

Automated Extraction of Spin Coupling Topologies from 2D NMR Correlation Spectra for Protein ^1H Resonance Assignment

Jun Xu[†] and B. C. Sanctuary*

Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, PQ, Canada H3A 2K6

B. N. Gray

Research School of Chemistry, The Australian National University, GPO Box 4,
Canberra, ACT 2601, Australia

Received October 27, 1992

For assignment of proton resonances for a protein from two dimensional NMR, spin coupling topologies should first be identified. These spin coupling topologies can then be mapped to specific amino acids. As the size of a protein increases, the large number of the spin coupling topologies defies manual analysis due to the overlap of resonances. In this paper algorithms and programs which automatically extract protein spin coupling topologies from different two dimensional NMR experiments and deal with the overlap problems are reported. These techniques have been tested on a raw data set of protein melittin consisting of 26 amino residues.

INTRODUCTION

The structural determination of proteins and nucleic acids has been a long-term goal of biochemists. Until recently, biomolecular structure has been determined by X-ray diffraction techniques. These, however, are limited by two requirements. One is that the sample has to be a monocrystal. The other is that biochemical activity takes place in solution, and solution conformations may differ from those in the solid state. In many cases, however, it is impossible to obtain a biomolecular monocrystal of sufficient size and good quality.

In the 1970s, Wüthrich recognized that the distance information inherent in NMR NOE experiments was the key to studying proteins and nucleic acid conformations in solution. Subsequently, multiple dimensional NMR (MD-NMR) has been a powerful tool for determining biomolecular structure in solution.^{1,2} The advantages of NMR spectroscopy are as follows: (1) it can determine solution-state biomolecular structure; (2) it requires only a few milligrams of purified sample.

NMR is presently limited to determining the structures of relatively small biomolecules (i.e., molecular weight < 20 000). Larger biomolecules produce heavy overlap which is difficult to analyze. Future technological developments, giving higher spectrometer resolution (for example, magnetic fields of a GHz), new MD-NMR experiments (for example, three or four dimensional NMR experiments),^{3,4} and the appearance of new and more efficient software to analyze molecular connectivity all will help eliminate these limitations.

In the field of biomolecular structural determination from MD-NMR, many computer-based processes are involved which can be briefly summarized as follows: (1) base line correction; (2) multiple dimensional Fourier transformation; (3) multiple dimensional spectral data representation and processing;⁵ (4) multiple dimensional spectral peak-picking or automated peak recognition;^{6,7} (5) ^1H resonance specific assignments; (6) gathering distance constraints, amide proton exchange rates, short range NOEs, and three bond scalar coupling constants, i.e., J constants; and (7) creating a biomolecular model, and refining it.

In the above steps, 1-3, 6, and 7 are reasonably well developed. Steps 4 and 5, however, remain the bottlenecks.⁸

Before the three dimensional structure of a biomolecule via NOE (NOESY) can be determined, the sequence-specific assignment of the ^1H resonances must be done. With these, the cross-peaks in NOESY spectra can be unambiguously assigned after which they can be used to obtain distance constraints for structure calculations.^{8,9}

The manual sequence-specific assignment procedure involves a considerable amount of bookkeeping¹⁰ but also requires special skills and heuristic knowledge. Even so, these assignments are tedious and time consuming if done manually. Therefore, in this area, many attempts to automate these procedures have been made with several programs resulting.¹¹⁻¹⁹ These programs are predominantly good for bookkeeping but cannot duplicate the skills of human assignment.

One way to automate sequential assignment involves following basic logical steps involved in manual assignments:²⁰

(1) Predict all possible spin coupling topological fragments based upon the cross-peaks of NMR correlation spectra, for example, multiple quantum filtered COSY spectra.

(2) Identify spin coupling topological fragments by comparing the predicted data set against the experimental spectra.

(3) Assemble spin coupling topological fragments into independent spin coupling systems; in most cases these systems correspond to individual amino acid residues. This work can be aided by combining cross-peaks of TOCSY.

(4) Map spin systems to specific residues of the relevant protein with spin coupling topological pattern recognition and chemical shift rules.

(5) Determine the sequence-specific assignments by checking the dipole couplings between amide protons, α -protons, and β -protons (Proline has no amide proton) which come from different residues. NOESY spectra can provide the information to do this.

In this work, concepts from graph theory^{21,22} are used to study protein molecular connectivity. The correct spin coupling topological analysis is based upon the correct interpretation of cross-peaks in correlated spectra. A COSY cross-peak pattern may have many possible interpretations when viewed in solution. Each should be checked with data from another

[†] Present address: New Methods Research Inc./Tripos Associates, Inc., 6035 Corporate Dr., East Syracuse, NY 13057-1016.

* To whom the correspondence should be addressed.

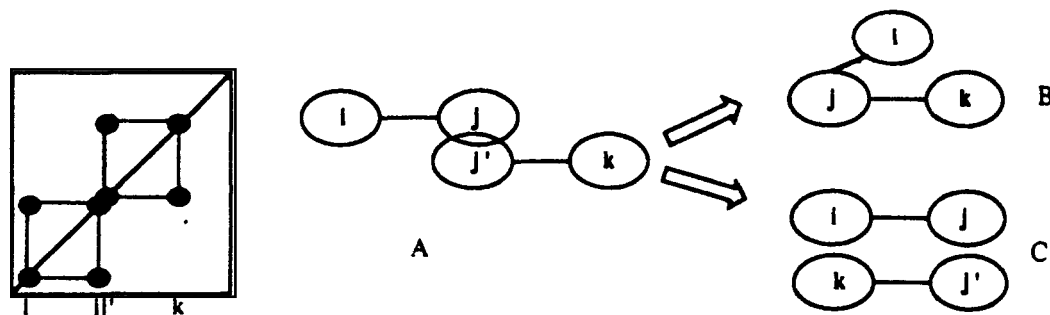


Figure 1. Overlap problem. Left is a part of 2QF-COSY spectra which implies two spin coupling edges represented by A, where $|\omega_j - \omega_{j'}| \leq \text{tolerance}$. It is difficult to tell if A is interpreted either as B or C. That is, spins i, j , and k may belong in the same spin coupling system, or spins i and j, j' , and k may belong in two different spin coupling systems which are overlapped at spin j .

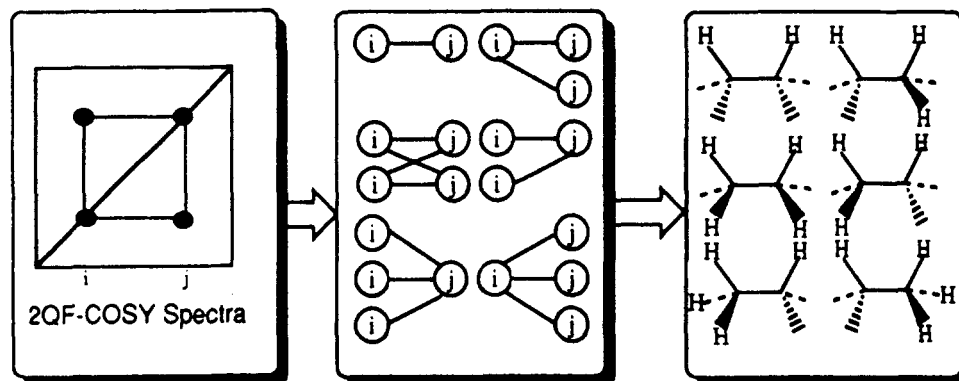


Figure 2. Magnetic equivalent problem. The overlap problem arising from degeneracies. Six chemical structural fragments produce six spin coupling topologies, but only two equivalent cross-peaks can be observed from 2QF-COSY spectra.

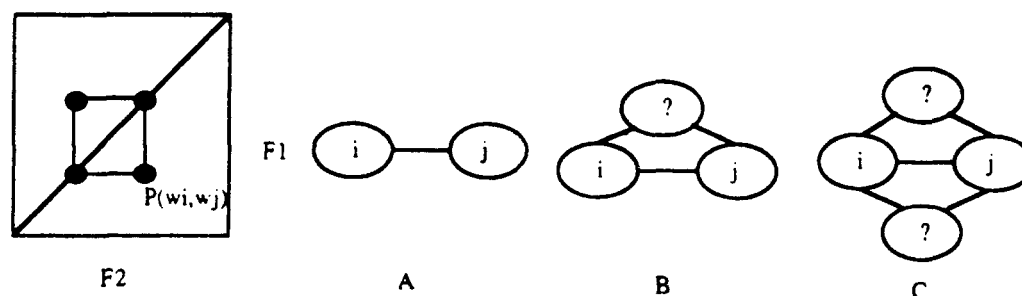


Figure 3. Spin topological fragments A, B, and C for p appearing in two, three, or four QF-COSY spectra, respectively.

data source. For example, TOCSY or MQ spectra must agree with a COSY pattern. Figure 1 shows an example.

Moreover, due to spectral degeneracies, magnetically equivalent spins give rise to the same peak in MQF-COSY, TOCSY, and NOESY spectra, from which it is not possible to distinguish if a peak represents $-\text{CHCH}_3$, $-\text{CHCH}_2-$, or $-\text{CHCH}<$, etc., unless the spectrum is well resolved and the cross-peak fine structure can be determined. Figure 2 shows some examples.

In this paper, a method which can extract and verify all possible spin coupling topologies from MQF-COSY and MQ spectra is proposed. With this method, spin coupling topologies including equivalent spins are identified directly from correlation spectra. These spin coupling topologies are important for a pattern recognition algorithm which maps spin coupling systems to specific amino residues of a protein in automated assignment of ^1H resonances studies.

DESCRIPTION OF THE METHOD

A MQF-COSY cross-peak, $p(\omega_i, \omega_j)$, can be considered as an edge of a spin coupling topological graph which represents a coupling between spin i and j . In a MQF-COSY spectrum of order M , a cross-peak between spin i and j signifies i and

j are mutually coupled to an order of at least $M - 2$ other spins (see Figure 3). These other spins however cannot be identified from a MQF-COSY spectrum alone. On the other hand, if the peak p appears in a MQF-COSY spectra, theoretically, it should also appear in the lower quantum order filter COSY spectra. For a variety of reasons, p may not be observed in other quantum order filter COSY spectra.

The amino acid residues in a protein due to the separation by carbonyls have the structure



Each amino acid in a protein will give rise to separate and independent coupling topology or topologies. Two like amino acids which have the same chemical and covalent structure do not necessarily have the same spin coupling topology. Some amino acids can give rise to more than one spin coupling topological system. Different amino acid types can give rise to the same spin coupling topology; e.g., Trp, Ser, Cys, Phe, Asp, Asn, Tyr, and His all contain a triangular three spin coupling topology. The positions of these resonances can overlap. For example, the protein melittin has three Lys residues; their spin topological networks and the chemical shifts are similar to each other. Figure 4 shows two residues

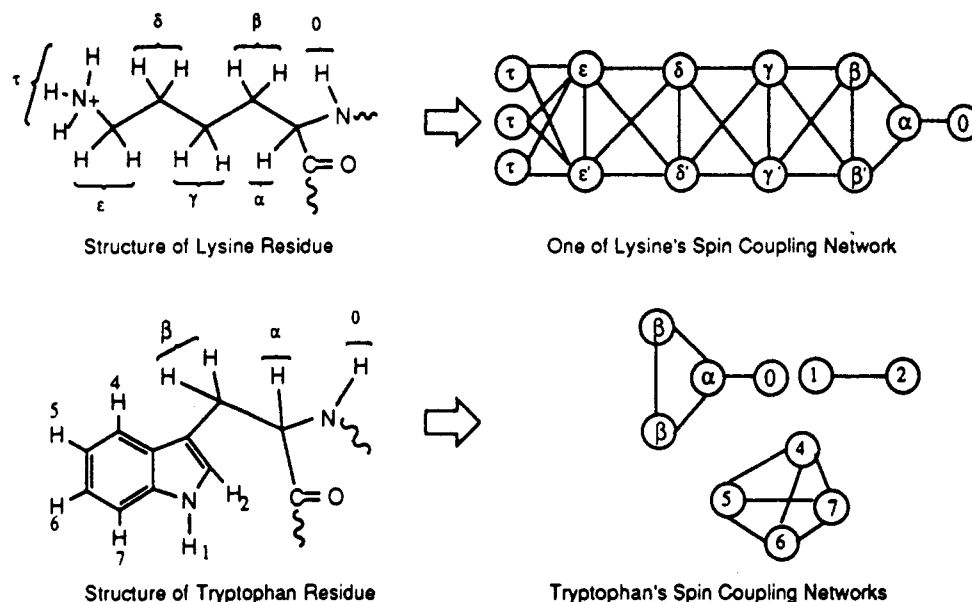


Figure 4. Two examples of amino acid residues and their spin coupling topological networks. In the most cases, three bond couplings are taken into account; Trp, however, has an aromatic ring, and therefore, four bond couplings between spins 4 and 6 and also spins 5 and 7 and also five bond couplings between spins 4 and 7 could be observed.

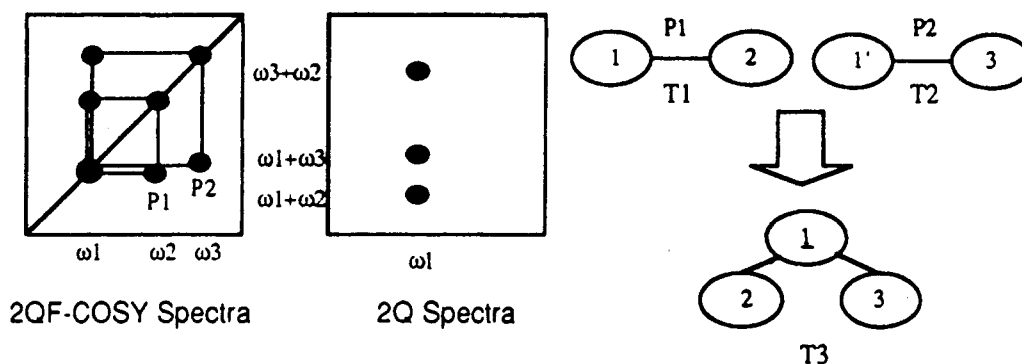


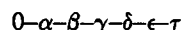
Figure 5. Building more complex spin coupling topologies. If peaks P_1 and P_2 belong in the same spin system, then the spin topological fragment T_3 is created. Peaks: $(\omega_1, \omega_3 + \omega_2)$, $(\omega_1, \omega_1 + \omega_2)$, and $(\omega_1, \omega_1 + \omega_3)$ in 2Q spectra must exist where $\omega_1 = (\omega_1 + \omega_1)/2$. It is possible that some of these 2Q cross-peaks cannot be observed due to overlap, incorrect threshold, or other experimental condition problems. However, if none of them are observed, P_1 and P_2 can only be considered as two partially overlapped peaks which may belong to different spin coupling systems but accidentally have a common frequency ω_1 .

and their theoretically possible spin coupling topological networks.

In Figure 4, the β -, γ -, δ -, and ϵ - protons of lysine can be magnetically equivalent or inequivalent due to the influence of the chemical environment. Therefore, the couplings β - β' , γ - γ' , δ - δ' , and ϵ - ϵ' may appear or not. Hence, the same kind of amino residue can produce different spin coupling topological networks when it appears in different proteins or in different positions in the same protein.

Some amino acid residues, such as tryptophan, histidine, tyrosine, and phenylalanine, have some protons separated by more than three bonds. Therefore this type of residue has more than one spin coupling system.

Although amino acid residues generate many different spin coupling topological patterns, for each type, there is a fixed spin coupling topological frame and chemical shift distribution.¹ Lysine, for example, no matter where it occurs, it should give rise to the following spin coupling topological frame



This frame comes from collapsing edges in Figure 4 which correspond to degeneracies. The collapsed graph however is too concise to properly distinguish different amino residues. Moreover, due to experimental data incompleteness, the above

frame may be split into fragments, thereby leading to incorrect assignments. To avoid such problems, the original spin coupling network should be recovered by using the information from other 2D NMR experiments, such as MQ NMR.



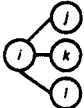
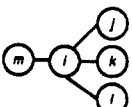

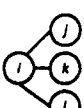
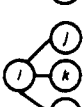
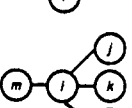
Each amino acid residue also has a distinguishing set of chemical shifts.^{1,23} Therefore, once a spin coupling topological network is created, it can be mapped onto an amino acid residue by a Fuzzy spin coupling topological Pattern Recognition Algorithm (FPRA) which can choose candidate spin coupling topological networks for residues of a protein. In turn, a tree search algorithm (TSA) will complete the sequence specific assignment by using NOESY cross-peaks.²⁶

Spin coupling networks can be extracted directly from COSY and MQ spectra. The usual strategy is to predict these on the basis of COSY cross-peaks. COSY cross-peaks can be considered as edges of a spin coupling topological graph. A spin coupling graph can be created by connecting these edges within a tolerance. If there are two COSY cross-peaks $P_1(\omega_1, \omega_2)$ and $P_2(\omega_1', \omega_3)$, and the following condition is met by them:

$$|\omega_1 - \omega_1'| \leq T$$

where T is a given tolerance, then within this tolerance a preliminary connection is made implying that spins 1, 2, and

Table I. All Possible Interpretations for 2-, 3-, and 4Q Cross-Peaks^a

quantum order	MQ peak ^b		spin coupling topology
	F_2	F_1	
2	$\omega_i + \omega_j$	ω_i	
2	$\omega_k + \omega_j$	ω_i	
2 ^c	$\omega_i + \omega_j + \omega_k - \omega_l$ or $\omega_i + \omega_j - \omega_k + \omega_l$ or $\omega_i - \omega_j + \omega_k + \omega_l$ or $-\omega_i + \omega_j + \omega_k + \omega_l$	ω_i	
2 ^c	$\omega_m + \omega_j + \omega_k - \omega_l$ or $\omega_m + \omega_j - \omega_k + \omega_l$ or $\omega_m - \omega_j + \omega_k + \omega_l$ or $-\omega_m + \omega_j + \omega_k + \omega_l$	ω_i	
3	$\omega_i + \omega_j + \omega_k$	ω_i	
3	$\omega_l + \omega_j + \omega_k$	ω_i	
4	$\omega_i + \omega_j + \omega_k + \omega_l$	ω_i	
4	$\omega_m + \omega_j + \omega_k + \omega_l$	ω_i	

^a This table is based upon weak J -coupling. ^b F_1 denotes resonances observed at single quantum domain, and F_2 denotes multiple quantum domain. ^c For example, a two quantum coherence during mixing time period may arise from four active spins. In this case, the observed frequency is the frequency combinations of the four spins. These are called combinational lines, although these lines are usually of lower intensity and may not be observed experimentally. If combination lines are not taken into account, incorrect interpretations may occur. Combination lines however, give information useful in identifying complicated spin coupling topology.

3 are part of the same graph. This hypothesis must be further verified with evidence from for example 2Q spectral peaks. If confirmed, it is reasonable to merge the two edges at ω_i and $\omega_{i'}$. In this way more complex spin coupling topologies are built. This procedure is shown in Figure 5.

A MQ cross-peak can have many possible interpretations. For 2-, 3-, and 4Q cross-peaks, these interpretations are listed in Table I for normal coherence transfer cases. In some situations, however, particularly in biomacromolecules, because of the asymmetrical chemical environment, relaxation violation lines can be observed (multiple dimensional NMR spectral selection rules have been discussed by Gray and Brown²⁷). For example, a spin i which couples to three other magnetically equivalent spins $j, j',$ and j'' can give rise to more peaks in MQ spectra; the relaxation times of magnetically equivalent spins $j, j',$ and j'' can be slightly different from each other (see Figure 6).

Normally, combinational lines may be observed in 2Q spectra. Higher multiple quantum experiments are less sensitive, and the combinational lines are too weak to observe. Therefore, in this paper only 2Q combination lines are considered. On the other hand, combinational lines do appear more often in 3D NMR experiments.³

To elucidate MQ spectra, we have developed the following prediction rules which are used in spin coupling topological network predictions and identifications:

(1) Normal prediction (NR), that is, prediction is based only upon the spin coupling topology made from merging several spin coupling edges (COSY cross-peaks). The merging order can be 2, 3, and 4. For example, (see Figure 7), a 3-order merging may mean that the merging can be carried out when there are three COSY cross-peaks which have a common frequency.

(2) Normal case and equivalent spin hypothesis (NREQ), i.e., assuming a COSY cross-peak is produced from degenerated lines, the hypothesis gives all possible spin coupling interpretations simply based upon given COSY cross-peaks. See also Figure 2.

(3) Normal case and combination lines (NRCL). Here, combination lines are predicted on the basis of the original spin coupling topology, and the equivalent spin hypothesis is not taken into account. The spin coupling topology on which the prediction is based should involve more than three spins.

(4) Normal case, equivalent spins and combination lines (NREQCL), i.e., combination lines are simultaneously predicted on the basis of the original spin coupling topology, and equivalent spin hypothesis is also taken into account at the same time. The spin coupling topology on which the prediction is based can involve two spins or more.

(5) Relaxation violation lines in combination with equivalent spin hypothesis (RVEQ). In this case, the relaxation violation lines are predicted on the basis of the spin coupling topologies, which include the equivalent spin hypothesis.

(6) Relaxation violation times in combination with equivalent spin and the combination lines hypothesis (RVEQCL). In this case, the relaxation violation combination lines are predicted on the basis of the spin coupling topologies which include the equivalent spin hypothesis.

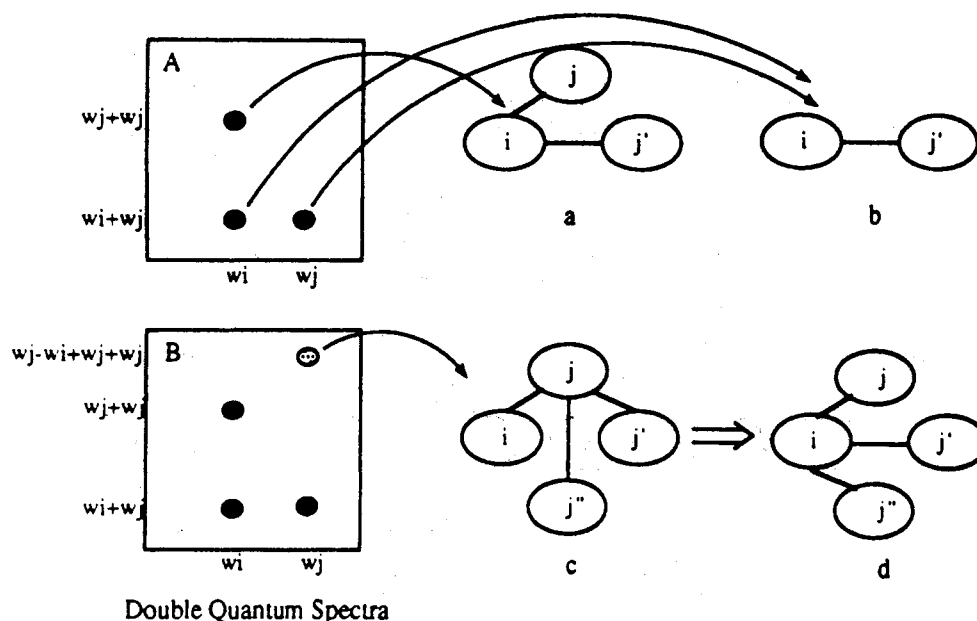
Suppose there are three COSY spectral peaks: $P_1(\omega_1, \omega_j)$, $P_2(\omega_2, \omega_k)$, and $P_3(\omega_3, \omega_l)$, where $\omega_1 < \omega_2 < \omega_3$, and $|\omega_3 - \omega_1| < \text{tolerance}$. This tolerance is chosen in accordance with the resolution of the spectrometer. Let $\omega_i = (\omega_1 + \omega_2 + \omega_3)/3$, and obtain a hypothetical spin coupling topological fragment H as shown in Figure 7 by merging spins 1, 2, and 3.

From fragment H in Figure 7, all possible double quantum spectral cross-peaks are predicted in Table II, and relevant spin topological fragments are listed in Figure 8.

Hypothetical spin coupling topologies, such as those shown in Table II, used to predict 2Q peaks, cannot always include those produced by a lower order. For example, three-edge spin coupling topology cannot always include two-edge spin coupling topology. This can be illustrated by the example in Figure 9.

The higher the merging order, the more complicated and numerous the hypothetical topologies become. As a result, the number of possible spin coupling topologies increases. On the other hand, when the merging order is more than four, due to poor resolution, the results may not be reliable. Hence, it is not necessary to do four-order or higher merging. The number of three-order mergers is more than two-order mergers. Four-order mergers, however, are less than three-order mergers due to fewer cross-peaks that satisfied the merging condition (see e.g. Table VII). On the other hand, more cross-peaks meet the conditions of two-order merging rather than three-order merging, but each three-order merging can produce more mergers than two-order merging due to combinational mathematical formula $C(2, n) < C(3, n)$, where n is the number of edges to be merged, and $n \geq 5$ (where $C(m, n)$ means the number of m elements chosen from n objects).

Clearly attempts to apply three hypotheses, prediction, and conformation are extremely tedious. To extract and identify

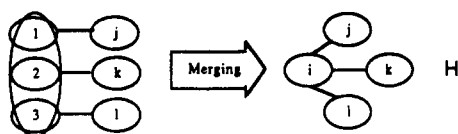


Double Quantum Spectra

Figure 6. Combination line from relaxation violation. In the normal case, spin topological fragments a and b may be deduced from double quantum spectra A. Due to an asymmetrical chemical environment, a combination line may be observed as in spectra B of this figure, which can be interpreted as spin topology c, and c implies spin topology d.

Table II. All Possible 2Q Spectral Peaks for Spin Coupling Fragment *H*

NR		NREQ		NRCL		NREQCL		RVEQ		RVEQCL	
no.	peak	no.	peak	no.	peak	no.	peak	no.	peak	no.	peak
1	(i, i+j)	7	(i, j+j)	10	(i, i+j+k-l)	14	(i, i+j+j-k)	35	(j, j+j)	38	(j, j+j+j-i)
2	(i, i+k)	8	(i, k+k)	11	(i, i+j-k+l)	15	(i, i+j+j-l)	36	(k, k+k)	39	(k, k+k+k-i)
3	(i, i+l)	9	(i, l+l)	12	(i, i-j+k+l)	16	(i, j+j+k-i)	37	(l, l+l)	40	(l, l+l+l-i)
4	(i, j+k)			13	(i, j+k+l-i)	17	(i, j+j+l-i)				
5	(i, j+l)					18	(i, i+k+k-j)				
6	(i, k+l)					19	(i, i+k+k-l)				
						20	(i, k+k+j-i)				
						21	(i, k+k+l-i)				
						22	(i, i+l+l-j)				
						23	(i, i+l+l-k)				
						24	(i, l+l+j-i)				
						25	(i, l+l+k-i)				
						26	(i, j+j+j-i)				
						27	(i, k+k+k-i)				
						28	(i, l+l+l-i)				
						29	(i, j+j+j-k)				
						30	(i, l+l+l-k)				
						31	(i, k+k+k-l)				
						32	(i, j+j+j-l)				
						33	(i, l+l+l-j)				
						34	(i, k+k+k-j)				

**Figure 7.** Initial, unconfirmed, spin coupling topology *H* on which MQ spectral prediction is based.

spin coupling topological fragments from the above predicted peak space, a computer based comparison procedure can be implemented to undertake this job. The comparison procedure compares MQ experimental spectral peaks (EP) against MQ predicted spectral peaks (PP). If an experimental peak matches a predicted peak, the following cases must be considered: (1) unique matching (that is, this experimental peak has a physical interpretation as a spin coupling topological fragment); (2) more than one predicted peak matches (that is, this experimental peak corresponds to several spin coupling topological fragments due to overlap); (3) no match (this experimental peak has no interpretation; for example, missing

MQF-COSY cross-peaks, or tolerance too narrow, etc.).

The experimental cross-peaks with multiple interpretations can be ranked with comparison deviation and verified with other data sources (for example, TOCSY spectra) to choose a correct interpretation. Sometimes, chemical properties of the investigated compound may be used, which are sometimes used to get rid of the impurity peaks before doing further analysis.

TOCSY spectra can be used to assemble the spin topological fragments into complete spin topologies, and by means of NOESY spectra, sequence-specific assignments of the protein are done.

ALGORITHMS AND PROGRAMS

For implementation of the above hypotheses, predictions, and comparisons, a number of computer algorithms are needed. These are discussed in this section.

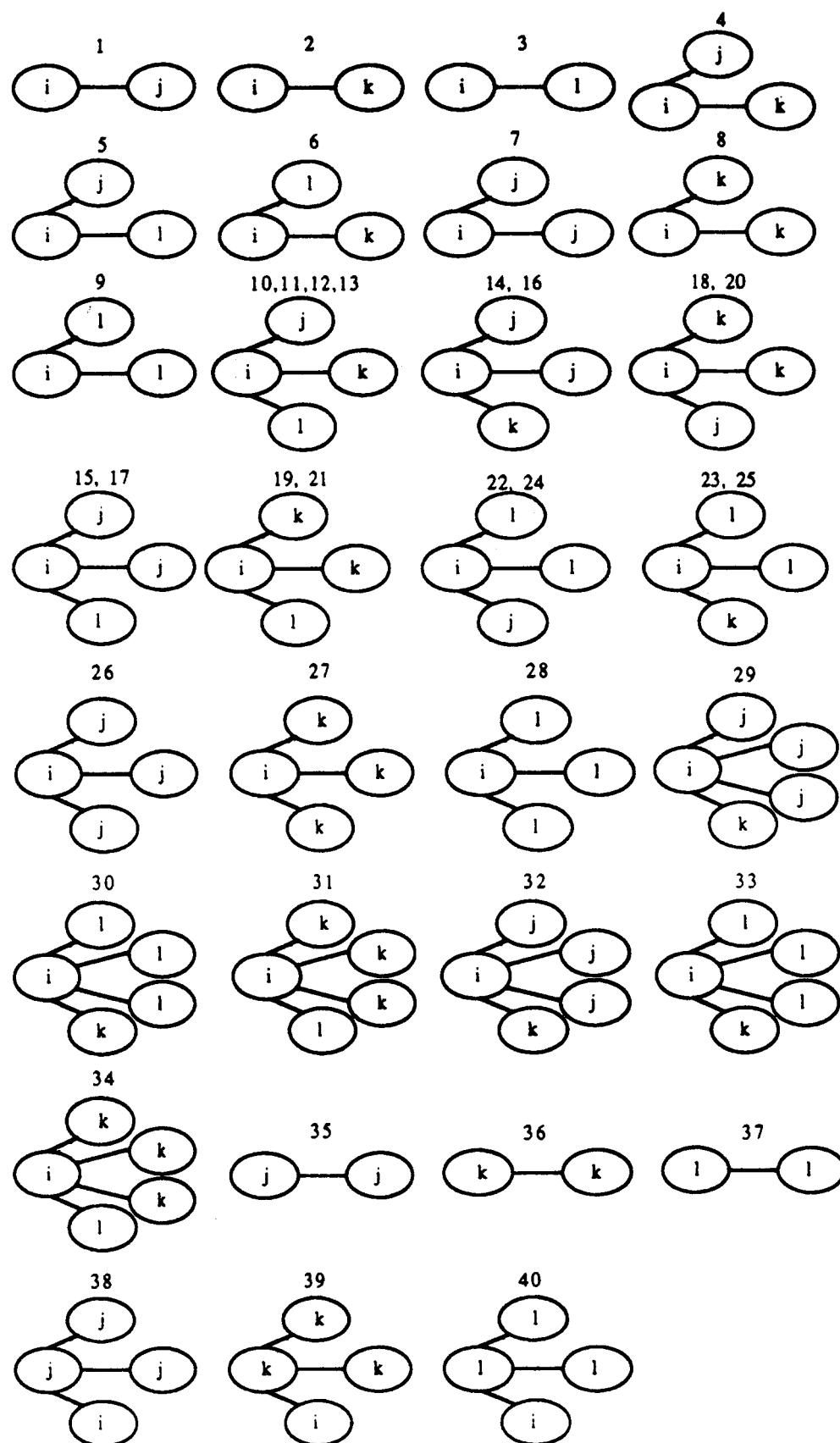


Figure 8. All possible spin topological fragments which can be identified from double quantum spectra. This prediction is based upon the hypothetical topology *H* in Figure 7.

Algorithm Merger. A multidimensional NMR spectral peak is defined as consisting of

$$\omega_1, \omega_2, \dots, \omega_m, \quad [\text{intensity, integral, etc.}]$$

where *m* is the number of dimensions, and entries in square

brackets are optional. For inclusion of all possible spin coupling topological edges from MQF-COSY, the algorithm receives more than one peak file as input. These files can be 2QF-COSY, 3QF-COSY, and 4QF-COSY spectra. In order to ensure all possible spin couplings are predicted, TOCSY cross-

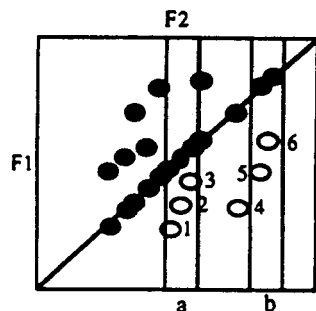


Figure 9. Three-order merging can be carried out in area "a" but not in area "b", while two-order merging can be carried out in both areas a and b ($|\omega_1 - \omega_3| < T$, $|\omega_4 - \omega_6| > T$, $|\omega_5 - \omega_6| < T$).

peaks are also combined to compensate for possible missing MQF-COSY cross-peaks. A peak observed in 4QF-COSY spectra may be present in 3QF-COSY and 2QF-COSY; these duplicates are filtered before the merging procedure is carried out. For example, if 2QF-COSY cross-peak $P_D(\omega_1, \omega_2)$, 3QF-COSY cross-peak $P_T(\omega_1, \omega_2)$, and 4QF-COSY cross-peak $P_Q(\omega_1, \omega_2)$ all exist, only peak $P(\omega_1, \omega_2)$ is retained, where $\omega_1 = (\omega_1 + \omega_1' + \omega_1'')/3$ and $\omega_2 = (\omega_2 + \omega_2' + \omega_2'')/3$. In addition, the cross-peaks (ω_1, ω_2) and (ω_2, ω_1) are considered as equivalent. So either one, but not both, is kept. After all the cross-peaks are entered, the peak set is transformed into a topological

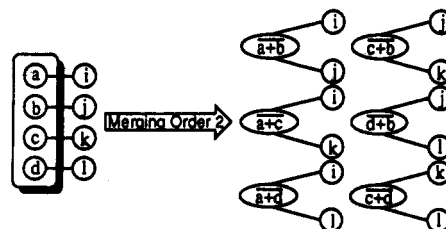


Figure 10. Example of four peaks within a tolerance. The possible two-order mergers are C(2, 4), that is six two-branched trees.

adjacency table called A_1 ; the subscript "1" indicates it contains one-branched trees. The data structure for A_1 is defined as follows:

```
struct spin-topology {
    chemical-shift: real;
    adjacency-table: array [0..N] of integer;
};
```

A_1 : array [0..M] of spin-topology;
the entry "chemical-shift" is sorted, such that the i th chemical shift is always less than or equal to $(i-1)$ th. The algorithm scans A_1 from the i th chemical shift to find all other chemical shifts which have approximately the same value of i th chemical shift within a given tolerance. The scope of scanning is in the range from the $(i+1)$ th chemical shift to the last one. The i th scanning is taken as an example; supposing there are n

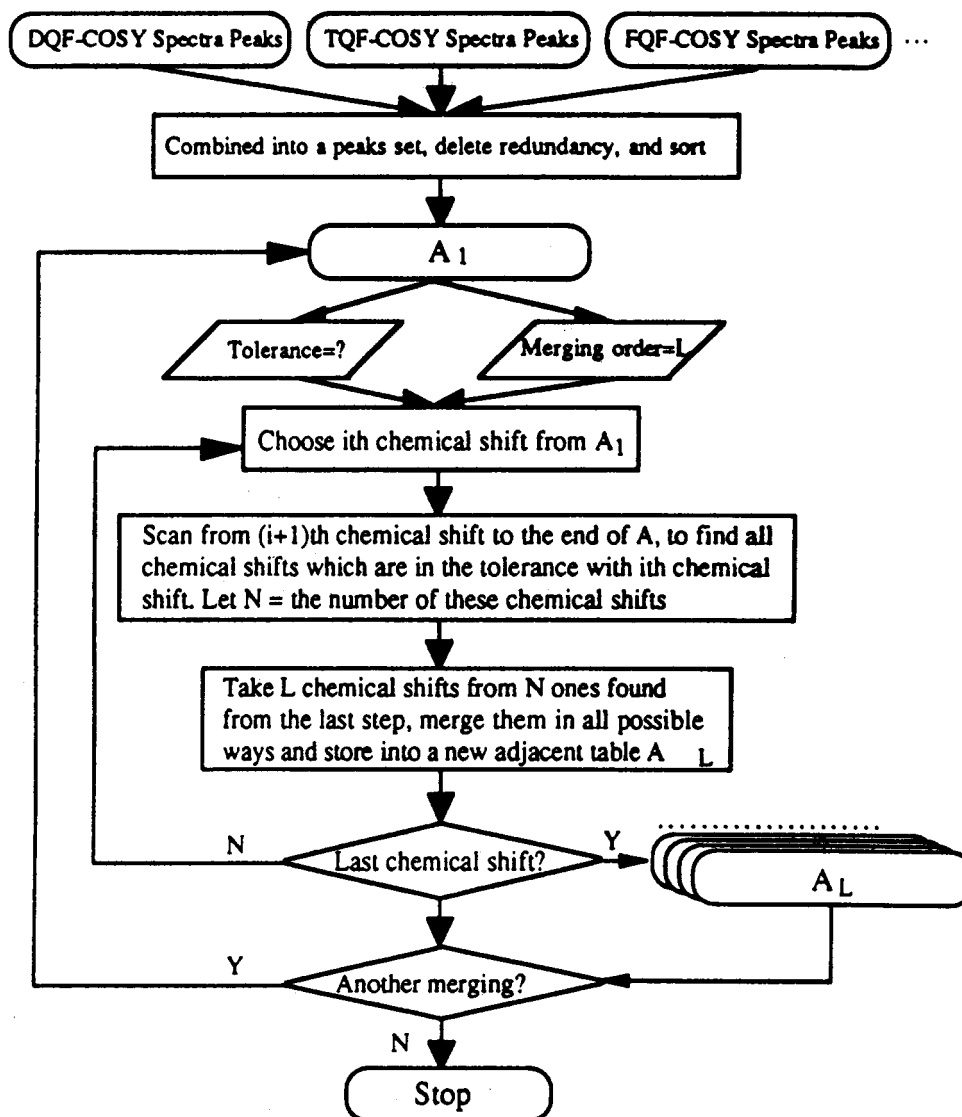


Figure 11. Flow chart of algorithm merger.

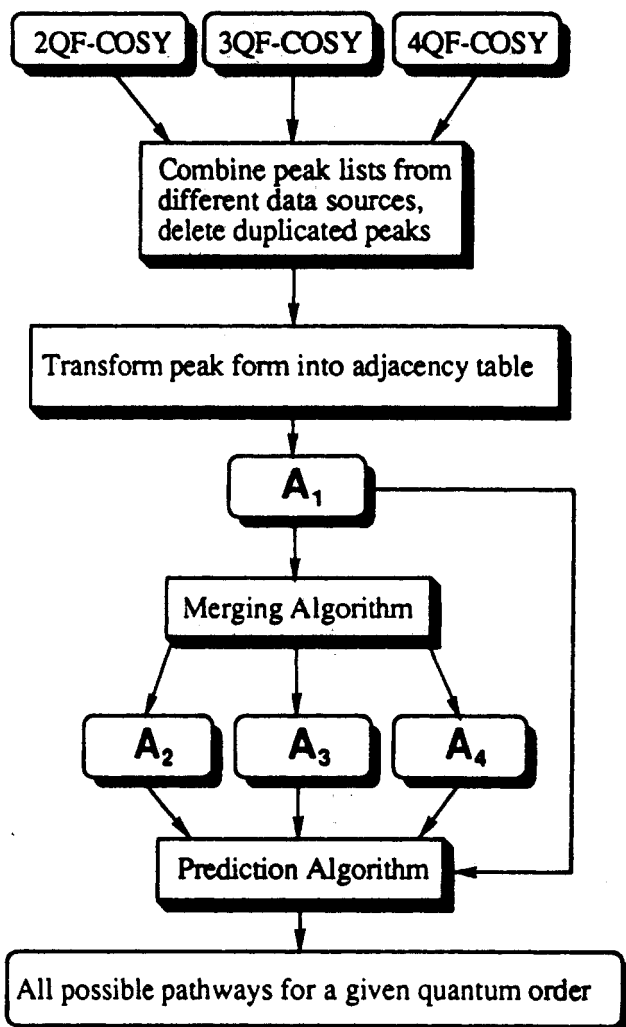


Figure 12. Data flow chart of merging and prediction algorithms.

chemical shifts found to be within the tolerance, there will be $C(2,n)$ two-order mergers, $C(3,n)$ three-order mergers, and $C(4,n)$ four-order mergers. n is greater than or equal to the merging order. Figure 10 gives an example of a merging order 2 involving four edges.

If L is the merging order and N is the number of peaks within the given tolerance, then the number of mergers is $C(L,N)$, i.e., choosing a subset of L elements from a set of N objects. Figure 11 is the flow chart of algorithm "merger".

Algorithm MQ-PREDICTION. The main task of algorithm MQ-PREDICTION is to predict all possible coherence pathways which can be observed from MQ experiments and to calculate the MQ spectra. The input of the algorithm is A_L , where L is the merging order. A_L contains a group of

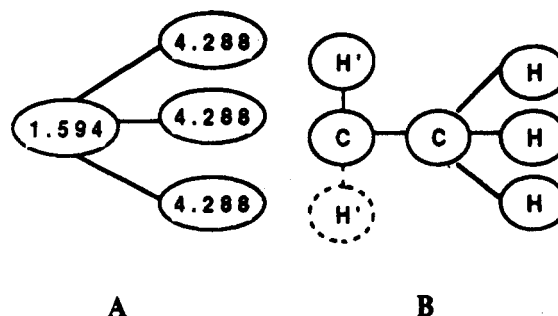


Figure 13. Possible pathway and corresponding covalent structure.

L -branched trees. The number of the possible coherence transfer pathways for MQ spectra are calculated from the following formulas:

normal pathways

$$\text{NRP}(Q,L) = N(C_L^{Q-1} + C_L^Q) \quad (1)$$

normal and equivalent spin pathways

$$\text{NREQP}(Q,L) = N \sum_{i=1}^{i=M} (C_L^{Q-1-i} + C_L^{Q-i}) \quad M \leq 3 \quad (2)$$

normal and combination line pathways

$$\text{NRCLP}(Q,L) = N(C_L^{Q+1} + C_L^{Q+2}) \quad (3)$$

normal, equivalent spins, and combination line pathways

$$\text{NREQP}(Q,L) = N \sum_{i=1}^{i=M} (C_L^{Q+1-i} + C_L^{Q+2-i}) \quad M \leq 3 \quad (4)$$

relaxation rules violation and equivalent spin pathways

$$\text{RVEQP}(L) = NC_L^1 \quad (5)$$

relaxation rules violation, equivalent spins,
and combination line pathways

$$\text{RVEQCLP}(L) = NC_L^1 \quad (6)$$

Formulas 1–6 are classified into two types: combination line type (formulas 3, 4, and 6) and noncombination line type (formulas 1, 2, and 5). The method of predicting the former is slightly different from the latter. When $M = 0$, formula 2 becomes formula 1. Therefore, a general algorithm to predict all six cases can be briefly described as follows.

Algorithm MQ-Prediction(adjacency-table A_m ,
int Merger-order, int Quantum-order,
Transmitter-offset)

Table III. Format of the Predicted Data Set for MQ Spectra

labels and frequencies form						frequencies form			
L_1	L_{2s}	L_3	F_1	F_2	F_3	F_1	F_{2s}	F_3	F_2
i	i, j, k, \dots	i	f_i	$f_{i+j+\dots}$	f_i	f_i	f_i, f_j, f_k, \dots	f_i	$f_{i+j+\dots}$
i	j, k, l, \dots	i	f_i	$f_{j+k+\dots}$	f_i	f_i	f_j, f_k, f_l, \dots	f_i	$f_{j+k+\dots}$
...

Table IV. Data Structure of Comparison Object

experimental peak		predicted peak and pathway							
ω_2	ω_1	best possible matched peak		status	D_{mq}	D_{sq}	best possible matched pathway	candidate pathways	
float	float	ω_2	ω_1						
float	float	float	float	integer	float	float	PATH	PATH array	

Table V. Data Structure of Path Object

status	D_{mq}	D_{sq}	spin topology
integer	float	float	float array

float

/* Here A_m is the merger set from m-order merging procedure, which hold n depth=1 and m-branched trees */

```

{
for (i = 1, i ≤ n, i++) /* n is the length of  $A_m$  */
{
take node i and its adjacent nodes, put them into table
ith-branched-tree; for (j = 0, j ≤ Quantum-order,
j++)
{
if ((j < 3) && ((Quantum-order-j) < 3))
{
enumerate all possible pathways which involve
equivalent spins from ith-branched-tree;
/* Case j = 0; Normal pathways;
Case j = 1: Pathways involving double equivalent
spins;
Case j = 2: Pathways involving triple equivalent
spins;
Case j ≥ 3 is rare in organic molecules */
};
};
};

```

calculate MQ spectra from above pathways;

/* The method is that the frequency pointed by the first component of the pathway is to be ω_2 , and $\omega_1 = \sum_k (\text{Transmitter_offset} - \omega_k)$,

$k \leq \text{Quantum_order}$, k is the kth component of the pathway. When the first component of the pathway is not i, this pathway will be the relaxation violation pathway. */

};

} //end of the algorithm

To predict combination lines, the above algorithm must be slightly modified. That is, for 2Q spectra, the input parameter *Quantum-order* is replaced by *Quantum-order* + 2, and the rule for calculating the multiple quantum frequency domain is changed. For example, to produce a 2Q combination line, a two quantum coherence transfer pathway during t_2 should involve four active spins, e.g., i, j, k, and l; therefore, all possible frequency combinations in multiple quantum frequency domain are $\omega_i + \omega_j + \omega_k - \omega_l$, $\omega_i + \omega_j - \omega_k + \omega_l$, $\omega_i - \omega_j + \omega_k + \omega_l$, and $-\omega_i + \omega_j + \omega_k + \omega_l$. When, however, some of these frequencies are equivalent, the possible combination lines are sharply reduced due to degeneracies. For example, if j, k, and l are equivalent spins, then only one frequency can be observed, i.e., $-\omega_i + \omega_j + \omega_k + \omega_l$. This case corresponds, for example, to a proton coupling with the protons of a methyl group.

The spectral data processing of merging and prediction algorithm, are shown in Figure 12.

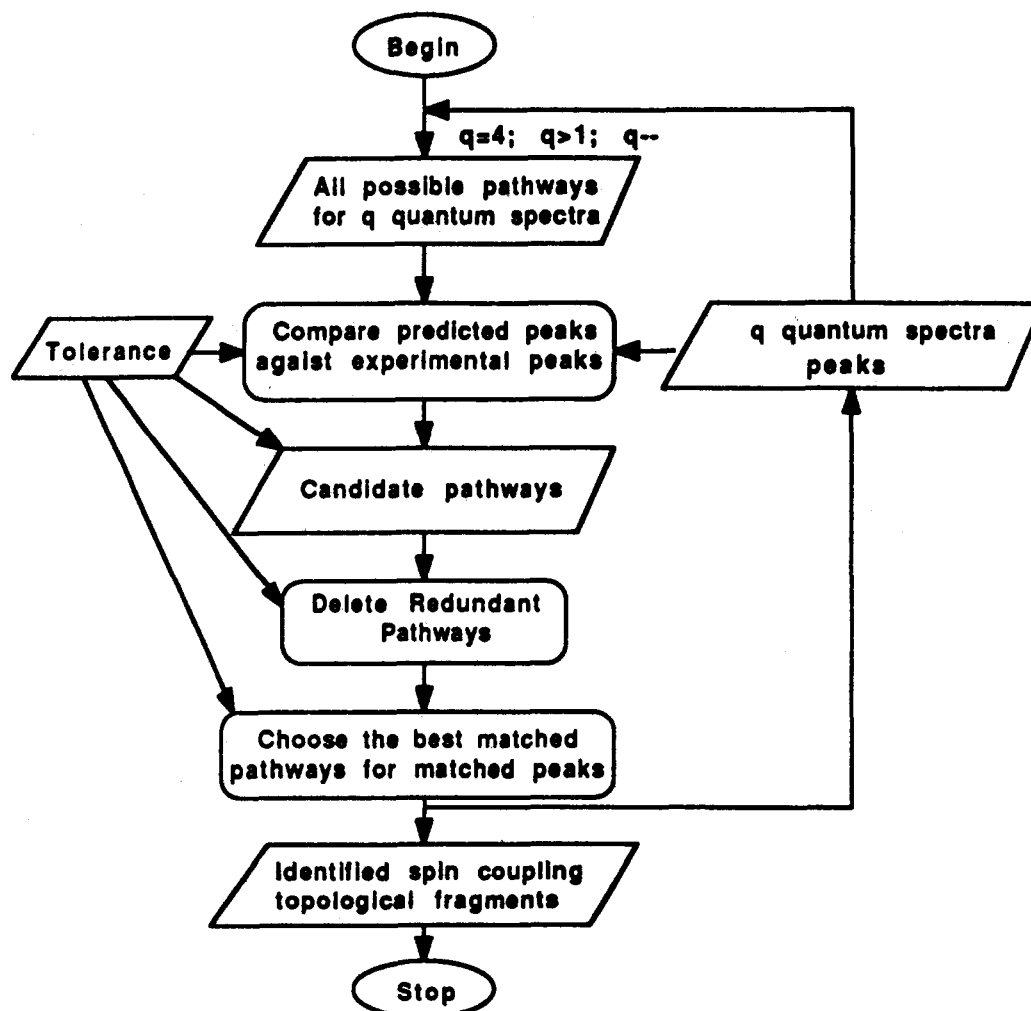


Figure 14. Flow chart of the comparison.

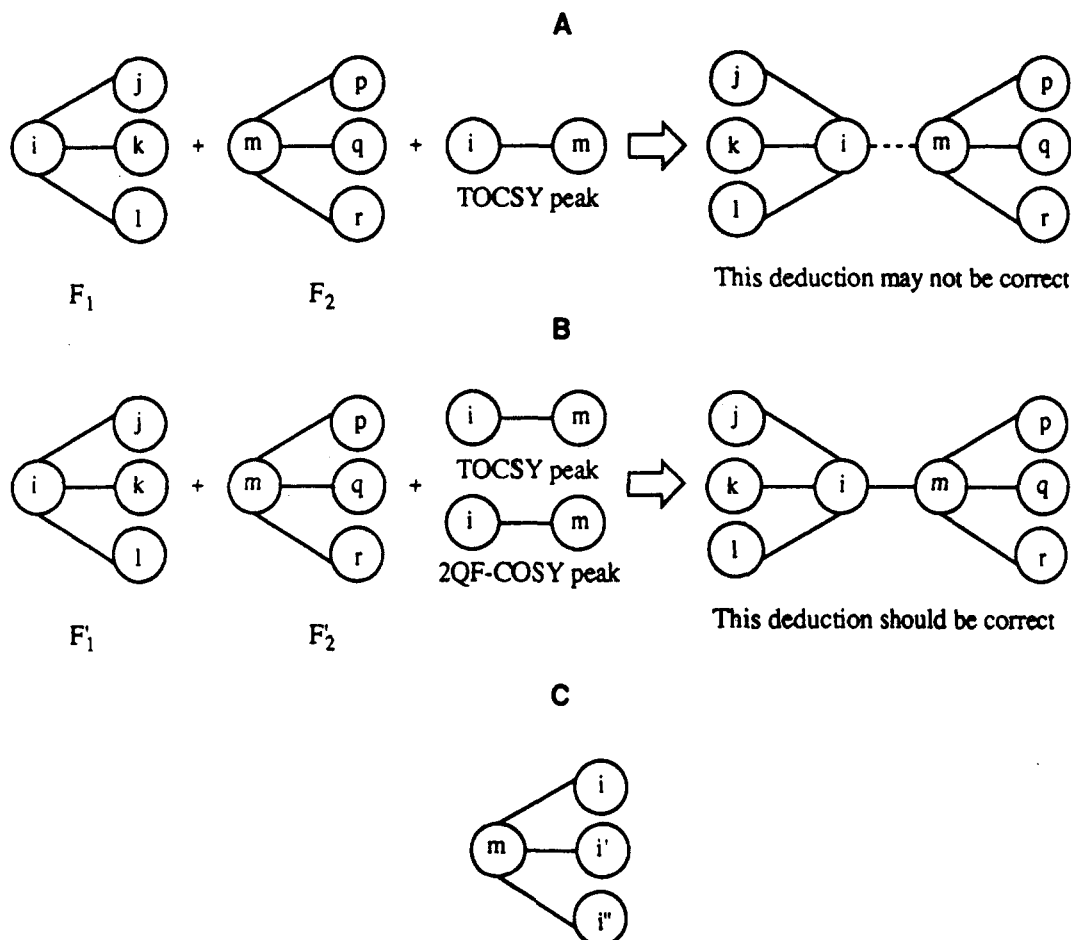


Figure 15. Assembling spin coupling topological fragments with TOCSY spectra.

Algorithm Comparison. Algorithm COMPARISON verifies that the spin coupling topological fragments from the predicted data set and screens them by comparing predicted MQ spectral peaks against experimental MQ spectra.

The data set from algorithm MQ-prediction is stored in the format given by Table III. Where F_1 , F_2 , and F_3 are active spins which correspond to the three experimental time periods in the general description for homonuclear NMR experiments, i.e., excitation period t_1 , mixing period t_2 , and detection period t_3 .²⁷

Each frequency found from COSY spectra is labeled by an integer, and the float number chemical shift value is also retained. It is in the right columns of Table III. Column F_1 contains spins excited in the experimental period t_1 ,²² and column F_2 contains spins which involve coherence transfer in the experimental period t_2 . Column F_3 contains spins which are expected to be observed at the single quantum frequency domain from the MQ spectra, and column F_2 contains frequencies observed in various multiple quantum frequency domains. Columns F_2 and F_3 actually consist of predicted peaks for multiple quantum experimental spectra, which will be compared with the experimental spectral peaks. Column F_2 contains coherence pathways, which arise from spin scalar coupling topologies. For example, if a predicted peak $P(f_i f_{j+k})$ matches a double quantum experimental spectral peak, within a given tolerance, then a coherence-transfer pathway $f_j \leftarrow f_i \rightarrow f_k$ is identified; i.e., the spin coupling topological fragment $f_j - f_i - f_k$ is identified.

The algorithm COMPARISON performs all comparisons of predicted peaks and experimental peaks but also transforms identified coherence-transfer pathways into adjacency tables.²⁴

The spectral resolution of the multiple quantum frequency domain and the single quantum one can be different due to the different digital resolution in each dimension. Therefore, different tolerances are applied in different frequency domains in the algorithm comparison.

The result of the algorithm comparison may give redundant spin coupling topologies due to following situations:

(1) One spin coupling topology may give rise to more than one coherence-transfer pathway which in turn may give rise to different peaks. For example, spin coupling $i-j$ may produce two double quantum spectral peaks: $(f_{i+j}f_i)$ and $(f_{i+j}f_j)$.

(2) Given a spin coupling topology, many coherence-transfer pathways can be predicted. Each pathway can be observed in different quantum order spectra. For example, coherence-transfer pathways involving three spins can be observed in double and triple quantum spectra, while coherence-transfer pathways involving four spins can be observed in triple and four quantum spectra, etc.

(3) Normal coherence-transfer pathway and relaxation violation coherence-transfer pathway may be interpreted as the same spin coupling topology.

(4) COSY cross-peaks may be reconfirmed from multiple quantum experimental spectra.

(5) Some spin coupling topologies may appear simultaneously due to the fact that some initial edges in A_1 can be merged in merging order 2-4; therefore, more predictions based upon homomorphic spin coupling topologies are possible.

Before assembling these spin topological fragments, these redundant spin coupling topologies are deleted.

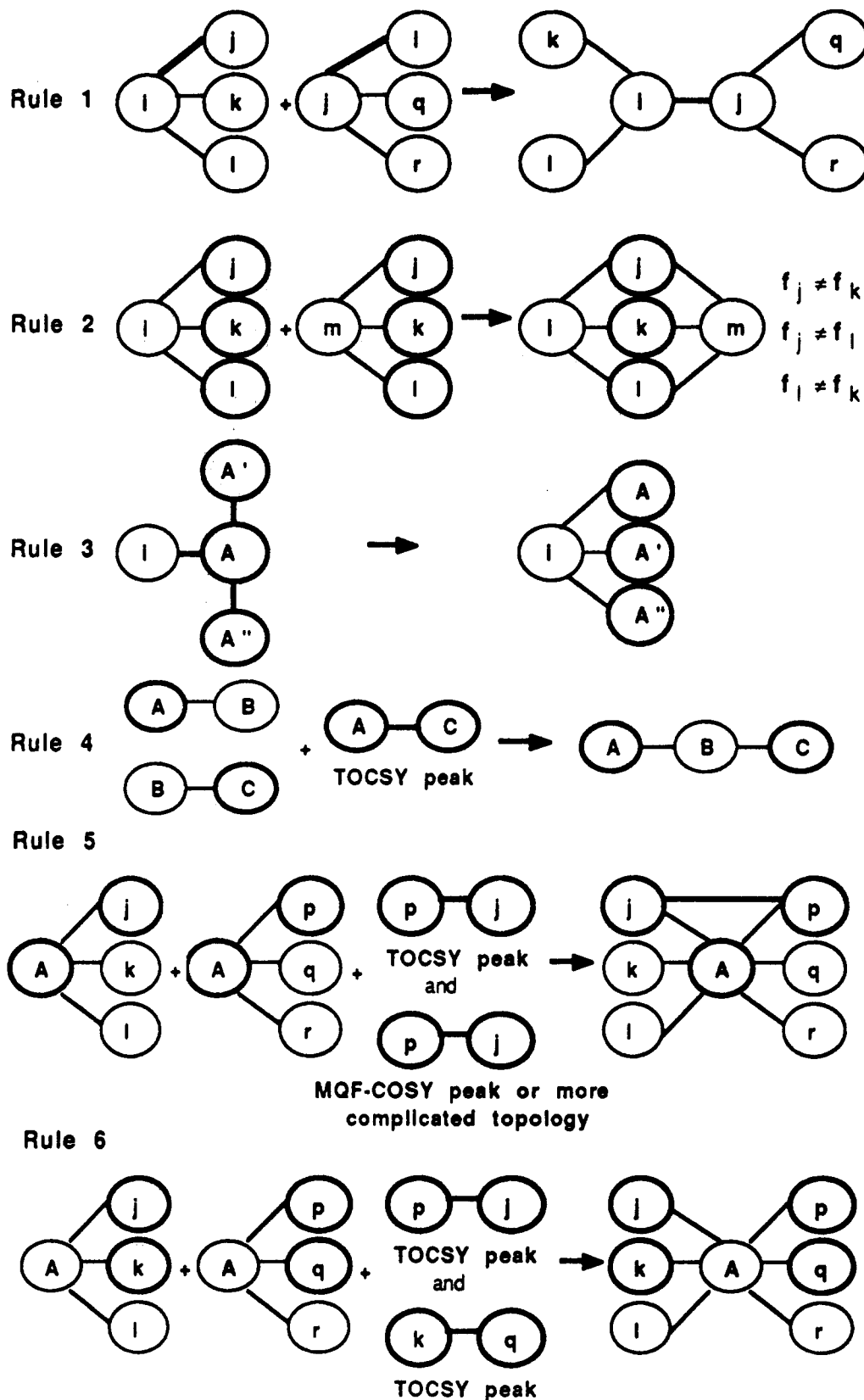


Figure 16. Graphic explanation of assembling rules.

After comparison, some multiple quantum experimental peaks have unique spin coupling topological interpretations. Some multiple quantum experimental peaks, however, have more than one interpretation due to overlap of chemical shifts. To select the correct spin coupling topologies from these, they are ranked by a criterion which minimizes a match distance Δ , which is defined as follows:

$$D_{mq} = |\omega_{\text{prediction}} - \omega_{\text{experiment}}|_{\text{multiple quantum domain}}$$

$$D_{sq} = |\omega_{\text{prediction}} - \omega_{\text{experiment}}|_{\text{single quantum domain}}$$

$$\Delta = D_{mq} + D_{sq}$$

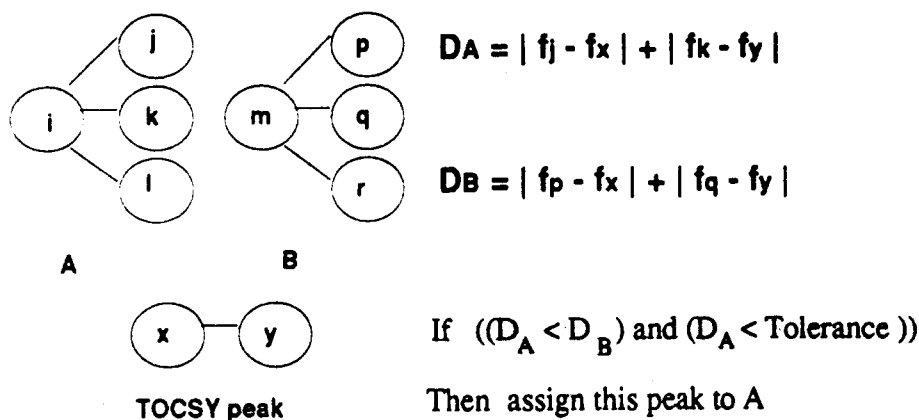


Figure 17. Assignment of the nearest fragments to a TOCSY peak.

Table VI. Data Structure for Assembling Algorithm^a

TOCSY peak		assignment message									next peak
ω_1	ω_2	BF ₁	BF ₂	status	chosen	D ₁	D ₂	D _c	CL ₁	CL ₂	assembling
float	float	integer	integer	integer	integer	float	float	float	integer array	integer array	

^a BF, best matched fragment. Status, same spin topology assignment, set 1, otherwise set 2; chosen, when the peak is assigned, set 1; D₁, difference of ω_1 and the corresponding frequency in BF₁; D₂ difference of ω_2 and the corresponding frequency in BF₂; D_c, difference of frequency in BF₁ and the corresponding frequency in BF₂, which are most closed; CL, candidate list.

Table VII. Merging Results of Melittin MQF-COSY and TOCSY Spectra

merging order	merging tolerance	length of A _M
1		552 ^a
2	0.03	2631
3	0.03	8646
4	0.03	6514

^a Comes from combined MQF-COSY and TOCSY cross-peaks directly.

Table VIII. Numbers of Predicted MQ Spectral Peaks

adjacent table	quantum order	no. of peaks
A ₁	2	3 840 ^a
	3	3 846
	4	3 294
A ₂	2	31 504
	3	52 520
	4	60 438
A ₃	2	189 877
	3	354 128
	4	475 711
A ₄	2	227 460
	3	461 973
	4	697 441

^a Including combinational lines for methyl group.

In addition, chemically acceptable criteria may also be considered. For example, in the protein melittin 2Q spectra, there is a peak P_{2q} (-3.394, +4.301), which, according to the chemical shifts, should be assigned as a coupling of $\alpha - \beta$ (4.32, 4.28), this belongs in thr-10 of protein melittin. This peak (4.32, 4.28), which could conform to this coupling, is, however, not observed in MQF-COSY spectra. Therefore, the prediction algorithm cannot produce the pathway $\alpha \alpha \beta$ or $\beta \beta \alpha$ from this 2Q spectra. After comparison, a unique match for P_{2q} implies the spin coupling topology like Figure 13A is due to an overlap.

According to Figure 13A, a spin, which could be two equivalent proton spins, could be coupled to three other protons. According to a chemical shift rule, these three protons should

be a methyl group as shown in Figure 13B. Considering it is a protein sample, however, and the chemical shift for a methyl group proton is not consistent with the value 4.228 ppm, the methyl group interpretation is not acceptable. The assignment should therefore be made on the basis of further evidence.

To identify the best possible interpretation for every experimental peak and subsequently transform coherence-transfer pathways into a spin coupling topological adjacent table, complex data structures are created and shown in Tables IV and V.

In Table V, "status" identifies the origin of the pathway. The best possible matched pathways are transformed into spin coupling topologies and accepted as the input for further assignments. Candidate pathways are kept for examination, in the event that the best possible match is not accepted by estimation criteria.

The validity of results from the algorithm COMPARISON results also depends on a suitable tolerance. When the range of the tolerance is too narrow, many peaks cannot have assignments. If the range is too broad, however, many experimental peaks may have multiple interpretations. This can change previously found matched pathway. In order to speed up the comparison procedure, the above data structures are implemented with dynamic link lists in the main memory which can be quickly reset and processed. The flow chart of the algorithm comparison is shown in Figure 14.

Algorithm Assembler. The algorithm assembler undertakes the job of finding different spin coupling topological fragments which belong in the same spin system and connects them by using cross-peaks from TOCSY. Theoretically, if there is a TOCSY cross-peak P_{TOCSY} (ω_i, ω_j), and spin ω_i belongs to spin coupling fragment F₁, while spin ω_j belongs in spin coupling fragment F₂ (see Figure 15), then the conclusion that F₁ and F₂ belong to the same spin system is possible. Practically, however, this deduction may not be correct, due to possible heavy overlap. To confirm this deduction, more evidence is needed.

Table IX. Comparison of Computer Assignment and Manual Assignments for Melittin

experiment	total peaks	computer assignment		manual assignment		identity/%
		assigned peaks	unassigned peaks	assigned peaks	unassigned peaks	
2Q	165	138	27	121	45	82.6
3Q	75	72	3	57	18	86.3
4Q	15	15	0	14	1	85.7

Table X. Assignment for Melittin MQ Spectra

spectra	peaks assigned by unique match and agree with manual assignment	peaks assigned by best match and agree with manual assignment	multiple matched peaks
2Q	53	43	81
3Q	12	21	56
4Q	2	10	13

In Figure 15A, due to overlap between spins i and m , a TOCSY peak can arise from a fragment such as shown in Figure 15C. In fact, F_1 and F_2 may belong to different spin systems. Hence, deducing rules to merge fragments are formulated as follows.²⁵

Rule 1: Two topologies which contain a common edge may be merged at the common edge to give a single topology.

Rule 2: Two topologies which contain three common distinct vertices can be merged at these three vertices to give a single topology.

Rule 3: If evidence indicates that vertex A represents degenerate hydrogens A, A', and A'', then vertices which are connected to A are said to be also connected to vertices A' and A''.

Rule 4: The two topologies A-B and B-C, can be merged at vertex B if there exists in a TOCSY spectrum, a peak correlating the frequencies represented by vertices A and C, and if there does not exist a similar peak in multiple quantum filtered COSY spectra.

Rule 5: Two topologies with a common vertex, A, can be merged at A, if there exists TOCSY peaks from a vertex in a third topology, to at least one vertex in each of two topologies, which are with a radius of two edges from A.

Rule 6: Two topologies with a common vertex, A, can be merged at A, if there exists at least two TOCSY peaks connecting vertices between the two topologies (including peaks connecting to A).

These rules are illustrated by graphic examples of Figure 16.

Frequencies of a TOCSY cross-peak may come from the same spin coupling topological fragment or two different spin coupling topological fragments which belong to the same spin system. Therefore, the assignment of TOCSY cross-peaks is divided into two steps: (1) assign peaks from the same spin coupling topological fragment; (2) assign peaks from the different spin coupling topological fragments.

Also, a TOCSY cross-peak may have multiple interpretations; only the best matched fragments are chosen to be the assignment. Figure 17 shows a simple example to deal with this case.

It is simpler to assign a TOCSY cross-peak in which frequencies belong to the same spin coupling topological fragment. In this case, to assign a TOCSY cross-peak (ω_1 , ω_2), all spin coupling topological fragments which contain frequencies ω_1 and ω_2 within a given tolerance are checked as candidates, and the nearest candidate is assigned. It is complicated, however, to assign a TOCSY cross-peak in which

frequencies belong to different spin systems. So the above-mentioned assembling rules should be applied. This procedure can be divided into the following steps.

If there is a TOCSY cross-peak (ω_1, ω_2) which is not assigned to a spin coupling topological fragment, then (1) find all spin coupling topological fragments which contain ω_1 within a given tolerance, and put them into a candidate list CL₁; (2) find all spin coupling topological fragments which contain ω_2 within a given tolerance, and put them into a candidate list CL₂; (3) search fragment pair P₁ and P₂, where P₁ is in CL₁, and P₂ is in CL₂. If P₁ and P₂ agree with the assembling rules of Figure 17, merge P₁ and P₂ into spin system P. Hence the TOCSY cross-peak (ω_1, ω_2) is assigned.

A TOCSY cross-peak can be assigned to more than one spin coupling topological fragment pair, and the best matched pair is chosen as the assignment. The data structure for algorithm, assembler is listed in Table VI.

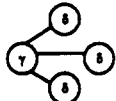
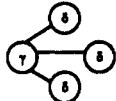
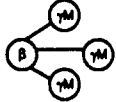
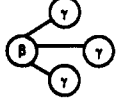
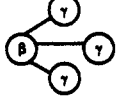
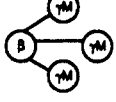
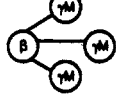
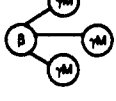

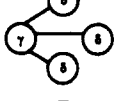
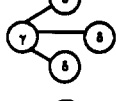
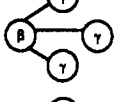
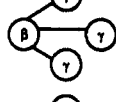
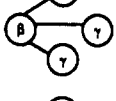
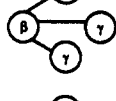
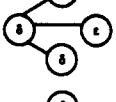
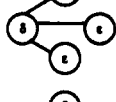
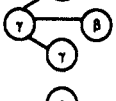
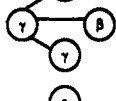
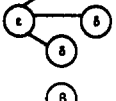
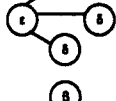
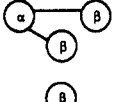
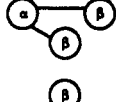
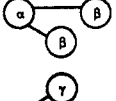
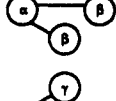
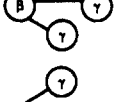
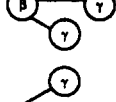
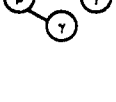
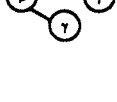
Program Architecture. The programs are implemented independently and can be called and run by a main program. The main program reads two-dimensional NMR spectral peaks, transforms them into a spin coupling topological table A₁. A₁ can be merged with order 2, 3, or 4, and A₂, A₃, and A₄ are created for prediction. Program PREDICTOR predicts all possible MQ spectra, where M is quantum order of 2, 3, and 4. When quantum order > 4, the spectral intensities are usually too low to be detected. All predictions are based upon A₂, A₃, and A₄. Normal lines, combinational lines, and relaxation violation lines are all predicted. A₁, A₂, A₃, and A₄ are stored as an adjacency table, while predicted data are stored in the form of coherence-transfer pathways and multiple quantum peaks (see Table III). Results from COMPARISON have two formats: one is the user's checking, and the other is for the program ASSEMBLER's processing. The latter result identifies spin coupling topologies in the form of adjacency tables. All output data are stored as a text file, which is easily checked by a user.

This software package consists of more than 25 000 C++ statements and runs on a SUN Sparc Station supported by the OpenWindows system.

RESULTS AND DISCUSSION

The complete melittin MQF-COSY, MQ, TOCSY, and NOESY experimental spectra have been obtained from Varian 500-MHz NMR spectrometer. All experiments were run at 40 °C, on a micelle bound melittin sample which was p¹H adjusted to 3.0. A melittin concentration of 5 mM was used, with a ratio of detergent (DDPC) to melittin molecules of 45:1. Presaturation was used to suppress the H₂O signal in the experiments conducted in H₂O.²⁵

Table XI. Comparison of 4Q Peak Assignments of Melittin

peak		computer assignment		manual assignment	
ω_2	ω_1	residue	spin topology	residue	spin topology
5.932	0.850	Leu 6 ^a		Leu 6	
5.638	0.967	Ile 2			not assigned
5.150	0.995	Val 8		Val 8	
5.464	0.995	Ile 17		Ile 17	
5.501	1.003	Ile 2		Ile 2	
5.569	1.039	Leu 16		Leu 16	
4.788	1.107	Val 5 ^b		Val 5	
4.673	1.156	Val 8 ^b		Val 8	
1.574	1.722	Lys 21 ^a		Lys 21	
1.418	2.418	Gln 26 ^a		Gln 26	
1.253	2.920	Lys 21		Lys 21	
1.620	4.203	Ala 4		Ala 4	
1.432	4.288	Ala 15		Ala 15	
2.210	4.288	Thr 10		Thr 10	
2.094	4.351	Thr 11		Thr 11	

^a Not the best match. ^b Unique match.

The 2-, 3-, and 4-quantum filtered COSY spectra are combined. The result of merging is listed in Table VII.

On the basis of A_1 – A_4 , all possible MQ spectra are predicted. The numbers of possible peaks are listed in Table VIII.

The comparison result shows that the prediction has included all spin coupling topologies found by manual assignments. There are some MQ peaks, however, having multiple interpretations due to overlap. Table IX gives the comparison of computer assignment and manual assignment. In Table IX, the entry identity is calculated as follows:

$$\text{identity} = A/M$$

where M is manual assignments and A is the number of assignments which are identical to manual assignments. For example, in 4Q spectrum, the program assigns 15 peaks, 12 of them being identical to the manual assignments, giving an identity of 12/15, namely, 80.0%.

More peaks can be assigned by the "best match" rule, although this strategy is considered inadequate for automated assignments. These cases are listed in Table X.

The details of assignment for 4Q spectra of protein melittin are listed in Table XI.

From Table XI, the following conclusions are made:

(1) On the basis of the correlation spectra, COSY and TOCSY, for example, all possible spin coupling topologies and corresponding spectra can be completely predicted by this program.

(2) Assignments based upon "the best match rule" work in most situations but are not accurate for resolving heavy overlap. For bigger protein molecules, these problems can be serious, so that other algorithms must be developed. Recently a CONSTRAINED PARTITIONING ALGORITHM (CPA), FUZZY GRAPH PATTERN RECOGNITION ALGORITHM (FPRA), and TREE SEARCH ALGORITHM (TSA), which solve most of these problems, have been implemented in our laboratory and tested on the raw data of four proteins; the biggest one has 58 residues.²⁶

(3) Assignment with a computer's aid appears to produce better results more quickly. For example, our program assigns six 2Q peaks and one 4Q peak correctly, while these peaks were not assigned by manual methods.

(4) On the basis of the methods proposed in this paper, a spin coupling topological network involving equivalent spins can be directly extracted from MQF-COSY and MQ spectra, by which the heavy overlap can be partial resolved.

(5) MQ experiments are important for ^1H resonances assignments; however, a higher quantum order experiment has less sensitivity. Improved instrumentation can help overcome this.

CPA has also been used in our new program PSE (Protein Structure Elucidation), which has successfully assigned four peptides and a protein at the residue range of 21–58. Details will be reported in future work.

ACKNOWLEDGMENT

The authors thank Dr. Larry R. Brown for many exciting and initial ideas. This work is supported by the Natural

Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES AND NOTES

- (1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
- (2) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon: Oxford, U.K., 1986.
- (3) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. Three-Dimensional Fourier Spectroscopy. Application to High-Resolution NMR. *J. Magn. Reson.* **1989**, *85*, 186–197.
- (4) Gronenborn, A. M.; Clore, G. M. Applications of Three- and Four-Dimensional Heteronuclear NMR Spectroscopy to Protein Structure Determination. *Prog. NMR Spectrosc.* **1991**, *23*, 43–92.
- (5) Hoffman, R. E.; Levy, G. C. Modern Methods of NMR Data Processing and Data Evaluation. *Prog. NMR Spectrosc.* **1991**, *23*, 211–258.
- (6) Kleywegt, G. J.; Boelens, R.; Kaptein, R. A Versatile Approach toward the Partially Automatic Recognition of Cross Peaks in 2D ^1H NMR Spectra. *J. Magn. Reson.* **1990**, *88*, 601–608.
- (7) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. A Common Sense Approach to Peak Picking in Two-, Three-, and Four-Dimensional Spectra Using Automatic Computer Analysis of Contour Diagrams. *J. Magn. Reson.* **1991**, *95*, 214–220.
- (8) James, T. L.; Basus, V. J. Generation of High-Resolution Protein Structures in Solution from Multidimensional NMR. *Annu. Rev. Phys. Chem.* **1991**, *42*, 501–542.
- (9) Majumdar, A.; Hosur, R. V. Simulation of 2D NMR Spectra for Determination of Solution Conformations of Nucleic Acids. *Prog. NMR Spectrosc.* **1992**, *24*, 109–158.
- (10) Kleywegt, G. J.; Boelens, R.; Cox, M.; Llinas, M.; Kaptein, R. Computer-assisted Assignment of 2D ^1H NMR Spectra of Proteins: Basic Algorithms and Application to Phoratoxin B. *J. Biomol. NMR* **1991**, *1*, 23–47.
- (11) van de Ven, F. J. M.; PROSPECT, a Program for Automated Interpretation of 2D NMR Spectra of Proteins. *J. Magn. Reson.* **1990**, *86*, 633–644.
- (12) Billeter, M.; Basus, V. J.; Kuntz, I. D. A Program for Semi-automatic Sequential Resonance Assignments in Protein ^1H Nuclear Magnetic Resonance Spectra. *J. Magn. Reson.* **1988**, *76*, 400–415.
- (13) Cieslar, C.; Clore, G. M.; Gronenborn, A. M. Computer-Aided Sequential Assignment of Protein ^1H NMR Spectra. *J. Magn. Reson.* **1988**, *80*, 119–127.
- (14) Weber, P. L.; Malikayil, J. A.; Mueller, L. Automated Elucidation of J Connectivities in ^1H NMR Spectra. *J. Magn. Reson.* **1989**, *82*, 419–426.
- (15) Eads, C. D.; Kuntz, I. D. Programs for Computer-Assisted Sequential assignment of Proteins. *J. Magn. Reson.* **1989**, *82*, 467–482.
- (16) Zolnai, Z.; Westler, W. M.; Ulrich, E. L.; Markley, J. L. Drafting Table and Light-Box Software for Multidimensional NMR Spectral Analysis (PIXI). The Personal Computer Workstation. *J. Magn. Reson.* **1990**, *88*, 511–522.
- (17) Kleywegt, G. J.; Lamerichs, R. M. J. N.; Boelens, R.; Kaptein, R. Toward Automatic Assignment of Protein ^1H NMR Spectra. *J. Magn. Reson.* **1989**, *85*, 186–197.
- (18) Pfändler, P.; Bodenhausen, G. Automated Analysis of Two-Dimensional Correlation Spectra. Assembly of Fragments into Networks of Coupled Spins. *J. Magn. Reson.* **1990**, *87*, 26–45.
- (19) Kraulis, P. J. ANSIG: A Program for the Assignment of Protein ^1H 2D NMR Spectra by Interactive Computer Graphics. *J. Magn. Reson.* **1989**, *84*, 627–633.
- (20) About the strategy for determining biomolecular structure from NMR, the following paper is interesting: Mirau, P. A Strategy for NMR Structure Determination. *J. Magn. Reson.* **1992**, *96*, 480–490.
- (21) Pfändler, P.; Bodenhausen, G. Topological Classification of Fragments of Coupling Networks and Multiplet Patterns in Two-Dimensional NMR Spectra. *J. Magn. Reson.* **1988**, *79*, 99–123.
- (22) Gray, B. N.; Brown, L. R. Detection of Topologies in Homonuclear Pseudo-3D NMR. *J. Magn. Reson.* **1991**, *95*, 320–340.
- (23) Gross, K. H.; Kalbitzer, H. R. Distribution of Chemical Shifts in ^1H Nuclear Magnetic Resonance Spectra of Proteins. *J. Magn. Reson.* **1988**, *76*, 87–99.
- (24) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic Press: New York, 1976.
- (25) Gray, B. N. Doctoral Thesis, The Australian National University, 1991.
- (26) Xu, J.; Straus, S. K.; Sanctuary, B. C. Automation of Protein 2D Proton NMR Assignment by Means of Fuzzy Mathematics and Graph Theory. To be published.
- (27) Gray, B. N.; Brown, L. R. Detection of Topologies in Homonuclear Pseudo-3D NMR. *J. Magn. Reson.* **1991**, *95*, 320–340.