Automated Processing of Two-Dimensional Correlation Spectra

Helmut Sengstschmid,* Heinz Sterk,* and Ray Freeman†

*Institut für Organische Chemie, Karl-Franzens-Universität Graz, Austria; and †Department of Chemistry,
Cambridge University, Lensfield Road, Cambridge, United Kingdom

An automated scheme is described which locates the centers of cross peaks in two-dimensional correlation spectra, even under conditions of severe overlap. Double-quantum-filtered correlation (DQ-COSY) spectra have been investigated, but the method is also applicable to TOCSY and NOESY spectra. The search criterion is the intrinsic symmetry (or antisymmetry) of cross-peak multiplets. An initial global search provides the preliminary information to build up a two-dimensional "chemical shift grid." All genuine cross peaks must be centered at intersections of this grid, a fact that reduces the extent of the subsequent search program enormously. The program recognizes cross peaks by examining the symmetry of signals in a test zone centered at a grid intersection. This "symmetry filter" employs a "lowest value algorithm" to discriminate against overlapping responses from adjacent multiplets. A progressive multiplet subtraction scheme provides further suppression of overlap effects. The processed two-dimensional correlation spectrum represents cross peaks as points at the chemical shift coordinates, with some indication of their relative intensities. Alternatively, the information is presented in the form of a correlation table. The authenticity of a given cross peak is judged by a set of "confidence criteria" expressed as numerical parameters. Experimental results are presented for the 400-MHz double-quantum-filtered COSY spectrum of 4-androsten-3,17-dione, a case where there is severe overlap.   © 1998 Academic Press

INTRODUCTION

Three very popular two-dimensional NMR experiments are COSY (correlation through scalar coupling), TOCSY (identification of an entire coupled spin system), and NOESY (a test for spatial proximity). However, the proliferation of more and more complex two-dimensional correlation spectra makes processing by hand a tedious and time-consuming procedure. The principal aim is to chart the correlations that determine the topology of the coupling network, or the through-space nuclear Overhauser interactions. A lesser goal is the extraction of those spin–spin coupling constants that are not readily accessible from the corresponding one-dimensional spectra. Processing the experimental data by computer frees the spectroscopist for the more important and complex task of *interpretation*.

Several schemes have already been proposed for auto-

matic processing of two-dimensional correlation spectra (*1–14*). These have been particularly successful in the study of spectra from very large molecules where the cross-peak structure tends to be rather simple. Complex interpenetrating multiplets in spectra from molecules of intermediate size present a much more severe challenge, and this is where attention is focused. The automatic data processing scheme described here relies heavily on the concept of a symmetry filter (*3, 5*), a device that examines the local symmetry over a suitable test zone and modifies the intensity of each pixel to force it to satisfy the symmetry requirements. One early implementation of this idea was used to attenuate t_1 noise in two-dimensional spectra by comparing intensities at two symmetrically related locations and retaining only the lower value (*15, 16*). Perhaps the most general description of the use of symmetry properties in two-dimensional spectra is that given by Hoch *et al.* (*5*) using the concept of projection operators. Symmetry filters possess an advantage over pattern recognition methods that employ a variable template, in that they are scale invariant and do not need to search for unknown coupling constants.

Symmetry Filters

The proposed new program is designed to process COSY or double-quantum-filtered COSY (DQ-COSY) spectra of small-to-medium sized molecules which have relatively narrow linewidths and where individual cross peaks can be quite complicated. It should also be applicable to TOCSY and NOESY, after allowing for the change from multiplet antisymmetry to symmetry. We consider the lower symmetry that one obtains in modified correlation experiments of the type E-COSY (*17*), z-COSY (*18*), or COSY-45 (*19*) to be less suitable for the symmetry filter approach.

Two-dimensional multiplets from spectra of the conventional COSY or DQ-COSY type possess D_2 symmetry. They can be considered a basic square pattern generated by the active splitting, with sign alternation of the intensities in both frequency dimensions, further split by in-phase passive splittings that do not alter the basic antisymmetry of the architecture. A typical experimental COSY cross peak of this kind is shown in Fig. 1; it has vertical planes of antisym-
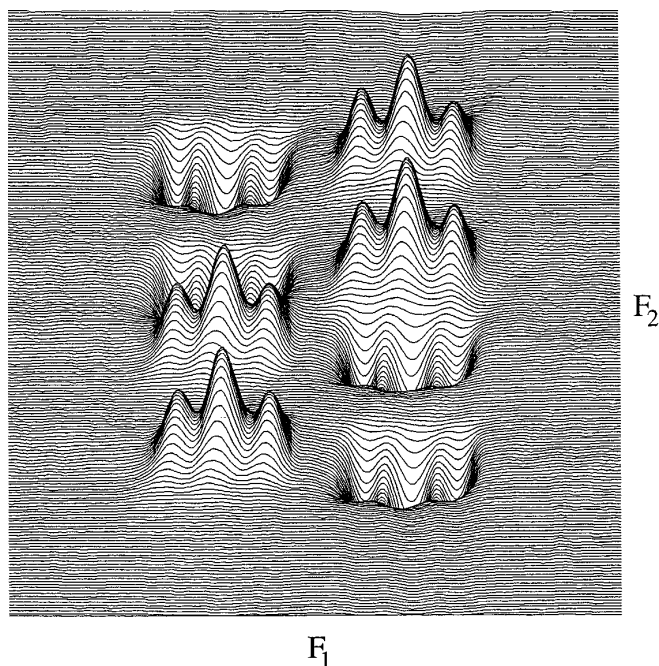
FIG. 1. A typical experimental 400-MHz COSY cross peak (24 × 24 Hz) from protons in nicotine with an active splitting of 7.9 Hz in both dimensions, and passive splittings of 2.2 and 1.7 Hz in the F_1 dimension and 4.8 Hz in the F_2 dimension. It has two planes of antisymmetry normal to the F_1 and F_2 axes, intersecting at the center.

metry normal to the F_1 and F_2 axes, intersecting at the center. We can notionally construct such a cross peak by multiplying together two orthogonal one-dimensional multiplets, each consisting of one antiphase splitting and any number of in-phase splittings (*20*). The symmetry of the diagonal peaks of a COSY spectrum is irrelevant for the present program because a prespecified region straddling the principal diagonal is excluded from the search routine.

THE COMEX PROGRAM

The main features of the new program COMEX (*CO*SY *m*ultiplet *ex*traction) are set out in Fig. 2 in the form of a simplified flowchart. Several other approaches were explored before it was decided to follow this particular processing scheme. The choice was based on the principle that two consecutive search routines using different symmetry criteria are probably more powerful than a single search, and that the main search should be restricted to the intersections of a ''chemical shift grid'' constructed from a relatively rapid determination of chemical shifts in an initial global ''strip search.''

Concerns about overlap are partly answered by the use of a lowest value algorithm (*15, 16*) that rejects responses that fail to meet the symmetry criterion (*21, 22*). Residual overlap and subsidiary centers of symmetry are addressed by

progressively subtracting the strongest two-dimensional multiplet from the experimental data, thus removing both types of artifact. This is a technique first used by radioastronomers (*23*) and later applied to two-dimensional NMR spectroscopy (*24*).

Below we examine the main stages of the COMEX program as illustrated in the simplified flowchart of Fig. 2.

PREPROCESSING

The search routines can be speeded up to some extent by preprocessing the raw experimental data so as not to waste time examining regions where no signals appear. In cases where very weak cross peaks might be anticipated, this stage would be omitted on the grounds that it is likely to overlook such marginal responses. This routine is essentially a cluster analysis (*4–6*) which designates a roughly rectangular area around any close group of lines as a potential cross peak and builds up a ''search mask'' that can be superimposed on the experimental spectrum to limit the area to be investigated.

This algorithm operates by allocating a value of unity
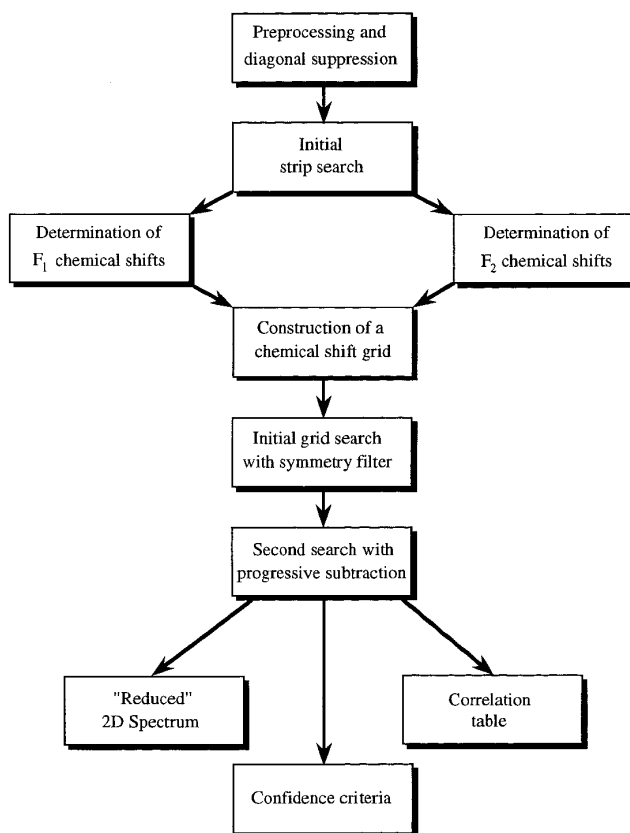
FIG. 2. A schematic flowchart of the COMEX program for automatic processing of two-dimensional correlation spectra. The ''reduced'' 2D spectrum shows cross peaks with no spin–spin splittings.

("white") to each pixel with an intensity exceeding a pre-scribed threshold, and zero ("black") to all other pixels. Then this two-dimensional array of white and black pixels is convoluted with a white square the size of a typical cross peak, so that all the white pixels are included within one of several white clusters. This routine is called the *area-filling algorithm.* Search routines are restricted to the white regions.

An alternative scheme, called the *minimum-area algorithm,* reduces the boundary region surrounding each multiplet and thus speeds up the subsequent search routine. It starts with the same two-dimensional array of black and white pixels but, instead of convolution with a white square, it fills in any intervals between white pixels by converting black to white, as long as the frequency separations are less than a prescribed value, of the order of the largest spin–spin coupling constant The resulting search mask is smaller than that generated by the area-filling algorithm.

The choice of the intensity threshold is important—too high a value introduces the possibility of overlooking weak signals, and too low a value picks up artifacts and noise, and extends the search mask to almost entire spectrum. Normally the threshold would be related to the peak-to-peak noise level measured in a region known to be empty of all signals.

Although diagonal peaks in a double-quantum-filtered COSY spectrum are less obtrusive than in conventional COSY, the fact that they can be of mixed phase is a decided disadvantage in an automated processing program. Furthermore, in the region close to the diagonal there are occasionally cross peaks from strongly coupled protons, with their well-known distortions of frequencies and intensities. It is therefore prudent to exclude from the search mask a band of pixels straddling the principal diagonal, thereby accepting a slight risk that a chemical shift might be missed altogether.

THE INITIAL STRIP SEARCH

The next stage sets out to determine chemical shifts in the F_1 and F_2 dimensions through a relatively rapid examination of the experimental spectrum $S(F_1, F_2)$ limited to the white areas of the search mask. COMEX makes use of the fact that cross peaks involving coherence transfer from a spin with a shift δ_1 in the F_1 dimension will all share the same coordinate in F_1. The search is implemented by examining the data within a long narrow test strip (Fig. 3). This strip runs the entire length of the F_2 axis and is just wide enough to encompass the largest spin multiplet (typically 50 Hz). The program examines a short search trace and tests for antisymmetry of intensities with respect to the middle of the trace. In practice this antisymmetry is disturbed by signal overlap from adjacent multiplets, and by any artifacts or noise.
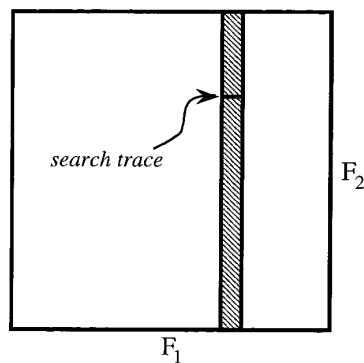
FIG. 3. The initial strip search to discover the chemical shifts in the F_1 dimension. A short search trace is examined for antisymmetry with respect to its center. This is repeated for the entire strip (shaded) and then the strip is moved in horizontal steps until the entire two-dimensional array has been checked. The process is repeated with F_1 and F_2 interchanged, giving chemical shifts in the F_2 dimension.

Similarity Algorithm

The similarity test represents the left half of a typical horizontal search trace by a vector L and the right half by a vector R, each starting at the center of the trace and running toward the ends. The degree of antisymmetry may then be written as the vector product (2)

$$\sigma = -\frac{L \cdot R}{|L| \ |R|}. \qquad [1]$$

The parameter σ only approaches $+1$ if there is a high degree of antisymmetry between the two halves of the search trace, it is equal to -1 for complete symmetry, and it is zero where there is no symmetry. This value of σ is entered at the frequency coordinates defining the center of the trace. The process is repeated for each horizontal trace, and then again for all possible vertical strips, creating a similarity map covering the entire two-dimensional array $\sigma(i, j)$.

Fuzzy Logic Algorithm

The similarity parameter $\sigma(i, j)$ normally lies between zero and unity; it does not reflect the relative intensity associated with a given center of symmetry. A measure of intensity is obtained by processing each search trace with a symmetry filter, since this largely conserves the intensity of the constituent resonances. We found it useful to adopt a "fuzzy logic" variation of the lower-value algorithm by defining an adjustable parameter $0 \geq D \geq 1$. This works as follows: a pairwise comparison is made between ordinates symmetrically located with respect to the center of the search trace. If these intensities have the same sign, both are replaced by zero. If they have opposite signs, they are both replaced by the lower (absolute) intensity, retaining the original signs. This is the "hard" version of the algorithm ($D = 1$) and it has the

property that it strongly suppresses overlapping signals from adjacent multiplets. On the other hand, it is intolerant of multiplet distortions due to overlap, poor digitization, or strong coupling effects, and this could adversely affect the final intensity. The other extreme version of the algorithm ($D = 0$) leaves both intensities unchanged, thus mitigating these difficulties, but allowing many artifacts to pass the filter. The intermediate ''soft'' version of the test reduces the higher (absolute) intensity by D times the difference between the higher and lower (absolute) intensities. In this manner a compromise is struck between a very strict and an overtolerant filter. The ''fuzziness parameter'' D is chosen by the operator and usually lies between 0.5 and 0.7. The integral of the absolute values of the intensities over the entire search trace is represented as $I(i, j)$, the total intensity at the frequency coordinates of the center of the trace.

The Two-Dimensional Symmetry Map

This process is repeated for each search trace along the strip, and then, by moving the strip stepwise in the F_1 dimension, for all points in the array $S(F_1, F_2)$. The intensity parameter $I(i, j)$ is then multiplied by the symmetry parameter $\sigma(i, j)$ at each location, building up a two-dimensional map $M(i, j)$. The projection of $M(i, j)$ onto the F_1 axis has positive peaks at each chemical shift value, symmetrically flanked by weaker negative peaks representing symmetric subpatterns (subsidiary centers of symmetry) of the spin multiplets (Fig. 4b). The negative peaks reflect the fact that $\sigma(i, j)$ is negative for a symmetrical subpattern.

Subsidiary Centers of Symmetry

A typical spin multiplet contains not only the principal center of antisymmetry but also some subsidiary centers of antisymmetry (or symmetry) arising from a subset of the total number of resonance lines. Subsidiary centers lie on either side of the principal center, separated by a J coupling, and have at most only 50% of the intensity of the principal response. This difference in intensity is used as a means of eliminating responses from subsidiary centers of symmetry. Any ordinate in the map $M(i, j)$ is set to zero if there is a more intense ordinate within a range $\pm\Delta$ Hz, where Δ is a predetermined parameter of the order of a typical spin–spin coupling constant. This leaves only the principal centers of antisymmetry. It also replaces the linewidths by delta functions at the chemical shift frequencies (Fig. 4c).

The entire process is then repeated with strip searches parallel to F_1 in order to obtain chemical shift values in the F_2 dimension. Owing to various imperfections in the raw data, and shortcomings in the data processing, the F_1 and F_2 chemical shift spectra may not be identical. For example, the initial search program may sometimes discover an artifact in the F_2 chemical shift spectrum that has no counterpart in the F_1 spectrum. This is not a serious problem since it merely

creates additional intersections on the chemical shift grid, at locations where there are no genuine cross peaks. The subsequent search routine eliminates these possibilities. In cases of doubt, it is better to include rather than exclude questionable chemical shift values. If a shift appears in the F_1 dimension but not in the F_2 dimension it is advisable to add it to the F_2 list so as not to miss a possible cross peak.

THE CHEMICAL SHIFT GRID

Knowledge of the chemical shift values permits the construction of a two-dimensional chemical shift grid, made up by drawing vertical and horizontal lines through the chemical shifts in the F_1 and F_2 dimensions. All cross peaks in the correlation spectrum must be centered on intersections of this grid. This leads to an enormous reduction in the time expended by the subsequent search algorithm. For example, a conventional global search would need to examine almost k^2 possible locations for cross-peak centers, where k^2 is the number of data points defining the two-dimensional spectrum. On the other hand, a grid search reduces this to a much smaller number $n(n - 1)$ of possible centers, where n is the number of chemical shifts. The ratio $k^2/[n(n - 1)]$ would typically lie in the range 10^3 to 10^4.

Grid Search

All off-diagonal grid intersections are candidates for the center of a cross peak. A square test zone is extracted from the experimental correlation spectrum centered at one of the grid intersections. Typically, an initial test zone might be a 50×50-Hz square, selected to encompass any possible two-dimensional multiplet. As in the initial strip search, two tests are performed, a two-dimensional similarity test to give a measure of $\sigma(i, j)$, and a symmetry filter in combination with integration to give a measure of the intensity $I(i, j)$.

Similarity Test

The similarity test is in two parts. We select a square test zone of side $2m$ units centered at an intersection of the chemical shift grid at the coordinates (i, j) (Fig. 5a). First a typical horizontal trace is checked for antisymmetry with respect to its center $(i, j + l)$ using the vectorial expression similar to Eq. [1]. This operation is repeated for all horizontal traces within the test zone and the results are summed together to give the similarity parameter for the F_1 dimension,

$$\sigma(F_1)$$
$$= -\sum_{1=-m}^{m} \frac{\sum_{k=1}^{m} Z(i + k, j + l)Z(i - k, j + l)}{\sum_{k=1}^{m} Z^2(i + k, j + l) \sum_{k=1}^{m} Z^2(i - k, j + l)},$$

$$[2]$$
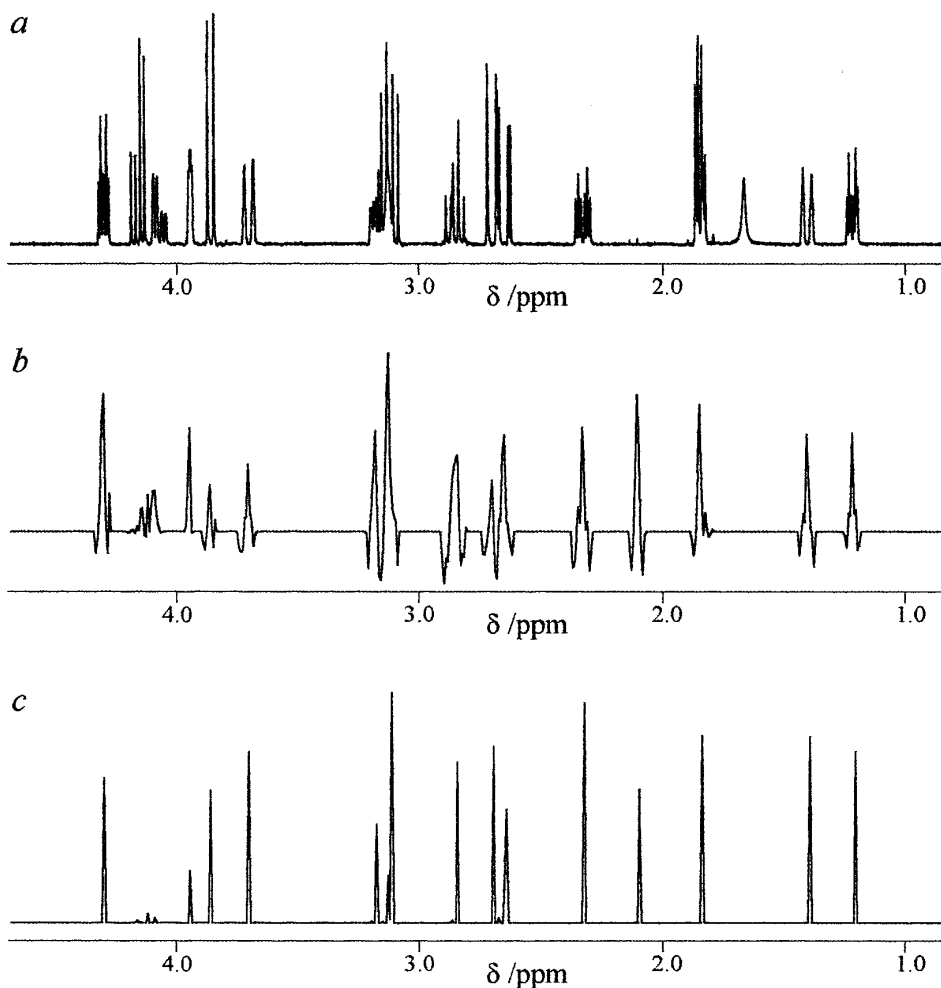
FIG. 4. (a) Part of the conventional 400-MHz spectrum of strychnine. (b) Projection of the symmetry map obtained at the end of the strip search. In addition to the peaks at chemical shift values there are smaller negative peaks at each side, attributable to subsidiary centers of symmetry. (c) Projection of the symmetry map after removal of the subsidiary centers.

where $Z(i + k, j + l)$ represents the intensity at a pixel with the general coordinates $(i + k, j + l)$.

The process is then repeated for all vertical traces (Fig. 5b) and the results are summed together to give the similarity parameter for the F_2 dimension:

$$\sigma(F_2) = - \sum_{k=-m}^{m} \frac{\sum_{l=1}^{m} Z(i + k, j + l)Z(i + k, j - l)}{\sum_{l=1}^{m} Z^2(i + k, j + l) \sum_{l=1}^{m} Z^2(i + k, j - l)}.$$

[3]

The overall measure of similarity is given by M_{sim} defined as the product $\sigma(F_1)\sigma(F_2)$ if $\sigma(F_1)$ and $\sigma(F_2)$ are both positive; otherwise $M_{sim} = 0$. This similarity parameter has a range of values between zero and unity; later it is used to multiply an intensity factor derived from a symmetry filter.

Symmetry Filter

The symmetry filter employs the same test zone centered at (i, j). It examines the intensities at pixels located at the four frequency coordinates $(i + k, j + l)$, $(i - k, j + l)$, $(i + k, j - l)$, and $(i - k, j - l)$. If these intensities satisfy the antisymmetry rules,

$$Z(i + k, j + l)Z(i - k, j + l) < 0$$
$$Z(i + k, j + l)Z(i - k, j - l) > 0$$
$$Z(i - k, j + l)Z(i + k, j - l) > 0,$$

[4]

then the lowest (absolute) intensity is placed at all four locations, retaining the original signs associated with these locations; otherwise zero is placed at all four locations. This is the hard symmetry filter ($D = 1$). A soft filter reduces
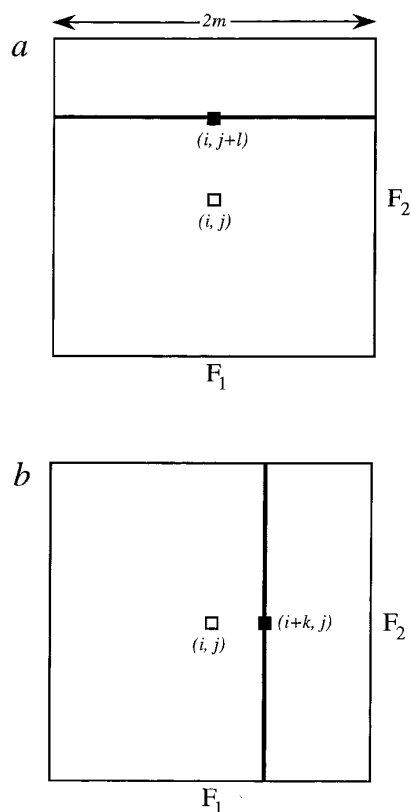
FIG. 5. (a) The similarity test examines the antisymmetry of a typical trace with respect to its center $(i, j + l)$ and repeats this calculation for all horizontal traces within the test zone, summing the results to give the parameter $\sigma(F_1)$ of Eq. [2]. (b) An analogous process is carried out for vertical traces, giving a corresponding parameter $\sigma(F_2)$ of Eq. [3]. The product $\sigma(F_1)\sigma(F_2)$ is the similarity measure M_{sim} unless either $\sigma(F_1)$ or $\sigma(F_2)$ is negative, in which case $M_{\mathrm{sim}} = 0$.

the intensity at each location by D times the difference between the modulus of the actual intensity and the modulus of the lowest value. Setting $D = 0$ degenerates the procedure into a mere sign check. Typically D was set near 0.5 for this application. This procedure is repeated for all sets of symmetrically related pixels in the test zone. The result is a symmetrized version of the test zone, $Z^*(i + k, j + l)$.

This symmetrized two-dimensional array $Z^*(i + k, j + l)$ may or may not represent a genuine cross peak, and the principal criterion for making this distinction is the integral of the absolute intensities over the entire test zone multiplied by the similarity parameter M_{sim}. This is compared with the corresponding measure taken over a comparable zone in an empty region of the spectrum. The integral is largely determined by the degree of mutual interference between antiphase signals, a function of the ratio of the active coupling divided by the linewidth. Small couplings generate relatively weak cross peaks, and if the coupling is much smaller than the linewidth, the cross peak is undetectable. Identification of a cross peak as ''authentic'' requires that

the integrated intensity exceed the user-defined threshold intensity. In rare cases it may still represent an artifact caused by a subsidiary (or accidental) center of symmetry that happens to fall at a grid intersection. The resolution of such ambiguities relies on a set of ''confidence criteria'' to be discussed later.

Shrink-to-Fit

In practice an extravagantly large test zone may prove undesirable if it includes extraneous signals from adjacent multiplets. On the other hand, a too-conservative choice for the size of the test zone may sacrifice intensity from the multiplet under investigation. For such cases a ''shrink-to-fit'' algorithm can be employed to adapt the size of the test zone to the size of the experimental two-dimensional multiplet. Starting with the standard (large) test zone, the routine applies the symmetry filter and evaluates the integral over the zone. Then it contracts the zone slightly and recalculates the integral, repeating the contraction cycle in each frequency dimension until the integral is reduced to about 80%, indicating that the perimeter of the test zone is just skirting the edges of the two-dimensional multiplet and clipping a small amount of intensity. While this procedure is probably unnecessary for most multiplets, it can be important for cross peaks with an extreme aspect ratio, such as those from long-range couplings to a methyl group.

Progressive Subtraction of Multiplets

At this stage the results are bedeviled by two serious problems—subsidiary centers of symmetry formed from subsets of the resonance lines that make up a given spin multiplet, and accidental centers of symmetry generated by the close approach of two different multiplets. In most cases these artifacts are not centered on grid intersections and are therefore rejected, but it is important to eliminate the remaining ''accidents.'' The technique used is ''progressive subtraction,'' a procedure adapted from radioastronomy (23). It operates on the results of the initial grid search, made up of responses from genuine cross peaks and from artifacts. The symmetrized two-dimensional multiplet with the most intense integral is subtracted from the remainder of the spectrum. Choosing the strongest response essentially guarantees that it is a genuine cross peak. In this operation it is advisable to avoid too much encroachment on adjacent multiplets that might perturb their intensities, so the shrink-to-fit procedure is favored. Removal of a multiplet automatically removes its subsidiary centers of symmetry and any accidental centers involving that multiplet. After subtraction of the strongest cross peak, all nearby multiplets are reprocessed, noting whether there is any reduction of intensity caused by the removal of overlapping lines. The process is then repeated by subtracting the next strongest multiplet, and so on until all the ''occupied'' grid intersections have
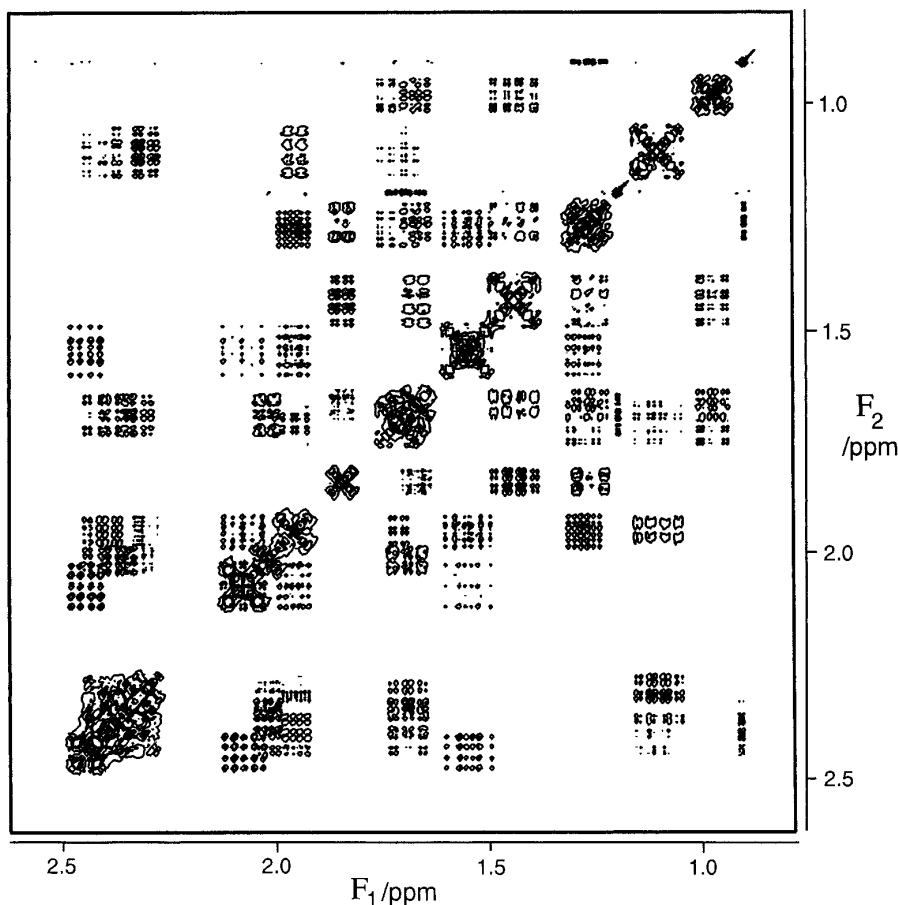
FIG. 6. Part of the unprocessed experimental 400-MHz DQ-COSY spectrum of 4-androsten-3,17-dione, the starting point of the COMEX program. Note the interpenetrating cross peaks from protons near 2.0 and 2.4 ppm.

been reexamined. The result is a much more reliable indication of the correlations.

The symmetry filter for this operation is the soft version with a low value of D (typically 0.3) in order to allow for any distortion of the multiplets through overlap effects. Slightly better results have been achieved in test cases by a modified symmetry filter that finds the best three of the four pixels used in the symmetry test (*21*). Here "best" means the three pixels that most closely satisfy the symmetry test; the intensity of the fourth pixel is assumed to be falsified by overlap effects and it is replaced by the lowest (absolute) intensity of the other three pixels, retaining the original sign. Although such a filter allows artifacts to slip through relatively easily, and should not be used in the initial search routine, it is well suited to the task of progressive subtraction.

The integral of the absolute intensities over the test zone, multiplied by M_{sim}, is then represented as a single positive ordinate at the center of symmetry, taken to indicate the chemical shift coordinates. This correlation "peak" is essentially a delta function, being the pixel that best satisfies the symmetry test. No linewidths (natural or instrumental) are

involved. In cases of strong coupling, there could be a slight discrepancy between the center of symmetry and the chemical shift, as is well known for AB spectra in conventional spectroscopy, but this is unavoidable. Many strongly coupled features appear in responses close to the principal diagonal, a region that is excluded from the COMEX search.

PRESENTATION OF RESULTS

Reduced Correlation Spectra

There is something rather reassuring about a two-dimensional correlation spectrum displayed in the form of an intensity contour plot. It is a clear, direct way to indicate correlations. Clarity is significantly enhanced if all cross-peak multiplets are reduced to a singlet at the chemical shift coordinates. It also helps sensitivity by gathering into a single peak all the intensity that was originally distributed throughout the multiplet. One output of COMEX is a "reduced" correlation spectrum where each cross peak is indicated by a small circle or square at the appropriate chemical
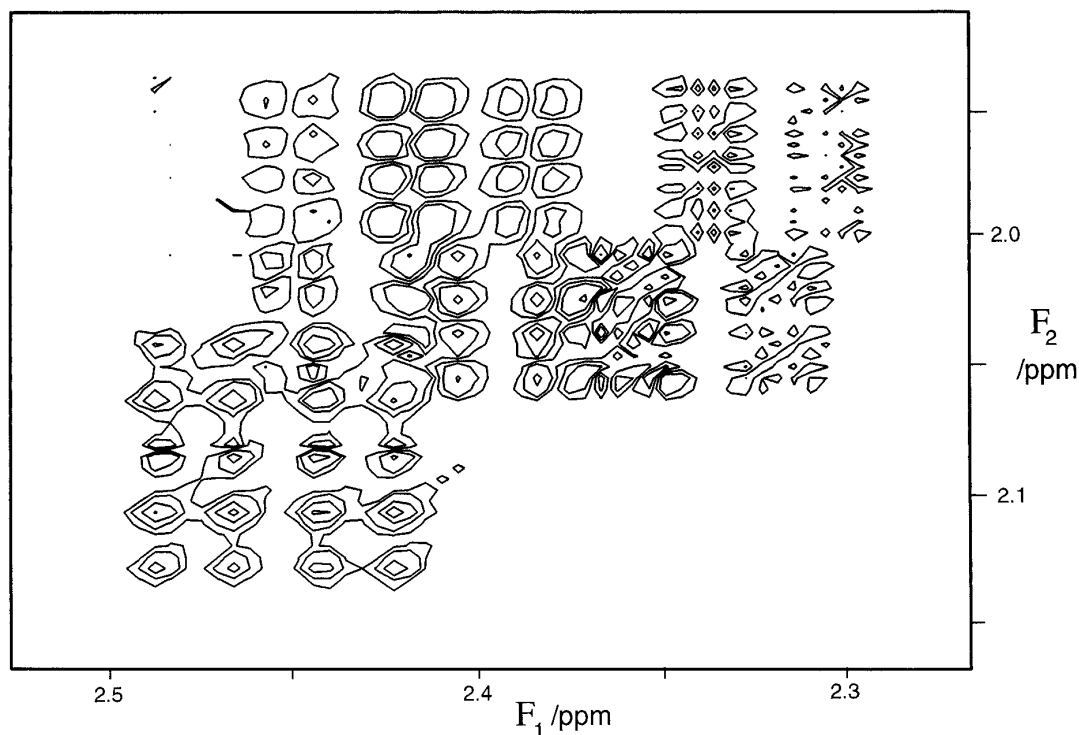
FIG. 7. A crowded region of five interpenetrating multiplets arising from protons near 2.0 and 2.4 ppm in the DQ-COSY spectrum of 4-androsten-3,17-dione.

shift coordinates. If required, a rough indication can be given of the relative cross-peak intensities, for example, by displaying filled, shaded, or open squares to represent strong, medium, or weak integrals.

Note that spin–spin coupling information is not lost. COMEX separates all the two-dimensional cross peaks, and the appropriate sections give the one-dimensional multiplets. Alternatively these can be obtained by a projection procedure adapted to deal with patterns of antiphase signals. The two-dimensional multiplets have been symmetrized and they are isolated from any previously overlapping responses. They provide a good starting point for the measurement of the coupling constants, even for poorly resolved splittings (*25*).

Chemical Shift Spectra

Almost as a by-product of the program, a one-dimensional spectrum of chemical shifts is obtained, with no spin–spin splittings. This has long been a goal of NMR spectroscopists ever since the pioneering article by Aue *et al.* (*26*) based on the idea of 45° projection of two-dimensional *J* spectra. The COMEX program lists the centers of all the cross peaks, and these frequencies can be taken as chemical shifts, provided that strong coupling effects can be neglected. A chemical shift spectrum may then be constructed. Since the relative intensities of the cross peaks are appreciably perturbed by destructive interference arising from the antiphase nature of

the active splitting, the simplest procedure is to assign unit intensity to all the peaks in the chemical shift spectrum.

Correlation Tables and Confidence Criteria

A time may well come when the two-dimensional correlation spectrum itself is no longer required for publication purposes, the essential information being presented in a table that lists all the chemical shifts in one column, with the shifts of correlated sites in a second column. The first column may be conveniently ordered in terms of increasing chemical shielding. The important question here is the degree of certainty that a given correlation is authentic. Cross peaks become unreliable when the active coupling constant is small compared with the linewidth because of mutual cancellation of antiphase signals. Consequently the correlation table should also contain ''confidence criteria'' indicating how certain we can be that a given cross peak is genuine.

The prime criterion is the relative intensity (the integral of the absolute intensities within the test zone) since this reflects the main cause of intensity loss—the self-cancellation of antiphase multiplet components. Sometimes a very crude indication of intensity would be sufficient, such as strong, medium, or weak. More generally, the intensity *M* is expressed as a percentage of the intensity of the strongest cross peak in the spectrum. Responses are excluded as unreliable if they fall below a predefined threshold integral. This
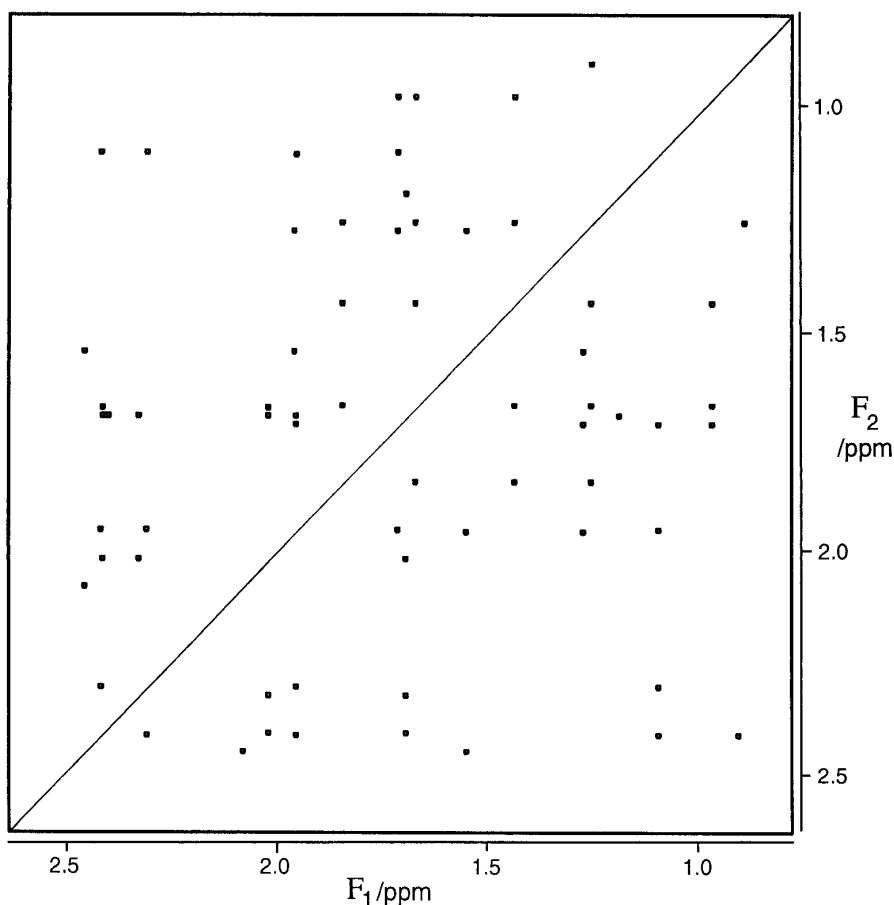
FIG. 8. The reduced two-dimensional DQ-COSY spectrum of 4-androsten-3,17-dione, processed by COMEX, showing 65 correlation peaks represented by small squares at the chemical shift coordinates.

is normally set just above the integral measured over a standard test zone in an empty region of the spectrum.

There are also some supplementary criteria which can act as guides to authenticity but allow no absolute conclusions to be drawn. One is the overlap number N, defined as the number of adjacent multiplets that are so close that their test zones intersect. If $N = 0$ the multiplet is completely isolated, and this is a good indication that the corresponding correlation is probably genuine. Another criterion is the subtraction ratio R, defined as the ratio of integrals after and before the subtraction stage of the final grid search. If $N = 0$, then the intensity of a given response should not change upon subtraction of any other multiplet, and R should be close to 100%. If R is below about 85% it is an indication that this particular feature is either an artifact or a severely distorted cross peak. A related criterion is the symmetrization ratio S, defined as the integral over the symmetrized test zone divided by the integral over the unmodified test zone. If S is a high percentage, it indicates a high degree of confidence that this cross peak is genuine; a low value is cause for scepticism.

There are two further qualitative tests that can be applied. One is the well-known check to see whether a putative cross peak has its expected counterpart symmetrically positioned on the opposite side of the diagonal. This global symmetry test may break down if folding has been permitted in one or both frequency dimensions, or in regions where there is t_1 noise. It is also undermined if different digitization steps or different spectral windows are used in the two frequency dimensions. Unfortunately some artifacts also satisfy this symmetry test if they arise from subsidiary symmetry centers or from accidental overlap. The second test is the geometrical structure of the symmetrized cross peak; two-dimensional multiplets have characteristic patterns that are readily recognized by an experienced operator, whereas artifacts often have bizarre structures.

Clearly it is not always possible to make a hard-and-fast decision about the authenticity of a given correlation, although if all of these confidence criteria are taken into account, ambiguities should be minimized. In molecular structure investigations both false positive and false negative correlations could raise serious problems. What COMEX

does is remove the subjectivity associated with conventional processing by inspection. It supplies a numerical estimate of the relative intensity M of a given cross peak, together with the N, R, and S confidence parameters which act as further guides. It would certainly be possible to write a subroutine that weighed up all these criteria and then made its best estimate of authenticity, although many spectroscopists would probably prefer to make this judgment for themselves. In critical cases it may prove necessary to set up a fresh correlation experiment to confirm the presence of a dubious cross peak.

EXPERIMENTAL TEST

We chose as an example a proton correlation spectrum that includes a significant number of interpenetrating multiplets. This is the crowded region of the double-quantum-filtered COSY spectrum of 4-androsten-3,17-dione dissolved in CDCl$_3$ recorded on a Varian VXR-400 spectrometer. The raw experimental spectrum is shown in Fig. 6. It was acquired with 256 complex data points in the evolution (t_1) dimension and 1024 data points in the acquisition (t_2) dimension. The spectral width was 964 Hz in both frequency dimensions. A Gaussian sensitivity enhancement function was applied in the time domain, giving an additional broadening of the lines of 0.33 Hz. A nonspinning sample was used at 25°C and the total data gathering operation lasted 7.5 h. The spectrum in the form of a 512 × 512 square array was processed by COMEX in 1.5 min on a Sun Classic SPARC-station IPC.

There are two obvious regions of severe overlap in the conventional high-resolution spectrum, centered around 2.0 ppm (four protons) and 2.4 ppm (five protons). Figure 7 shows a cluster of five overlapping cross peaks from the double-quantum-filtered COSY spectrum; each individual multiplet contains between 24 and 64 component lines. This is a situation where assignment by hand can become very tedious indeed, and it provides a reasonably demanding test for COMEX.

The Chemical Shift Grid

The chemical shifts in F_1 and F_2 were determined by the strip search method described above, using the similarity routine to evaluate $\sigma(i, j)$ and the lowest value algorithm to determine $I(i, j)$. A two-dimensional map of the product $M(i, j) = \sigma(i, j)I(i, j)$ was then used to obtain chemical shift values by projection onto the F_1 and F_2 axes. These two projections are very similar except for markedly different intensities for the response at 1.11 ppm, and each gave 25 responses above the prespecified threshold. Of these, only three later turned out to be artifacts. The resulting chemical shift grid had $600 = 25 \times 24$ off-diagonal intersections that were candidates for centers of cross peaks.

TABLE 1
Frequency Coordinates of Completely Isolated Cross Peaks Found in the DQ-COSY Spectrum of 4-Androsten-3,17-dione

$\delta(F_1)$ (ppm)	$\delta(F_2)$ (ppm)	N	M (%)	R (%)	S (%)
1.967	1.284	0	100.0	100	68
1.269	1.856	0	91.7	100	75
1.284	1.967	0	90.7	100	72
1.856	1.269	0	84.1	100	68
1.109	2.318	0	80.9	100	68
2.318	1.109	0	78.6	100	70
1.448	1.856	0	66.0	100	67
1.109	1.964	0	62.2	100	76
1.856	1.448	0	60.1	100	69
1.964	1.109	0	56.0	100	74
1.560	2.464	0	53.4	100	68
2.464	1.560	0	50.8	100	66
1.448	1.681	0	47.1	100	64
1.967	1.560	0	43.7	100	64
1.560	1.284	0	41.9	100	63
1.560	1.967	0	40.4	100	62
1.681	1.448	0	32.9	100	54
1.448	0.980	0	26.7	100	64
1.269	0.905	0	25.5	100	62
1.284	1.560	0	25.1	100	56
0.905	1.269	0	24.9	100	63
0.980	1.448	0	22.2	100	56
1.109	1.722	0	21.9	100	68
2.318	2.432	0	20.0	100	31
1.109	2.432	0	19.2	100	51
2.432	2.318	0	17.5	100	29
1.681	1.856	0	16.6	100	49
1.856	1.681	0	15.7	100	49
1.722	1.109	0	15.1	100	56
2.432	1.109	0	14.5	100	44
0.919	2.432	0	8.3	100	43
1.448	1.269	0	4.3	100	27
1.269	1.448	0	3.3	100	30

Note. N is the overlap number. M is the intensity relative to the strongest cross peak. R is the ratio of intensities after and before subtraction. S is the ratio of intensities after and before symmetrization.

Grid Search

In the initial grid search, the chemical shift grid was examined by the similarity algorithm and the fuzzy lowest value algorithm ($D = 0.5$) was used to obtain the relative intensities of the responses. A second search with the three-point lowest value algorithm ($D = 0.3$) combined with the progressive subtraction technique minimized artifacts from accidental and subsidiary centers of symmetry. A total of 65 two-dimensional multiplets were identified, 34 above and to the left of the principal diagonal and 31 below and to the right (Fig. 8). Their relevant parameters are listed in Table 1 for responses with an overlap number zero, and in Table 2 for responses involving some degree of overlap and consequently some degree of uncertainty.

TABLE 2
Frequency Coordinates of Partially Overlapped Cross Peaks Found in the DQ-COSY Spectrum of 4-Androsten-3,17-dione

$\delta(F_1)$ (ppm)	$\delta(F_2)$ (ppm)	N	M (%)	R (%)	S (%)
1.704	2.338	1	79.1	100	63
2.338	1.704	3	76.6	100	63
2.464	2.090	1	69.0	100	58
2.090	2.464	1	63.4	100	57
2.031	1.704	3	62.1	100	57
1.704	2.031	1	57.5	100	55
2.423	1.964	2	46.5	100	63
1.964	2.432	2	45.9	100	62
1.681	0.980	1	34.1	100	56
0.980	1.681	1	33.9	107	50
1.681	1.269	2	33.8	105	45
1.269	1.681	2	31.6	133	41
0.980	1.722	1	26.8	100	32
2.031	2.419	3	26.7	98	33
1.722	1.284	2	23.4	100	41
1.284	1.722	2	21.4	100	32
1.722	0.980	1	20.6	78	38
2.031	2.338	3	19.0	100	32
1.704	2.419	1	18.6	100	46
2.419	1.704	3	18.5	100	47
2.338	2.031	3	17.6	100	31
1.964	1.722	3	14.7	89	43
1.722	1.964	1	14.5	94	39
1.201	1.704	2	13.9	94	47
2.318	1.964	1	12.8	101	42
1.964	2.318	1	12.2	98	40
2.419	2.031	3	11.3	97	31
1.704	1.201	2	10.5	86	36
2.031	1.687	3	5.7	82	20*
2.407	1.704	3	4.9	83	20*
1.964	1.704	3	3.6	116	18*
2.419	1.687	3	3.0	194	18*

*Note. N is the overlap number. M is the intensity relative to the strongest cross peak. R is the ratio of intensities after and before subtraction. S is the ratio of intensities after and before symmetrization. *Denotes artifacts.*

Levels of Confidence

Clearly we can be quite confident about all the correlations in Table 1 because all represent isolated multiplets ($N = 0$) with the subtraction ratio $R = 100\%$ throughout. Our interpretation of Table 2 is that all the correlations are genuine except the four marked with asterisks. These four ''false positive'' responses all have very low relative intensities M (below 6%), an overlap number of 3, and very low symmetrization factors S (20% or less). For the multiplets recognized to be genuine, the lowest M value was 10.5% and the lowest S value was 27%. In this example, the subtraction ratio R proved not to be a useful guide, in two cases exceeding 100%, probably due to statistical errors on the very low recorded intensities. It is only in these few cases that a judgment by the spectroscopist is necessary. If compelled to

generalize, we would suggest that the ambiguous cases would form a group with $N \neq 0$, $M < 10\%$, $R < 90\%$, and $S < 30\%$; the specific cases within this group would then be judged individually.

On the other hand, there are also four ''false negative'' correlations that were missed by COMEX because their responses fell below the specified intensity threshold. The evidence for this assertion comes from superimposing the reduced correlation spectrum (Fig. 8) over the raw experimental COSY spectrum (Fig. 6) and noting two additional pairs of cross peaks in Fig. 6. One of these correlations (between protons at 2.09 and 1.97 ppm) lies quite close to the diagonal and has its two multiplets severely distorted by strong coupling effects that reduce the local symmetry. The second false negative correlation (between protons at 2.09 and 1.56 ppm) also shows severely distorted multiplets attributed to coupling to two strongly coupled neighbors. These shortcomings of the program emphasize the dangers associated with strong coupling. There appears to be no simple solution to this problem short of running the spectrum a second time at a higher spectrometer field.

CONCLUSIONS

No computer program for reducing two-dimensional correlation spectra is likely to perform better than a skilled spectroscopist working by inspection. On the other hand, automated processing can relieve a great deal of tedium and can offer numerical criteria for the degree of confidence in the authenticity of a given correlation, removing the subjectivity inherent in conventional methods of assignment. Even in the most trying cases, COMEX can serve as a preliminary examination, guiding the spectroscopist in a conventional analysis of an overcrowded region of the spectrum. For more routine applications, a simple table showing the correlations is probably quite adequate, and it may not be necessary to display the actual spectrum. Judged on the basis of its performance on the proton spectrum of 4-androsten-3,17-dione, COMEX appears to fulfill these requirements.

REFERENCES

1. B. U. Meier, G. Bodenhausen, and R. R. Ernst, *J. Magn. Reson.* **60,** 161 (1984).
2. P. H. Bolton, *J. Magn. Reson.* **67,** 391 (1986).
3. P. H. Bolton, *J. Magn. Reson.* **70,** 344 (1986).
4. P. Pfändler and G. Bodenhausen, *J. Magn. Reson.* **70,** 71 (1986).
5. J. C. Hoch, S. Hengyi, M Kjaer, S. Ludvigsen, and F. M. Poulsen, *Carlsberg Res. Commun.* **52,** 111 (1987).
6. Z. Mádi, B. U. Meier, and R. R. Ernst, *J. Magn. Reson.* **72,** 584 (1987).
7. S. Glaser and H. K. Kalbitzer, *J. Magn. Reson.* **74,** 450 (1987).
8. B. U. Meier, Z. L. Mádi, and R. R. Ernst, *J. Magn. Reson.* **74,** 565 (1987).
9. Z. L. Mádi and R. R. Ernst, *J. Magn. Reson.* **79,** 513 (1988).

10. K. P. Neidig and H. R. Kalbitzer, *Magn. Reson. Chem.* **26,** 848 (1988).

11. K. P. Neidig and H. R. Kalbitzer, *J. Magn. Reson.* **88,** 155 (1990).

12. K. P. Neidig, R. Saffrich, M. Lorenz, and H. R. Kalbitzer, *J. Magn. Reson.* **89,** 543 (1990).

13. K. P. Neidig and H. R. Kalbitzer, *J. Magn. Reson.* **91,** 155 (1991).

14. M. Woodley and R. Freeman, *J. Magn. Reson. A* **118,** 39 (1996).

15. R. Baumann, A. Kumar, R. R. Ernst, and K. Wüthrich, *J. Magn. Reson.* **44,** 76 (1981).

16. R. Baumann, G. Wider, R. R. Ernst, and K. Wüthrich, *J. Magn. Reson.* **44,** 402 (1981).

17. C. Griesinger, O. W. Sørensen, and R. R. Ernst, *J. Chem. Phys.* **85,** 6837 (1986).

18. H. Oschkinat, A. Pastore, P. Pfändler, and G. Bodenhausen, *J. Magn. Reson.* **69,** 559 (1986).

19. A. Bax and R. Freeman, *J. Magn. Reson.* **44,** 542 (1981).

20. L. McIntyre and R. Freeman, *J. Magn. Reson.* **89,** 632 (1990).

21. M. Woodley and R. Freeman, *J. Magn. Reson. A* **109,** 103 (1994).

22. S. Simova, H. Sengstschmid, and R. Freeman, *J. Magn. Reson.* **124,** 104 (1997).

23. J. A. Högbom, *Astron. Astrophys. Suppl.* **15,** 417 (1974).

24. A. J. Shaka, J. Keeler, and R. Freeman, *J. Magn. Reson.* **56,** 294 (1984).

25. R. Freeman, ''Spin Choreography,'' Chap. 12, Spektrum Academic Publishing, Oxford (1997).

26. W. P. Aue, J. Karhan, and R. R. Ernst, *J. Chem. Phys.* **64,** 4226 (1976).