsville, NJ., 1987; Chapter 6, p 89.

(11) LaPlante, S. R.; Boudreau, E. A.; Zanatta, N.; Levy, G. C.; Borer, P. N. "Natural Abundance ^{13}C-NMR of Three DNA Oligomer Duplexes: Assignment Methods for the Base Carbons and Structural Features". *Biochemistry* (in press).

(12) Martin, G. G.; Lienhart, J.; Hyman, T.; Levy, G. C. "A Statistical Expert System for Data Analysis". *Am. Lab. (Fairfield, Conn.)* **1988**, *20*, 74.