Evaluation of a ^{1}H-^{13}C NMR Spectral Library

S. K. Smith,* J. Cobleigh, and V. Svetnik

Merck & Company, Inc., P.O. Box 2000, Rahway, New Jersey 07065

Received March 30, 2001

A simple database of $^{13}C/^{1}H$-^{13}C spectral lists for 11 673 natural products was created in standard commercial database format. Over 50% of the spectra were predicted using HOSE code descriptors derived from the 50% of spectra having experimental values. Prediction errors obtained by prediction of and comparison to the experimental spectra revealed an exponentially decaying dependence between the average absolute error and the depth of the matching HOSE codes. A subset of the library containing over 1000 ^{1}H-^{13}C assigned experimental spectral lists were used to test against eight alternate query data sets. These sets represent query data from various combinations of 1D-^{13}C, 1D-DEPT, and 2D-^{1}H-^{13}C spectra. Simulated query lists were generated using Monte Carlo methods. As expected, queries based on 2D-^{1}H-^{13}C data were more likely to find the correct match under unfavorable conditions.

INTRODUCTION

^{13}C NMR spectroscopy plays a significant role in the identification and classification of unknown organic compounds from natural products. This is largely due to the well-known and exquisite dependence of the ^{13}C chemical shift and proton splitting pattern of each carbon atom on its local chemical environment and its number of attached protons, respectively. Furthermore, the highly resolved spectra, afforded by a large chemical shift range and narrow peak width, easily convert to highly reduced lists of chemical shift positions with minimal loss of information—peak intensity and width are features not generally used in dereplication. Libraries of such spectral lists are common for synthetic organic compounds and are an invaluable tool for confirming the identity of known compounds. In the field of natural products, where >154 000 compounds have been reported in the literature,[1] most compounds are absent from commercially available spectral libraries. However, the predictive correlation between molecular structure and ^{13}C NMR spectra have provided a variety of ready vehicles to expand libraries of experimental spectra with calculated spectra.[2,3]

Even if it were reasonable to make a comprehensive library of predicted spectra, one is faced with the fact that a complete ^{13}C spectrum is still the most costly data to acquire in terms of time and/or sample. Newer and more sensitive NMR equipment have not yet fully alleviated this significant problem, but one can use even more sensitive techniques based on ^{1}H-detected 2D ^{1}H-^{13}C spectroscopy. An HSQC[4] spectrum provides both carbon and proton shifts in a fraction of the time required for a complete ^{13}C spectrum. However, there are two drawbacks to this technique. First, quaternary carbon centers are systematically absent from the HSQC spectrum. This poses an especially serious problem when quaternary carbons comprise over 50% of a molecule. Fortunately, quaternary carbons comprise only about 30% of the average natural molecule. Second, the precision of these 2D spectra is at least an order of magnitude lower than corresponding directly detected 1D spectra. It is relevant to ask then if an incomplete and imprecise ^{13}C list extracted from a 2D spectrum would return the correct compound when submitted as a query against a library of ^{13}C spectra? Furthermore, could the query be improved by adding correlated ^{1}H data extracted from the same 2D spectrum and would this offset the significant cost of rebuilding a ^{13}C spectral library as a ^{1}H-^{13}C library?

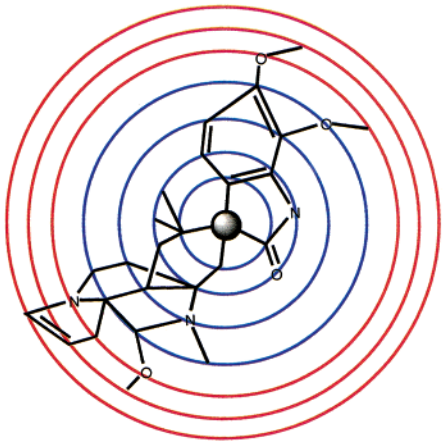
Using the in-house program, SIMSER,[5] as a starting point, a new application has been developed to provide general access both 2D ^{1}H-^{13}C and 1D ^{13}C data in a NMR spectral library of natural products. The new application, SimSearch, was coded in a variety of fourth generation programming languages, uses standard database software, and was the platform for the comparative evaluation of 1D and 2D NMR data for querying a small 2D NMR library, vida infra. SimSearch includes a new ^{13}C spectrum predictor, which also required extensive evaluation. The evaluations were carried out using general methodologies developed to measure prediction and search engine performance. Spectral prediction accuracy was tested by removing each spectrum record out one at a time and then predicting the missing spectrum from the structure of original molecule. Performance and response of the search engine to query types were evaluated using Monte Carlo simulated spectra. Simulated spectra were generated by perturbing the peak positions of an original parent spectrum. Small random deviations in chemical shift (ppm-noise) were added to each peak position. Searches were repeated using newly simulated spectra for each of eight combinations of virtual 1D and 2D NMR data.

EXPERIMENTAL SECTION

Over 17 000 natural product structure records were gathered from a variety of internal MDL ISIS databases. These were merged into a single database with 11 673 unique entries, 5733 of which contained ^{13}C assignments. Complete ^{1}H-^{13}C assignments were recorded for 1076 of the assigned entries. Inevitably, as a database of this nature grows, errors from the manual transcription process and even from the literature itself can lead to significant contamination of the

* Corresponding author e-mail: Scott_K_Smith@merck.com.

10.1021/ci010324m CCC: $20.00
Published on Web 09/13/2001



1 2 3 4 5 6 7
C/ CCCC / = O:N,*C*C,CC:C:,C / &,*C&,*C,C&,CN& / *CO,*&,C,N,CC: / *&O,C:,C&&,C&,O& / C:,C, = &,C: / = &

Figure 1. Complete HOSE code description (seven spheres) of a central carbon atom of paraherquamide.

data which is then propagated into the predicted data. ISIS/PL scripts were used to detect obvious chemical shift errors, such as shifts >350.0 ppm, as well as errors based on a few simple rules regarding proper ranges of chemical shift ranges for several easily identifiable functional groups. Structures were also checked for disconnected or five-coordinate carbons and inspected when flagged by an outlying chemical shift. The training structures and assignments were further refined by iterative submission of the full training set, stripped of its assignments, to the prediction module. Comparison of the estimated shifts with the assigned shifts, in the training set, provided another means to detect suspect assignments or structures.

HOSE code structure representation was chosen for the shift prediction model, despite this method's inability to encode for geometrical/stereochemical isomerism. The resulting bimodal distributions of chemical shifts for identical HOSE codes were handled in both the prediction stage and at the search engine. We chose to keep with HOSE code descriptors as they were used in the original SIMSER application with but one modification. The size limit of four spheres was removed from the HOSE code generator so that the new module generates an inclusive description of any given molecule (Figure 1). The chemical shift prediction module was written entirely in PERL 5 and consists of three modules: the HOSE code generator, a database of chemical shifts paired with their related HOSE code, and search engine for that database. To achieve inclusive molecular descriptors, the whole molecule must be traversed and accounted for, including ring closures. Perception of all ring closures in a structure gives rise to the smallest set of smallest rings from SDF connection tables. This was achieved by application of the *chain-message algorithm*, as described by Balducci and Pearlman.[6] In short, as the structure is traversed, atoms and bonds are recorded. When a branch point is reached, the traversal splits to follow all possible paths. Finally, if two separated traversals collide, the collision is detected signaling a ring. The object-oriented programming supported in PERL 5 was especially suited to handling and processing the ASCII input and output of HOSE code generation.

HOSE code-chemical shift pairs were loaded into a PERL accessible database file, CDB, and linked to a HOSE code search routine. Grouping by carbon multiplicity, alphabetical sorting, and a simple binary search algorithm was used to coarsely match HOSE codes from an unassigned molecule to a HOSE code in the CDB database. The chemical shift in the assigned database, associated with the best matching code or codes, as measured by the number of matching shells or "match-depth", was assigned to the query code. When the same number of matching shells are found, the algorithm continues searching for the best match, element by element within the first nonmatching shell. The protocol used to make the final selection varied depending on the number of discrete codes having the greatest and equal match-depth. Selections were grouped into three general categories in the following manner. One-to-one matching between the query code and a single training code was assigned the "match-type" category "best". One-to-many matching (i.e. isomers) between the query code and a set of training codes used two protocols. When the bimodal distribution of chemical shifts for a set of HOSE codes was small (<5 ppm), the distribution was reduced to its average for the subsequent assignment. When the distribution was large (>5 ppm), the extremes of its values were assigned, and the matching typed was set to "max" and "min", respectively. Proton assignments were propagated along with carbon assignments only when the match-type was best, thus avoiding interpolation errors in the proton dimension.

Thorough testing of these PERL modules was performed to verify its basic operation, to provide feedback for error correction, and to provide data for evaluating prediction performance. Refinement cycles consisted of submitting each of the training set structures to the prediction engine. Identical matches were discarded, forcing the search to provide a predicted shift from the remaining data. The effect of inclusive HOSE code generation was characterized using information retained from the last refinement cycle. This information included shift errors, match-depth, match-type, and molecule/atom identifiers.

The final predicted assignments were combined with the real assignments and loaded into a production Oracle database. A copy of the subset of real ^{1}H-^{13}C spectral assignments was extracted into an Access database. These data provided an opportunity to test the use of correlated 2D data in a chemical shift library. A search engine was written which accepts any combination of ^{13}C shifts, ^{1}H-^{13}C shift pairs, and multiplicity data. A resulting hit list was ranked by a normalized similarity index calculated from the minimum number of smallest errors arising from a peak-by-peak comparison of query to library spectrum. Although the entire database can be searched, it is clearly faster to delimit the number of spectra to compare by some selection rules such as total number of peaks or a maximum absolute error. The search is initiated with a list of carbon chemical shifts, a range for the number of peaks to look for in a spectrum and a maximum absolute shift error. The search engine, written as a stored procedure in PL/SQL for the Oracle version and Visual Basic for the Access version, makes an initial query for spectral records with the appropriate number of peaks. A subquery compares the query list to the spectra returned by the first query. Originally, the SIMSER application calculated a difference index D_S, from the set of smallest errors. As shown in eq 1, an explicit penalty for unmatched peaks is included, and the final total is partially normalized to the average of the peak counts in the query and library spectra. Search results were then ranked in ascending order starting from the smallest D_S.

$$D_S = \frac{\overbrace{C^{dmax} * (Q_n + L_n - 2H)}^{\text{penalty for misses}} + \sum \overbrace{\left|C^Q - C^L\right|}^{D_p}}{\underbrace{(Q_n + L_n)/2}_{\text{normalization}}}$$

C^{dmax} = adjustable maximum difference
C^Q = Query ^{13}C ppm; Q_n = #peaks in query list
C^L = Library ^{13}C ppm; L_n = #peaks in library spectrum
$D_p = 0$ if $\left|C^Q - C^L\right| > C^{dmax}$ H = count of ($D_p > 0$)

Equation 2 shown below mean centers and normalizes peak differences D_p individually before converting them to a similarity index and summing them. The final total is normalized to unity by an empirically derived function of the number of hits, query peaks, and library peaks. Search results are ranked in descending order of similarity index. Multiplicity information, when available, was used in the subsearch to delimit the number of query and library peaks to pair up for determining the minimum number of smallest errors.

$$SI_S = \overbrace{\left(\frac{2H}{Q_n^{\,2} + L_n^{\,2}}\right)}^{\text{normalization \& penalty function}} * \sum \overbrace{1 - \left(\frac{D_p - \overline{C^{QL}}}{C^{dmax}}\right)}^{SI_p}$$

$SI_p = 0$ if $\left|C^Q - C^L\right| > C^{dmax}$

$\overline{C^{QL}} = \frac{1}{H}\sum \frac{C^Q - C^L}{C^{dmax}}$ mean centering

^{1}H-^{13}C shift -pair and -triplet data obtained from 2D spectra can be represented by points in a pseudo-three-dimensional shift space where the carbon dimension is one axis, the first proton shift is the second axis, and the second proton shift is the third axis. (Asymmetric methylene protons usually have different ^{1}H chemical shift values, while methyl protons do not.) Equation 3 is a simple extension of eq 2 to accommodate ^{1}H-^{13}C spectral data and reduce it to a scalar using their sum, the Manhattan-Distance of the three dimensions. Thus, the similarity between a query point (C^Q, H_1Q, H_2Q) and a library point (C^L, H_1L, H_2L) is a function of the difference between the related shifts projected onto their relevant axis. Normalization of SI_p is maintained by the introduction of an adjustable parameter H^{dmax}, the maximum allowed absolute shift difference in the proton dimension, and a calculated parameter n (*where* $n = 1 +$ count of proton shifts correlated to the carbon shift).

$$SI_S = \left(\frac{2H}{Q_n^{\,2} + L_n^{\,2}}\right) * \sum 1 - \frac{1}{n}\left(\overbrace{\frac{\left|C^Q - C^L\right|}{C^{dmax}} + \frac{\left|H_1^Q - H_1^L\right|}{H^{dmax}} + \frac{\left|H_2^Q - H_2^L\right|}{H^{dmax}}}^{SI_p\ \text{(Manhattan Sum)}}\right)$$

$SI_p = 0$, if $\left|C^Q - C^L\right| > C^{dmax}$

or $\left|H_1^Q - H_1^L\right| > H^{dmax}$ or $\left|H_2^Q - H_2^L\right| > H^{dmax}$

$H_1^{Q/L}$ = First (Query/Library) proton shift

$H_2^{Q/L}$ = Second (Query/Library) proton shift

The normalization function outside the summation sign in eqs 2 and 3 also provides some degree of penalty in the event that peaks in the query do not match the number in the library. That is, it can be seen as a penalty function $(x-\epsilon)/\epsilon' x^2$, where $x = \min(Q_n, L_n)$, ϵ = miss count, and $\epsilon' = (1 + 2|Q_n - L_n|/x)$. When $Q_n = L_n$, it reduces to $(x-\epsilon)/x^2$. This combined normalization and penalty function is independent any chosen maximum shift difference and remains between 0 and 1; therefore, results from searches run under different conditions may be compared. Equation 3 reduces to eq 2, except for the mean centering term, when proton data are unavailable in the query spectrum list.

For completeness, eight virtual spectral-data combinations were chosen for the 1D/2D query comparison (Table 1). The ideal query mode was represented as a combination of 2D ^{1}H-^{13}C data available from an HSQC spectrum, multiplicity information available from a DEPT spectrum, and quaternary carbons available from a directly detected ^{13}C spectrum. The least ideal query mode was represented as a directly detected ^{13}C spectrum wherein it was assumed that quaternary carbons went undetected because of poor signal-to-noise, and no multiplicity data were obtained.

Quantitative comparison of these eight query data types was performed using Monte Carlo simulation methods on the 2D subset of the full database. Queries were based on spectral data for nemadectin, selected from this subset (Figure

Table 1. Query Modes Defined by Their Corresponding Spectral Data Sources, Distribution of Data Objects, and Total Number of Data Points

query mode		simulated query data (7-C_q, 15-CH, 6-CH_2, 7-CH_3)	mult > 0 (28)	mult = 0 (quaternary) (7)	^{1}H (34)	total (104)
1D	*MQ*	DEPT[8] & ^{13}C	+	+	-	63
	Mq	DEPT	+	-	-	56
	MQ	high S/N ^{13}C	-	+	-	35
	Mq	low S/N ^{13}C	-	-	-	28
2D	*MQ*	DEPT-edited HSQC & ^{13}C	+	+	+	97
	Mq	DEPT-edited HSQC	+	-	+	90
	MQ	HSQC & ^{13}C	-	+	+	69
	Mq	HSQC	-	-	+	62

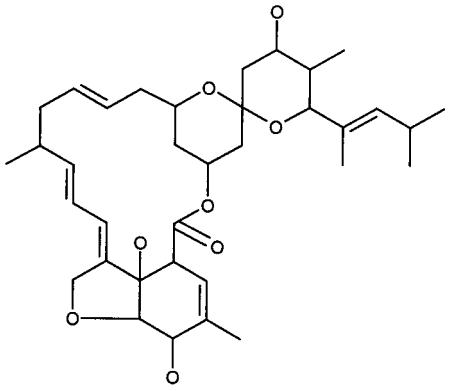


Figure 2. Nemadectin[9] a 16-membered macrolide was selected from the ^{1}H-^{13}C database as a representative structure having a variety of carbon types and functional groups.

2). Uniformly distributed noise was generated using Access (Visual Basic rand() function) and transformed using the requisite scaling and shift to obtain 10 different magnitudes of PPM noise centered on zero. Thus, the position of every peak for nemadectin was perturbed independently for every query. The resulting query spectra were submitted to the search engine against the 1076 record ^{1}H-^{13}C database. This process was repeated with at 10 noise levels from (1−10 ppm for carbon, 0.05−0.5 ppm for proton) to make one full cycle. The full cycle was repeated 60−90 times for each query mode. Match results (where the similarity index was > 0.25) were saved to a file for analysis. The top 60−90 hits (the number of simulation cycles) at each noise level were used to count the number of times the spectrum of nemadectin was found at the top of the ranked search results. The average of the similarity index for the top hit at each noise level for each query mode was also calculated.

RESULTS

To verify that the HOSE code generator was functioning properly it was tested on a variety of structures, including large multiring macrocycles. In every case the correct number of ring closures was determined, and under no circumstances did the module fail to produce a complete HOSE code. Subsequent conversion of the assigned structures and unassigned structures to HOSE codes produced 135 425 and 150 000 codes, respectively.

Exhaustive testing of the chemical shift prediction module was carried out ostensibly for two reasons; to obtain a measure of the quality of the structure/spectral data itself and to determine what correlation exists between prediction accuracy and the match-type and match-depth. The first attempt at the "leave-one-out" prediction test resulted in an average prediction error of 4.0 ppm with 4% of the errors greater than ±15 ppm. The largest errors were used to pinpoint drawing or assignment errors in the original data for correction. After three cycles of prediction and correction, the average prediction error dropped to 2.6 ppm, with only 2.3% of the outliers greater than 15 ppm, and no errors greater than 110 ppm. No further refinements were performed. This process was halted when 97.5% of the predicted assignments were within ±15 ppm of their correct values.

Table 2. Overall Prediction Results: Summarized by Quantile and Multiplicity

population (%)	error (ppm)	multiplicity	error (ppm)
50	<0.9	CH_3	±3.3
80	<4.0	CH_2	±2.6
90	<0.4	CH	±2.3
99	<21	C	±1.9

Prediction errors from the last refinement cycle of the training set were grouped into 5 ppm bins over the full carbon spectrum Figure 3. The box-plots clearly show a significant skew to the data such that the median error is approximately half the average error because of the occurrence of errors up to 100 ppm. These results demonstrate that the errors in the predicted shifts are independent of chemical shift position and the population distribution of the carbon chemical shifts in the library. The average error 2.6 ppm is comparable to the error reported originally for SIMSER. This is well within deviations observed in experimental data due to variations in sample preparation alone.[7] The average of the shift errors grouped by molecule for all 5733 assigned structures are also shown. The assigned shift database was organized by carbon type, but no statistically meaningful correlation between carbon type and prediction error was observed. Table 2 summarizes the results from the final prediction test.

The majority of codes (90%), which were not perfectly matched (i.e. isomers), were assigned chemical shifts according to the depth and type of match. As expected, match type and depth were strongly correlated with prediction error. The average error for a given match depth was fit using nonlinear regression to a three parameter exponential function. These fits were further refined by partitioning the match-sphere scores and chemical shift errors by match-type (Figure 4). Clearly, the min/max match-type demonstrated the least consistent prediction performance. The models were useful in estimating the error to expect when using the prediction engine on unassigned structures. The fitted parameters and estimates for the predicted library are summarized in Table 3. Finally, the completed set of 11 673 experimental and predicted spectra were loaded into an Oracle database.

The subset of fully assigned ^{1}H-^{13}C spectra was left in an Access database for all subsequent Monte Carlo simulated test searches. The purpose of these simulations was to

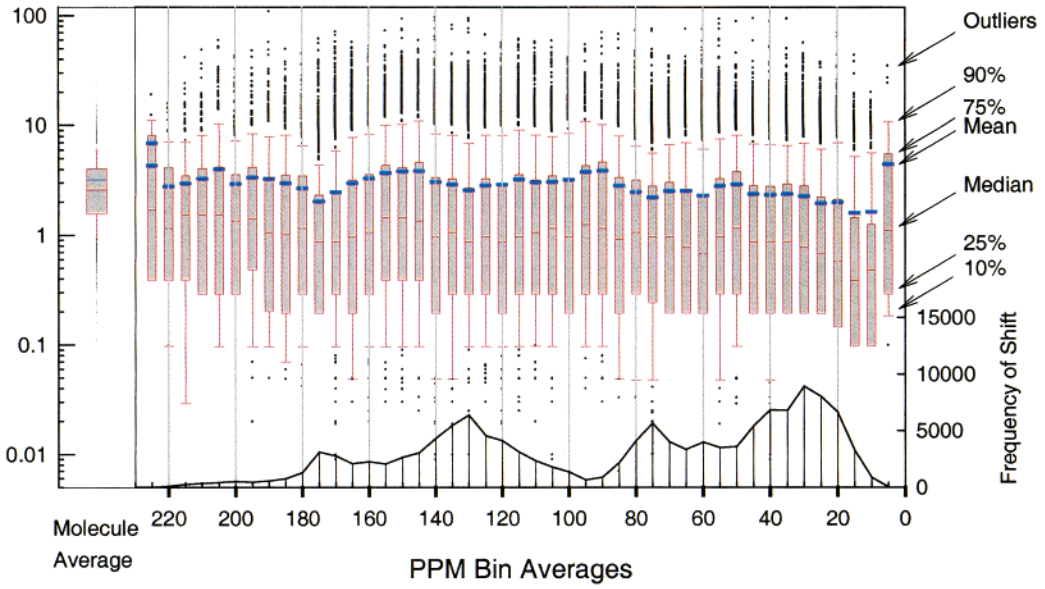


Figure 3. Prediction errors as a function of chemical shift. Box-plots were used to present the complete range and profile of errors in 5 ppm sections of the full ^{13}C spectrum. Outer wiskers are 10th and 90th percentile limits; the shaded region is bounded at the 25th and 75th percentile limits: heavy line - mean; light line - median. (The boundary of the box closest to zero indicates the 25th percentile. The lines within the box mark the median and mean; the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 10th and 90th percentiles, and data points shown above and below the whiskers are the outlying 10%.)

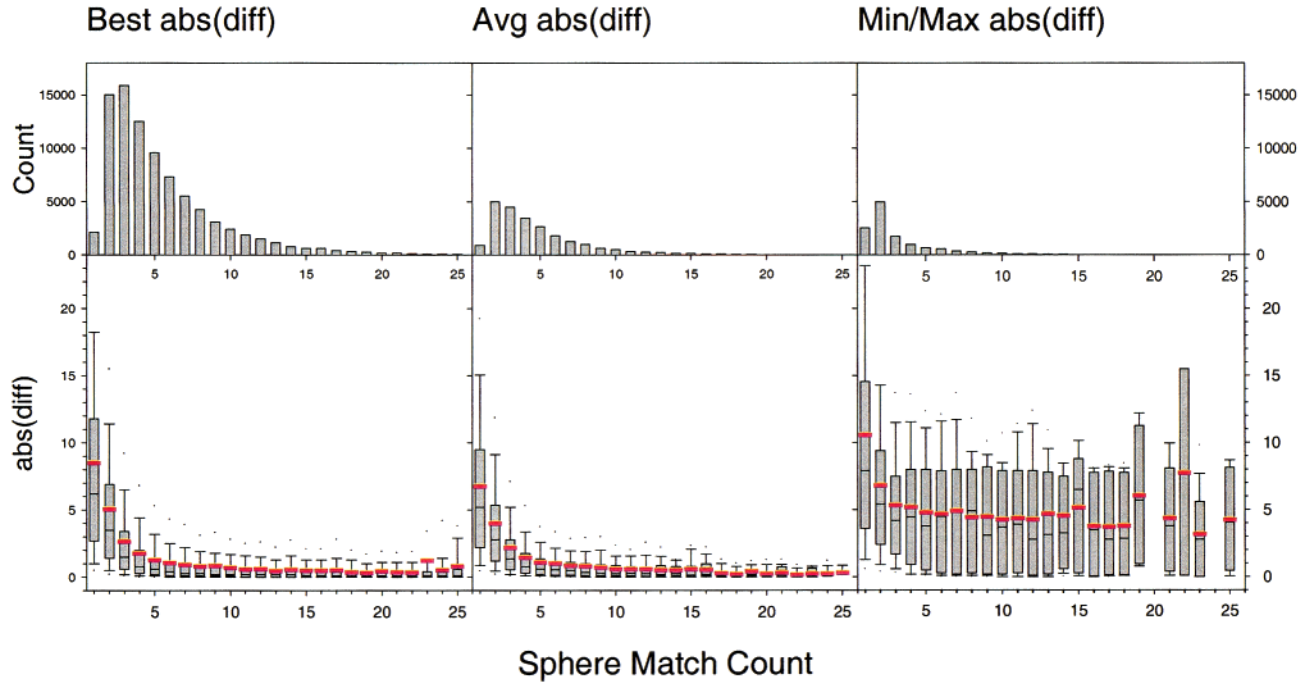


Figure 4. Prediction errors as a function of shell depth and match-type.

Table 3. Prediction Error Model[a]

match-type	D_0	D	K	error (spheres = 4)
best	6.0e-1	15	6.1e-1	1.9
av	4.9e-1	11	5.8e-1	1.6
min/max	4.6	16	9.6e-1	5.0

[a] Estimated error = $D_0 + D^{*}\exp^{-K^{*}\text{spheres}}$.

determine the noise response of eight query modes as described in Table 1. While typically a search would be truncated at the top 25−50 hits, for this work all similarity indexes greater than 0.25 were saved for additional processing and analysis. The relative change in the raw similarity indexes to increased ppm-noise for any given query mode was enhanced by a simple fourth power transform. Unit normalization of the similarity index permitted this operation to highlight differences between high and low SI values without changing the overall range of the SI data (0−1).

The transformed SI[4] data for each Monte Carlo simulation were grouped by noise level, sorted, and plotted for each query mode as shown in Figure 5. The trajectory of the noise response of each query mode show that at low levels of injected noise the resulting hits are resolved into two distinct groups, separated by a clear transition in the magnitude of SI. The transition coincides with the number of times Monte

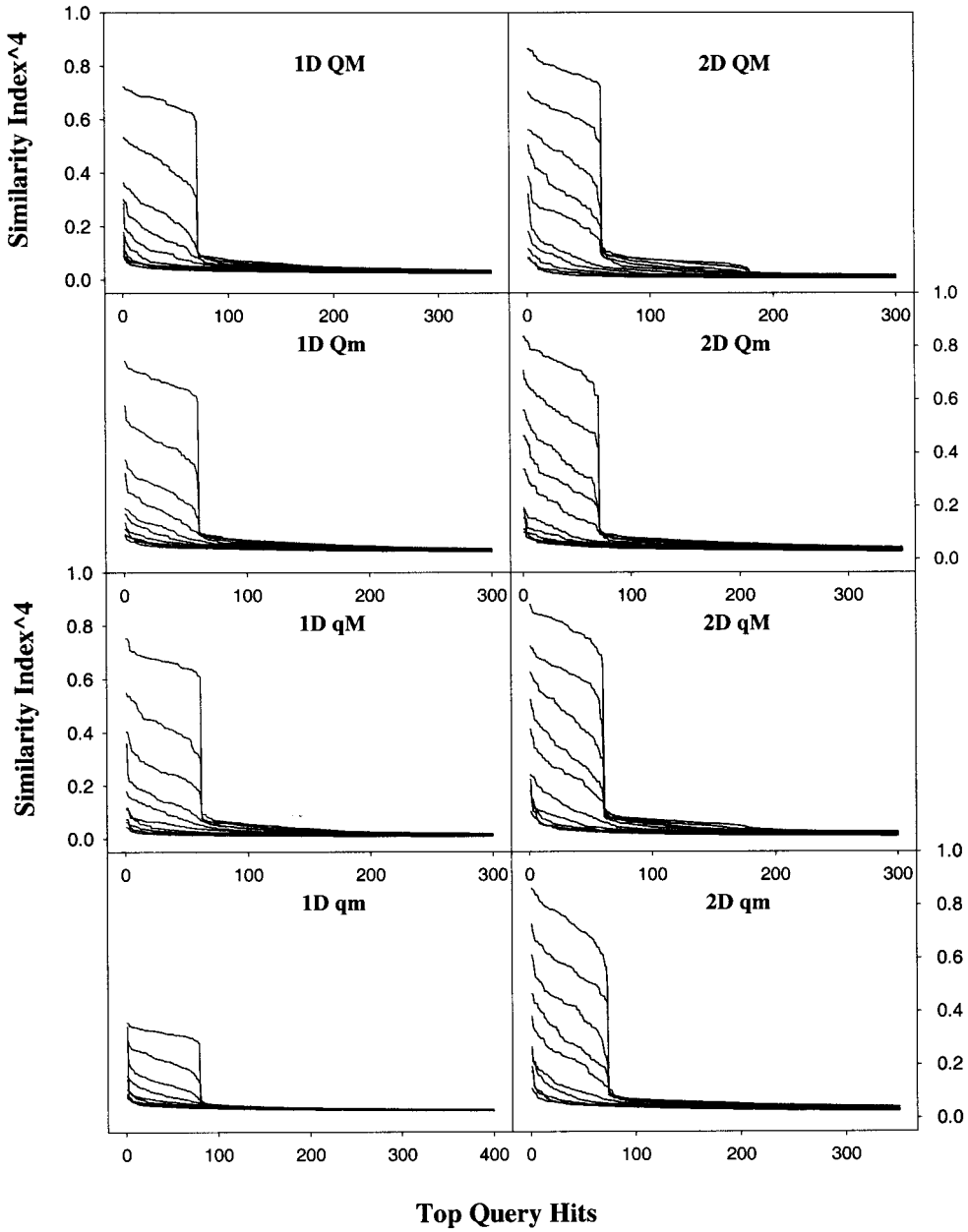


Figure 5. Trajectory plots of results from Monte Carlo simulated queries.

Carlo simulations were executed on the query mode in question, implying that the parent spectrum was the highest ranking hit in each of those queries. As the level of injected noise increases, the transition becomes less discernible, until at the highest levels there is no measurable difference between hits. The transitions in the 2D query modes survive down to 5 ppm's of injected noise, but those in the 1D query modes survive only to 3 ppm, suggesting that the 2D modes are, in general, more robust than the 1D modes. In fact, the 2D mode with the least amount of data (qm) for its query performs nearly as well as the ideal 1D mode (QM). The trajectory plots for the 2D-QM query mode, albeit the most expensive data set to acquire, shows that in some cases, similar spectra may be detected as a second plateau and transition in the plot.

Figure 6a shows the average value of SI^4 for the hits above the transition in each query mode. While the differences are not dramatic, the trend confirms differences implied by the trajectories shown in Figure 5. The 1D-qm query mode, where quaternary carbons are missing but not noted as being systematically absent, demonstrates that the least robust set of data to search is an incomplete set which one assumes to be complete. Figure 6b shows the $(\text{probability})^4$ of finding the parent spectrum in the set of hits above the transition. The large decrease in probability occurring at 5−6 ppm for all query modes coincides with the C^{dmax} error limit (5 ppm) used in these queries. This observation is probably an artifact of the simulation conditions, assuming that the Monte Carlo simulated spectra with >5 ppm noise should have more unmatchable peaks than those with <5 ppm noise. Nonetheless, the trend that 2D query modes perform slightly better than 1D modes is consistent with the rest of the results.

DISCUSSION

Careful analysis of shift prediction results from refinement of the experimental data confirmed that prediction accuracy depends on the matching depth of HOSE code spheres. It was also established that the dependence is an exponential

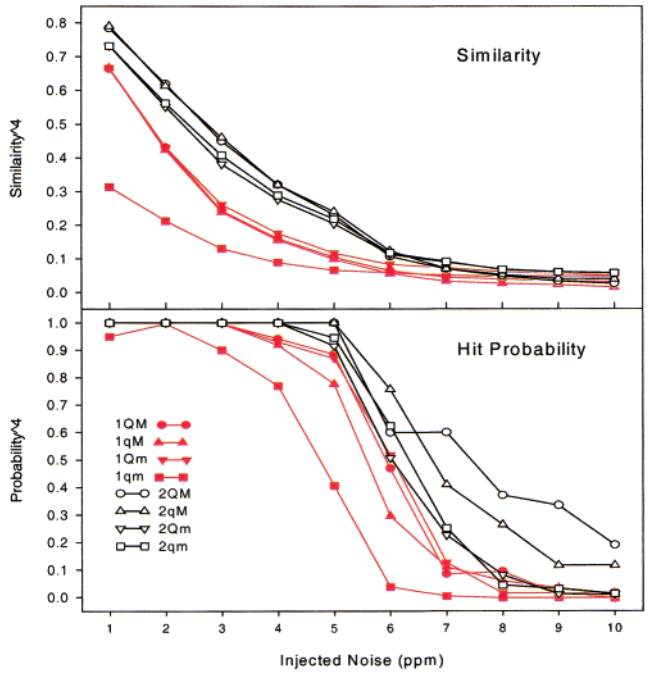


Figure 6. Summary results of Monte Carlo simulated queries: (a) averaged SI^4 for the first ranked matches in each search mode and (b) probability that the correct parent spectrum was found at the top of the ranking for each search mode.

function which asymptotically approaches a lower limit. It can be concluded that beyond seven spheres the dominant source of error originates from errors in the database such as, faulty transcription, publication errors, solvent/pH effects, and the bimodal distribution of shifts for stereo- and geometrical-isomers.

The match-depth score for a pair of HOSE codes provides an atom centric similarity index, which proved useful in estimating prediction errors in unassigned libraries. But the estimated errors should be viewed with some caution as they are expected to be systematically too low.

Monte Carlo simulated queries of a spectral database yielded useful insights regarding the overall performance of the database search-engine and could be applied generally to other NMR databases and search-engines. Ideally this would provide a uniform measure of performance so that different search algorithms and data-types could be compared. Plotting the SI^4 values for a large percentage of the hits returned from a query is not only useful for analysis of the Monte Carlo simulations but also would be useful in real queries for the purpose of grouping hits with similar spectra. The transition between success and failure to find the parent spectrum is a measure of the resolving power of the different query modes, and the noise level at which the transition is no longer discernible is a measure of the robustness of the query modes.

It appears that the added dimensionality of a query by the use of ^{1}H-^{13}C spectral data has only a modest but beneficial effect on search results. However, the most practical advantage of using 2D data comes from the significant savings in time afforded by indirect detection techniques. Furthermore, it was clearly demonstrated the least reliable searches are produced by 1D ^{13}C spectra with a low signal-to-noise ratios. This is consistent with the fact that those type of spectra product the fewest data points on which to search, and they give no information regarding the possible and probable absence of quaternary carbon data. With 1D-DEPT or ^{1}H-^{13}C spectra one can inform the search-engine that quaternary carbons are systematically absent. Indeed, when there are enough protonated carbons in the compound to compensate for lost quaternary carbon data, the HSQC spectrum should provide the most efficient route to dereplication of natural products by NMR because of its high data content relative to the time required to obtain those data.

ACKNOWLEDGMENT

We thank members of the Department of Natural Products Drug Discovery, Athanasios Tsipouras; the Biometrics Research Group, Bert Gunter; and the Molecular Systems Group, Simon Kearsley, for their help in this project.

REFERENCES AND NOTES

(1) *Dictionary of Natural Products on CD-ROM*; Chapman & Hall/CRC Press: London, 2000.
(2) Bremser, W. HOSE- A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.
(3) Patchkovskii, S.; Thiel, W. NMR Chemical Shifts in MNDO Approximantio: Parameters and Rsutls for H, C, N, and O. *J. Comput. Chem.* **1999**, *20*(12), 1220−1245.
(4) Bodenhausen, G.; Ruben, D. J. HSQC. *Chem. Phys. Lett.* **1980**, *69*, 185−188.
(5) Tsipouras, A.; et al. Using similarity searches over databases of estimated ^{13}C NMR spectra for structure identification of natural product compounds. *Anal. Chim. Acta* **1995**, *316*, 161−171.
(6) Balducci, R.; Pearlman, R. S. Efficient Exact Solution of the Ring Perception Problem. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 822−831.
(7) Grzonka, M.; Davies, A. N. Emperical Investigation on the Reproducibility of 13C NMR Shift Values. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1096−1101.
(8) Bendall, M. R.; Doddrell, D. M.; Pegg, D. T. DEPT. *J. Am. Chem. Soc.* **1981**, *103*, 4603−4605.
(9) Tsou, H.-R.; et al. Biosynthetic Origin of the Carbon Skeleton and Oxygen Atoms of the LL-F28249-alpha, A Potent Antiparasitic Macrolide. *J. Antibiotics* **1989**, *42*(3), 398−406.

CI010324M