J-Bio NMR 216

An evaluation of least-squares fits to COSY spectra as a means of estimating proton–proton coupling constants II. Applications to polypeptides

Ju-Xing Yang, Andrzej Krezel, Peter Schmieder, Gerhard Wagner and Timothy F. Havel*

Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.

Received 23 May 1994
Accepted 6 July 1994

Keywords: DQF-COSY; Correlated spectroscopy; Proton–proton coupling constants; Parameter estimation; Peptide and protein conformation; Decorsin; VDA

SUMMARY

A new computational method for simultaneously estimating all the proton–proton coupling constants in a molecule from COSY spectra [Yang, J.-X. and Havel, T.F. (1994) *J. Biomol. NMR*, **4**, 807–826] is applied to experimental data from two polypeptides. The first of these is a cyclic hexapeptide denoted as VDA (-D-Ala^1-Phe^2-Trp^3-Lys(Z)^4-Val^5-Phe^6-), in deuterated DMSO, while the second is a 39-residue protein, called decorsin, in aqueous solution. The effect of different data processing strategies and different initial parameter values on the accuracy of the coupling constants was explored. In the case of VDA, most of the coupling constants did not depend strongly on the initial values chosen for the optimization or on how the data were processed. This, together with our previous experience using simulated data, implies strongly that these values are accurate estimates of the coupling constants. They also differ by an average of only 0.36 Hz from the values of the 14 coupling constants that could be measured independently by established methods. In the case of decorsin, many of the coupling constants exhibited a moderate dependence on their initial values and a strong dependence on how the data were processed. With the most successful data processing strategy, the amide-α coupling constants differed by an average of 1.11 Hz from the 21 values that could be measured by established methods, while two thirds of the three-bond coupling constants fell within 1.0 Hz of the ranges obtained by applying the Karplus relation to an independently computed ensemble of distance geometry structures. The averages of the coupling constants over multiple optimizations using random initial values were computed in order to obtain the best possible estimates of the coupling constants. Most clearly incorrect averages can be identified by large standard deviations in the coupling constants or the associated line widths and chemical shifts, and can be explained by strong coupling and/or overlap with the water signal, the diagonal peaks or other cross peaks.

*To whom correspondence should be addressed.

INTRODUCTION

In the preceding paper (Yang and Havel, 1994), we presented a new computational method which can, in principle, simultaneously estimate *all* of the proton–proton coupling constants in a molecule. This method functions by finding a nonlinear least-squares fit between a simulated COSY spectrum and an experimental target spectrum. The results obtained by applying this method to a variety of simulated target spectra (Yang and Havel, 1994) demonstrated that the method can return reasonably accurate estimates of the coupling constants used to simulate the spectra, even in the presence of significant levels of noise and artifacts, and hence should yield useful results with experimental target spectra as well. Simulated test problems, of course, have the substantial advantage that not only the correct answer, but also the exact level and types of noise and artifacts in the spectra are known for certain. Nevertheless, there remains the possibility that the uncontrolled artifacts, extensive overlap and strong coupling effects present in typical experimental spectra will prevent the same method from being applied successfully to these spectra. The purpose of the present paper is to show that, although these problems do cause the method to fail for some coupling constants, it is nevertheless capable of producing useful results. In fact, the values that it appears to have correctly estimated probably constitute the most complete list of proton–proton coupling constants that has ever been obtained for a protein.

Our experience with the simulated test problems implies that a fairly high digital and signal resolution is required in order to obtain good results, particularly with large molecules for which the intrinsic line width is large. In addition, because we use a Gaussian line shape in our simulated spectra, it is clearly advisable to process the data so as to obtain a similar line shape in the experimental spectra; this has the additional benefit of improving the resolution of the spectrum (Bodenhausen et al., 1977; Pearson, 1987). By comparing the results obtained using different data processing strategies, it is possible to estimate just how sensitive the method is to artifacts in actual spectra. In addition, we repeated our calculations starting from several different sets of random values for the coupling constants, in order to identify those coupling constants that were not reliably determined by the fitting procedure. Their precision should be improved by averaging these coupling constants over several such runs.

EXPERIMENTAL METHODS

For our study, two high-resolution double-quantum-filtered COSY (DQF-COSY) spectra were collected (Piantini et al., 1982; Rance et al., 1983; Shaka and Freeman, 1983). The sample for the first of these was a cyclic hexapeptide in deuterated DMSO, retro-valine-D-alanine-008 (VDA), with the sequence -D-Ala^1-Phe^2-Trp^3-Lys(Z)^4-Val^5-Phe^6- (Kessler et al., 1988). The second was an aqueous solution of a 39-residue RGD-containing platelet aggregation inhibitor known as decorsin (Seymour et al., 1990). With VDA, the spectrum was of such high resolution and overall quality as to make it quite easy to derive the amide-alpha coupling constants by the method of Kim and Prestegard (1989), and the alpha-beta coupling constants by the DISCO method (Kessler et al., 1985). With decorsin, spectrometer drift during the three days required to collect the spectrum with sufficient resolution and signal averaging produced substantial artifacts, which in turn necessitated that special measures be taken in processing the data. The amide-alpha coupling constants were obtained by the method of Kim and Prestegard (1989), while bounds on all the coupling constants in the molecule were derived by applying the Karplus equation (De

Marco et al., 1978; Vuister and Bax, 1993) to an ensemble of 25 distance geometry structures that was computed independently (Krezel et al., 1994). The results of these measurements, together with the geminal and other fixed coupling constants, provide us with standards by which we can judge the quality of our results.

Data collection

The DQF-COSY spectrum of VDA was collected on a 500 MHz Varian spectrometer, using a 30 mM solution in deuterated DMSO at 298 K, a sweep width of 5800 Hz, and averaging over four scans with a standard four-step phase cycle. A total of 8192 FIDs were collected in TPPI–States format, where each FID contained 4096 complex points.

The DQF-COSY spectrum of decorsin was collected on a 500 MHz Bruker spectrometer, using a 5 mM solution in a 1:10 mixture of $D_2O:H_2O$ at 303 K and pH 4.7, with a sweep width of 7042 Hz. The 16-step phase cycle of Derome and Williamson (1990) was used, with CYCLOPS added to suppress quadrature artifacts, for a total of 64 scans. Water suppression was achieved by low-level presaturation during the relaxation delay of 1.5 s. A total of 2048 FIDs were collected in TPPI–States format, where each FID contained 2048 complex points.

Data processing for VDA

In order to find the optimum data processing strategy for use with our coupling constant estimation procedure, and to determine the sensitivity of our results to typical perturbations in the spectra due to differences in how they are processed, we prepared spectra using the following three procedures:

VDA(1): This spectrum was prepared by direct Fourier transformation of the 4K by 4K hypercomplex data set, with no apodization of any kind.

VDA(2): A Gaussian function with a variance designed to give a 2 Hz line width (Pearson, 1987) was used as the window function along t_2, with the maximum displaced by 200 points from the origin to approximately match the maximum of the FIDs. This same window function was also applied along the t_1 dimension after the first transformation.

VDA(3): The rms average of the absolute values of the FIDs was computed in order to obtain their decay envelope relatively free of interference effects. A linear least-squares fit was then calculated to the logarithm of this envelope over points 200 to 4096, to obtain the decay rate λ*. The FIDs were then divided by $exp(-\lambda t)$ to remove the Lorentzian component from the line shapes, and then processed exactly as described above for VDA(2), except that the interferogram was also divided by $exp(-\lambda t)$ before the second transform.

No baseline correction was used in any of these spectra, as this can distort the line shapes. The least-squares fits to these spectra were based on only one half of the full spectrum at a time, either the upper (VDA($*$, U)), lower (VDA($*$, L)) or either half of the symmetrized spectrum (VDA($*$, S)).

Data processing with decorsin

In the case of decorsin, spectrometer instability caused small shifts in the presaturation frequency during the experiment, with the two largest shifts occurring at about FID 800 and 1350.

*Due to the uneven distribution of coupling constants in polypeptides and the sinusoidal dependence of their contributions to each FID, significant interference effects remained. However, the average line widths implied by the exponential decay rate that we obtained by this procedure were reasonable.

This resulted in a significant water signal along t_1 as well as the usual signal along t_2, which had to be eliminated so that the entire amide region could be seen clearly. Eventually this was done by applying a 60-point binomial filter (i.e., the 'smo' command in the FELIX program) to the FIDs, and then scaling down the initial points of each FID by applying a window function to these points, of the form $sin^2((\pi/2)(i/i_0))$. Here $i \leq i_0$ is the point index in each FID, while i_0 was increased linearly from 10 at the first FID to 50 at the 800th FID, and kept at 50 for all subsequent FIDs. This procedure smoothly scales down the initial points of the FIDs to zero at the start of each FID, and will be called *initial point filtering* in what follows.

As with VDA, three different spectra were generated from this data set:

Dec(1): The standard cos^2 window function was applied to all 2048 points of each FID after 60-point binomial filtering, after which the FIDs were zero-filled to 4096 points and transformed. Instead of using initial point filtering as above, however, only the first 800 FIDs were used, i.e. the cos^2 window function was applied only to the first 400 points in t_1, followed by zero-filling to 4096 points and transformation. This is, of course, very similar to how DQF-COSY spectra are usually collected and processed.

Dec(2): In this case all 2048 FIDs were subjected to the initial point filtering procedure described above, and a Gaussian window function applied to all the FIDs that was designed to give a 7 Hz line width and was centered on the 200th point. Since this window function does not go smoothly to zero over a finite range, like a cos^2 window function does, zero-filling resulted in unacceptable artifacts. Instead, each FID was filled to 4096 points by concatenating its reverse onto its end prior to applying the Gaussian window function. This reverse concatenation was applied twice to the interferogram in the t_1 dimension to fill it out to 4096 points, and the same Gaussian window function was applied, centered this time on the 100th point, followed by transformation to obtain the final spectrum.

Dec(3): This spectrum was obtained by the same procedure as Dec(2), except that here a least-squares fit to the log of the rms absolute values of the FIDs over points 200 through 2048 was used to obtain their decay rate λ, after which the FIDs and interferogram were divided by $exp(-\lambda t)$ to eliminate the Lorentzian component from the line shapes, just as with VDA(3) above.

As with VDA, no baseline correction was used, but with decorsin the least-squares fits were performed on only the 'best' regions (Dec.(*, B)), i.e., those furthest from the water line and other intense artifacts. These regions included the above-diagonal fingerprint region and below-diagonal aliphatic region. The 30 points of each row nearest to the diagonal were not included in the regions.

Computational protocol

The computational procedure followed is described in detail in the companion paper (Yang and Havel, 1994). The starting values for the minimizations consisted of the measured chemical shifts, uniform line widths of 2.0 Hz for VDA and 6.0 Hz for decorsin, −14 Hz for the two-bond coupling constants and 7.0 ± 0.1 Hz for the three-bond. With the best spectra, i.e. VDA(3, S) and Dec(3, B), further minimizations were performed using random values in the range 7.0 ± 5.0 Hz for the three-bond coupling constants, in order to determine the severity of the local-minimum problem.

With experimental target spectra, the scale factor which gives the simulated spectra the same overall intensity as the target spectra is not known a priori, and as shown in the companion paper it is important to choose this scale factor correctly. Because differences in the cancellation of in-phase and antiphase peaks in a 2D spectrum make it difficult to compare the intensities of different spectra, the scale factor was computed from the pseudo-1D spectra, as follows: first, a region of the experimental spectrum relatively free of artifacts is chosen, in this case the fingerprint region. The rest of the spectrum is then set to zero, and the spectrum is back-transformed to an interferogram, after which the absolute values of the rows are added together to obtain the pseudo-1D spectrum. The sum of the values of all the points in this pseudo-1D spectrum is then taken as a measure of its intensity. This same procedure was applied to a spectrum simulated using a scale factor of one, and the ratio of the intensities of the two pseudo-1D spectra was then used as the scale factor for all subsequent simulations.

Another problem that surfaced during these calculations was that, in regions of the spectrum that were missing cross peaks (presumably due to exchange or water suppression), the minimization 'attempted' to do the same by making the line widths of those peaks much less than the frequency difference between the points of the spectrum. These sharp peaks resulted in essentially discontinuous changes in the sum of squares, which caused the minimizer to fail. Eventually, this problem was corrected by defining a pseudo-width β, which is related to the actual line width α as

$$\alpha = \begin{cases} \beta & \text{if } \alpha > \delta \\ \frac{1}{2}\left(\delta + \beta^2/\delta\right) & \text{otherwise} \end{cases} \qquad (1)$$

This limits the actual line width to at least half the difference δ between the points, and the actual line width smoothly becomes equal to the pseudo-width as the latter increases towards δ.

RESULTS AND DISCUSSION

The computer runs made with the spectra described in the Methods section are summarized in Table 1. This section discusses the results shown in the remaining tables and figures, first for the cyclic peptide VDA, and then for the protein decorsin. The VDA spectra are obviously very clean, while the decorsin data exhibits most of the problems that are typically present in the NMR spectra of biological macromolecules in aqueous solution, including significant impurities. Thus, except for their relatively high resolution, these two data sets roughly cover the range of quality that is commonly encountered with peptides and small proteins.

Results with VDA

It can readily be seen from Tables 1 and 2 that essentially the same results were obtained for VDA, regardless of how the data were processed as well as whether the lower half, the upper half or the symmetrized spectrum were used. The correlations among the absolute values of the 28 two- and three-bond coupling constants computed in these runs usually exceeded +90%, with average differences that were usually well under 1 Hz. The complete results obtained with the VDA symmetrized spectra are shown in Table 3. The first three columns of coupling constants (VDA(1, S), VDA(2, S) and VDA(3, S)) again demonstrate that very similar values were obtained, regardless of how the data were processed, with the exception of the coupling constants involving the γ-methylene protons of the lysine residue. These exceptions are easily explained by the fact that the chemical shifts of the γ-protons and the fully degenerate δ-protons of this residue

TABLE 1
DESCRIPTIONS OF THE CALCULATIONS PERFORMED AND THEIR ABBREVIATIONS[a]

Abbreviation	Molecule	Window function	Lorenzian line-shape correction	Regions taken	Starting couplings (Hz)
VDA(1, {U,L,S})		None	No	U = Upper half	7+/−0.1
VDA(2, {U,L,S})	VDA	Gaussian	No	L = Lower half	7+/−0.1
VDA(3, {U,L,S})	(6 residues)	Gaussian	Yes	S = Symmetrized	7+/−0.1
VDA(3, S; 1–5)		Gaussian	Yes		7+/−5.0
Dec(1, B)		Cosine square	No	B = Best regions	7+/−0.1
Dec(2, B)	Decorsin	Gaussian	No		7+/−0.1
Dec(3, B; 0)	(39 residues)	Gaussian	Yes		7+/−0.1
Dec(3, B; 1–5)		Gaussian	Yes		7+/−5.0

[a] The first column shows the abbreviation used in the text and other tables for the calculations described by subsequent columns. These columns give the name of the molecule, the window function, whether or not a Lorentzian line-shape correction was used, one-letter specifications of which regions of the spectra were used for computing the least-squares fits, and the starting values of the coupling constants used by the minimization. In the latter case, 7+/−0.1 means that the starting couplings were a uniformly distributed random number in the range [6.9,7.1] Hz. A more complete description of the spectra, the data processing methods, and the calculations is given in the text.

are less than 0.15 ppm apart, and hence are both strongly coupled and have cross peaks which heavily overlap the diagonal and each other.

The sixth column of Table 3 gives the maximum absolute difference between the coupling constants obtained for VDA(3, S), which used 7.0 ± 0.1 Hz as the starting values of all coupling constants for the minimization, with an additional five runs in which the starting values were chosen randomly in the range 7.0 ± 5.0 Hz (VDA(3, S; 1–5)). These differences were so small they had to be reported in thousandths of a Hz, which demonstrates that no significant problems with local minima were encountered. Finally, the amide-alpha and alpha-beta coupling constants that could be measured in the spectrum VDA(1, S) using the Kim and Prestegard (1989) and DISCO

TABLE 2
CORRELATIONS AMONG THE COUPLING CONSTANTS OBTAINED FROM THE VDA CALCULATIONS AND THE AVERAGE DIFFERENCES BETWEEN THEM[a]

	VDA(1, U)	VDA(1, L)	VDA(1, S)	VDA(2, U)	VDA(2, L)	VDA(2, S)	VDA(3, U)	VDA(3, L)	VDA(3, S)
VDA(1, U)		98.1	98.6	99.9	98.0	98.4	92.8	89.8	92.7
VDA(1, L)	0.43		99.6	98.0	99.9	99.4	85.5	83.4	85.9
VDA(1, S)	0.32	0.21		98.6	99.7	100.0	87.8	85.7	88.4
VDA(2, U)	0.09	0.44	0.28		98.0	98.5	93.2	90.4	93.2
VDA(2, L)	0.45	0.11	0.19	0.45		99.6	86.0	84.1	86.7
VDA(2, S)	0.39	0.26	0.08	0.33	0.20		88.0	85.8	88.6
VDA(3, U)	0.58	0.89	0.80	0.62	0.93	0.83		97.6	99.4
VDA(3, L)	0.69	0.82	0.80	0.71	0.86	0.85	0.38		99.2
VDA(3, S)	0.57	0.78	0.69	0.57	0.82	0.73	0.22	0.21	

[a] Correlations are given in percent in the upper half of the matrix and the average differences (in Hz) in the lower half.

(Kessler et al., 1985) methods, respectively, are reported in the last column of Table 3, together with the $C^{\delta}H$-$N^{\varepsilon}H$ coupling constant of the tryptophan residue that could also be measured by the method of Kim and Prestegard. The average difference between these values and those obtained by our least-squares fits was only 0.36 Hz. Most of this difference was due to the 1.24 Hz discrepancy between the values of the tryptophan $C^{\delta}H$-$N^{\varepsilon}H$ coupling constant, where the value obtained by both methods is clearly too small.

Figure 1 shows the amide-alpha and alpha-beta experimental and best-fit simulated cross peaks that were obtained with VDA(3, S). The fact that our method was able to accurately reproduce

TABLE 3
COUPLING CONSTANTS OBTAINED FROM THE VDA CALCULATIONS[a]

Amino acid	Coupling constant type	Coupling constant (Hz)			Differences to VDA(3, S; 1–5) (Hz × 10^{-3})	Manually measured value (Hz)
		VDA(1, S)	VDA(2, S)	VDA(3, S)		
Ala[1]	NH-$C^{\alpha}H$	5.45	5.36	5.66	0.14	6.01
	$C^{\alpha}H$-$C^{\beta}H^{*}$	7.17	7.19	7.10	0.05	
Phe[2]	NH-$C^{\alpha}H$	8.99	9.01	8.90	0.32	8.68
	$C^{\alpha}H$-$C^{\beta}H^{H}$	10.52	10.46	10.67	0.10	9.97
	$C^{\alpha}H$-$C^{\beta}H^{L}$	3.40	3.32	3.51	0.07	3.48
	$C^{\beta}H^{H}$-$C^{\beta}H^{L}$	−14.13	−14.20	−13.99	0.08	−13.85
Trp[3]	NH-$C^{\alpha}H$	9.40	9.43	9.24	0.23	9.09
	$C^{\alpha}H$-$C^{\beta}H^{H}$	9.41	9.34	9.65	0.04	9.75
	$C^{\alpha}H$-$C^{\beta}H^{L}$	4.35	4.27	4.53	0.16	4.48
	$C^{\beta}H^{H}$-$C^{\beta}H^{L}$	−15.69	−15.74	−15.49	0.05	−15.41
	$C^{\delta}H$-$N^{\varepsilon}H$	2.68	2.85	2.66	0.29	3.90
Lys[4]	NH-$C^{\alpha}H$	6.30	6.41	6.06	0.49	6.00
	$C^{\alpha}H$-$C^{\beta}H^{*}$	7.33	7.35	7.25	0.33	
	$C^{\beta}H^{*}$-$C^{\gamma}H^{H}$	7.62	7.66	7.99	4.45	
	$C^{\beta}H^{*}$-$C^{\gamma}H^{L}$	5.96	5.63	5.30	5.04	
	$C^{\gamma}H^{H}$-$C^{\gamma}H^{L}$	−12.19	−12.23	−7.56	27.83	
	$C^{\gamma}H^{H}$-$C^{\delta}H^{*}$	3.23	3.35	6.67	2.15	
	$C^{\gamma}H^{L}$-$C^{\delta}H^{*}$	0.00	0.00	6.63	2.08	
	$C^{\delta}H^{*}$-$C^{\varepsilon}H^{*}$	7.25	7.27	7.00	0.08	
	$C^{\varepsilon}H^{*}$-NH^{1}	5.73	5.73	5.76	0.15	
Val[5]	NH-$C^{\alpha}H$	8.79	8.97	8.31	0.16	8.24
	$C^{\alpha}H$-$C^{\beta}H$	4.81	4.65	5.07	0.12	4.16
	$C^{\beta}H$-$C^{\gamma}H^{H}$	6.78	6.76	6.76	0.02	
	$C^{\beta}H$-$C^{\gamma}H^{L}$	6.77	6.76	6.77	0.03	
Phe[6]	NH-$C^{\alpha}H$	4.16	4.08	4.17	0.37	4.56
	$C^{\alpha}H$-$C^{\beta}H^{H}$	7.94	7.86	8.14	0.10	8.39
	$C^{\alpha}H$-$C^{\beta}H^{L}$	5.16	5.06	5.36	0.18	5.56
	$C^{\beta}H^{H}$-$C^{\beta}H^{L}$	−13.72	−13.79	−13.53	0.10	−12.91

[a] The first column gives the amino acid and its sequence number in VDA. The second identifies the coupling constant. The next three columns give the value obtained for that coupling constant in the runs denoted by VDA(1, S), VDA(2, S) and VDA(3, S) in Table 1. The next column gives the average differences between the values obtained in runs VDA(3, S; 1–5 and VDA(3, S). The last column gives the values of those coupling constants that could be measured manually using either the Kim and Prestegard (1989) or DISCO (Kessler et al., 1985) methods, as described in the text.

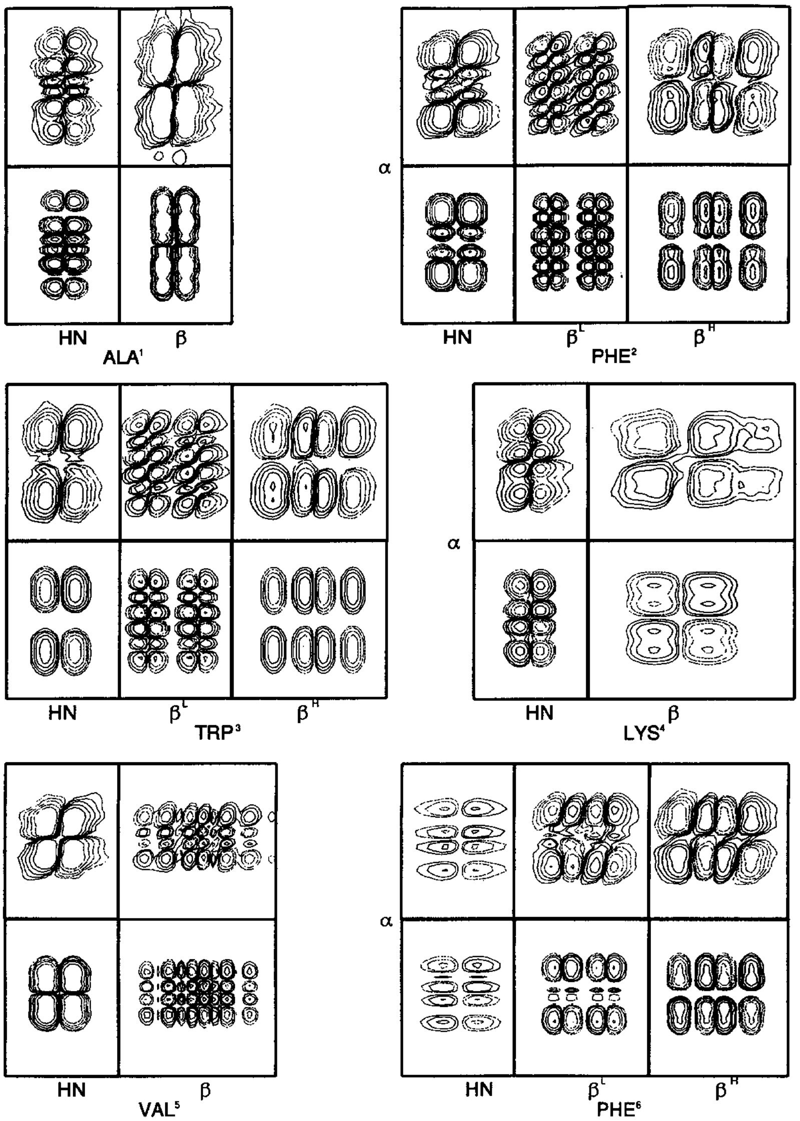

Fig. 1. The experimental (above) and best-fit (below) amide-alpha and alpha-beta cross peaks obtained in the run denoted by VDA(3, S) in Table 1. The axes correspond to the chemical shifts of the protons connected by each cross peak, and are labeled by the positions of the protons in the amino acid with the superscript 'L' or 'H' for low field and high field, respectively.

the fine structures of these cross peaks is obvious. The most apparent difference is in the bases of the cross peaks, which are clearly broader in the experimental spectra, probably because of a residual Lorentzian component in the line shape. Finally, we stress that, although we have displayed the cross peaks separately in this figure in order to enlarge them, in the actual calculations all these cross peaks were fitted simultaneously.

Results with decorsin

Table 4 shows the correlations and average differences among the coupling constants that were obtained for decorsin, using the data processing methods and minimization starting values summarized in Table 1 (and described fully in the Methods section). It is immediately apparent that the data processing strategy had a pronounced effect on the results obtained, while large random variations in the starting values of the coupling constants had a lesser but still noticeable effect.

For comparison, we also computed the ranges in the values of the three-bond coupling constants that were obtained on applying the Karplus relation to an ensemble of 25 structures (using the parameters given by De Marco et al. (1978), except for the amide-alpha parameters, which were taken from Vuister and Bax (1993)). This ensemble was computed independently from NOESY data by distance geometry methods, without any torsion angle constraints derived from coupling constants (Krezel et al., 1994). In addition, the values of all the amide-alpha coupling constants whose active cross peaks did not overlap the water line were determined by applying the method of Kim and Prestegard (1989) to the spectrum Dec(1, B).

Table 5 shows the percentages of the various types of coupling constants that fell within the ranges obtained by applying the Karplus relation to the distance geometry ensemble, as described above. The general trend toward better agreement with these independently established values that is observed on processing the data so as to obtain a more Gaussian-like line shape is highly encouraging, because this was expected on the basis of our experience with simulated test problems (Yang and Havel, 1994). With the best data processing strategies, about half the coupling constants fell within these ranges, and about two thirds were within 1.0 Hz of their ranges. These numbers do not depend greatly on whether we use all the coupling constants, only those involving stereospecifically assigned protons, or only the amide-alpha coupling constants. Using only those coupling constants involving stereospecifically assigned protons whose range was also less than 2.0 Hz over the entire distance geometry ensemble made these percentages somewhat smaller, with 23% in the range, 32% within 0.5 Hz, 47% within 1.0 Hz and 75% within 2.0 Hz.

In all the data processing strategies employed with decorsin, some of the two-bond coupling constants between nearly degenerate methylene proton pairs came out drastically too large (greater

TABLE 4
CORRELATIONS AMONG THE COUPLING CONSTANTS OBTAINED FROM THE DECORSIN CALCULATIONS AND THE AVERAGE DIFFERENCES BETWEEN THEM[a]

Abbreviation	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec
	(1, B)	(2, B)	(3, B)	(3, B; 1)	(3, B; 2)	(3, B; 3)	(3, B; 4)	(3, B; 5)
Dec(1, B)		61.3	63.7	69.0	61.9	63.0	60.2	65.8
Dec(2, B)	3.42		77.7	63.6	74.1	63.5	72.4	66.4
Dec(3, B)	3.49	1.23		90.5	97.4	94.1	98.3	96.2
Dec(3, B; 1)	3.31	1.72	0.39		89.2	90.7	85.5	95.9
Dec(3, B; 2)	3.59	1.49	0.20	0.41		89.6	96.6	92.3
Dec(3, B; 3)	2.04	1.78	0.34	0.33	0.47		92.7	95.8
Dec(3, B; 4)	3.63	1.57	0.17	0.48	0.25	0.34		92.5
Dec(3, B; 5)	3.44	1.81	0.30	0.23	0.40	0.26	0.40	

[a] Correlations are given in percent in the upper half of the matrix and the average differences (in Hz) in the lower half.

than −10 Hz) or too small (less than −20 Hz). As a rule, the cross peaks in which these two-bond couplings were active were under the diagonal and hence omitted from the fits, so that these coupling constants were determined by cross peaks in which they played a passive role. When these passive cross peaks were heavily overlapped with other cross peaks, however, the sensitivity of the fit to the two-bond coupling was poor and hence the value obtained was easily corrupted by noise and artifacts. Which two-bond couplings came out incorrect changed significantly depending on how the data were processed, with 15 out of 45 incorrect with Dec(1, B), seven incorrect with Dec(2, B) and only five incorrect with Dec(3, B). Because these obviously incorrect two-bond coupling constants may further corrupt the values of other coupling constants via cross peaks in which they play a passive role, in the final runs, designed to obtain the best possible values of all the coupling constants, the two-bond coupling constants that came out greater than −10 or less than −20 Hz in Dec(3, B) were fixed at a more reasonable value of −14 Hz.

In these final runs (Dec(3, B; 0–5)), the starting values for the coupling constants were chosen randomly in the interval 7.0 ± 5.0 Hz, except in Dec(3, B; 0) where values of 7.0 ± 0.1 Hz were used as before. The values obtained for the coupling constants over all six runs were averaged (subject to the line-width screen described below), and are reported along with their standard deviations in Table 6. As shown in Table 5, the average three-bond coupling constants agreed with the ranges obtained by applying the Karplus relation to the distance geometry ensemble about as well as the coupling constants from the Dec(3, B) run, from which they differed by only

TABLE 5
PERCENTAGE OF COUPLING CONSTANTS IN OR WITHIN GIVEN DISTANCES OF THE RANGES COMPUTED BY APPLYING THE KARPLUS RELATION TO THE DISTANCE GEOMETRY ENSEMBLE FOR DECORSIN[a]

Abbreviation	Coupling type	In range	<0.5 Hz	<1.0 Hz	<2.0 Hz
Dec(1, B)	Amide-alpha	25.7	40.0	54.3	68.6
Dec(2, B)	coupling constants only	45.7	60.0	65.7	80.0
Dec(3, B)		45.7	54.3	65.7	82.9
Dec(3, B; 0–5)		45.7	54.3	68.6	80.0
Dec(1, B)	Coupling constants	19.8	27.8	40.5	62.7
Dec(2, B)	involving only	33.3	49.2	59.5	80.2
Dec(3, B)	stereospecifically	38.1	49.2	61.1	79.4
Dec(3, B; 0–5)	assigned resonances	37.7	49.2	60.7	81.2
Dec(1, B)	All three-bond	30.7	36.4	46.0	65.3
Dec(2, B)	coupling constants	47.7	59.7	67.6	84.1
Dec(3, B)		52.8	60.8	69.3	83.5
Dec(3, B; 0–5)		53.5	61.8	70.6	85.3

[a] The ranges were computed by applying the Karplus relation to an ensemble of 25 distance geometry structures that were computed without using any coupling constant information, as described in the text. The percentages of each of the classes of coupling constants specified (second column) that are within their ranges (third column), within the range or 0.5 Hz of one of its end points (fourth column), 1.0 Hz (fifth column) or 2.0 Hz (sixth column) are given. For coupling constants involving nonstereospecifically assigned resonances, the smallest contiguous interval containing the ranges of all of the individual coupling constants was used. The statistics reported for Dec(3, B; 0–5) use the average coupling constants reported in Table 6.

0.37 Hz on average. Those coupling constants that are likely to be seriously in error should usually be revealed by their larger standard deviations, and possibly by fluctuations in the associated chemical shifts and line widths as well. We emphasize that the standard deviations reported in Table 6 should *not* be taken as estimates of the accuracy of the coupling constants reported there, but only as an indication of how reproducible the computational results were.

The fluctuations in the chemical shifts over runs Dec(3, B; 0–5) were less than 0.5 Hz for 85% of the resonances (data not shown in the tables). In a few cases the chemical shift fluctuations were above 5.0 Hz, and these were indeed associated with large fluctuations in the values of the coupling constants. In perhaps the most striking case, the γ-protons and one β-proton of Gln^{26} were all within 0.1 ppm of each other and hence heavily overlapped with the diagonal and each other. This led to chemical shift fluctuations of up to 8.5 Hz and fluctuations in the coupling constants of more than 7 Hz (see Table 6). Similar problems, with clearly incorrect coupling constants near 0 Hz, were encoutered in several other side chains (e.g. Pro^{5}, Lys^{19} and Pro^{36}), but were not associated with large fluctuations in the chemical shifts.

Although the larger fluctuations in the line widths also appeared to correlate with fluctuations in the coupling constants involving the same resonances, as a rule the fluctuations in the line widths were rather large anyway: the standard deviation in the line widths averaged 6.9 Hz over all resonances, and for 25% of the resonances the standard deviation exceeded 50% of the mean. It should be noted that these line-width fluctuations were an order of magnitude larger than those observed in our simulated test problems (Yang and Havel, 1994). Although in most cases the fluctuations in the line widths were not accompanied by comparable fluctuations in the values of the associated coupling constants, as described below we did observe some 'line width versus coupling constant' local minima, like those encountered in our simulated test problems (Yang and Havel, 1994). To screen out these and any other clear-cut problems, coupling constants with values less than 1.0 Hz or connecting protons with associated line widths less than 3 or greater than 30 Hz were not used in computing the averages and standard deviations reported in Table 6. With a few coupling constants, none of the values obtained in Dec(3, B; 0–5) passed this screen; these are marked 'discarded' in Table 6. Overall, however, the screen eliminated only 5% of the values that would otherwise have been used to compute the averages and deviations.

Figures 2a and b show a portion of the fingerprint region of the experimental and best-fit spectra obtained for Dec(3, B; 0), while Figs. 2c and d show a portion of the alpha-aliphatic region. The boxed cross peaks in the experimental spectra that are missing in the simulated spectra are due to the presence of impurities in the sample; their presence should affect the accuracy of the computed coupling constants only if they overlap one of the decorsin cross peaks, which they seldom seem to do. It can immediately be seen from these figures that the qualitative match between the experimental and best-fit cross-peak patterns is usually quite good. Nevertheless, there were also a few cases in which the fitting procedure failed to reproduce the basic shapes or sign patterns of the cross peaks, which are indicated by the boxed cross peaks in the best-fit spectra (Figs. 2b and d).

For example, the tetraplet cross peak at ca. $D_1 = 8.15$ and $D_2 = 4.35$ ppm in the experimental spectrum (Fig. 2a) that came out as an elongated octaplet (with an NH line width of 48 Hz) in the best-fit spectrum (Fig. 2b) is due to the N-terminal alanine, which was simulated as bearing an NH_3 group, but which is evidently either modified or overlapped with an impurity in the actual sample. The weak cross peak at ca. $D_1 = 2.85$ and $D_2 = 4.75$ ppm in the experimental spectrum

TABLE 6
AVERAGES AND STANDARD DEVIATIONS OF COUPLING CONSTANTS (Hz) OBTAINED FROM THE Dec(3, B; 0–5) CALCULATIONS[a]

Amino acid	NH-$C^{\alpha}H$	$C^{\alpha}H$-$C^{\beta}H$	Two-bond	Others
Ala^{1}	~~2.64~~ (0.00)	~~7.17~~ (0.00)		
Pro^{2}		$\underline{4.11}$ (0.00) 7.93 (0.00)	β: −15.31 (0.00) δ: ~~−10.83~~ (0.00)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{3.35}$ (0.01); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 4.79 (0.00) $C^{\gamma}H^{*}$-$C^{\delta}H^{H}$: 6.04 (0.00); $C^{\gamma}H^{*}$-$C^{\delta}H^{L}$: 5.29 (0.00)
Arg^{3}	9.08 (0.00)	6.62 (0.01) 4.66 (0.00)	β: −12.71 (0.01)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: 4.09 (0.00); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 7.30 (0.00) $C^{\gamma}H^{*}$-$C^{\delta}H^{*}$: 6.78 (0.00); $C^{\delta}H^{*}$-$N^{\epsilon}H$: 6.11 (0.00)
Leu^{4}	6.48 (0.00)	2.17 (0.00) $\underline{9.88}$ (0.00)	β: −15.00 (0.00)	$C^{\beta}H^{H}$-$C^{\gamma}H$: 4.53 (0.00); $C^{\beta}H^{L}$-$C^{\gamma}H$: $\underline{3.23}$ (0.00) $C^{\gamma}H$-$C^{\delta}H^{*H}$: 7.78 (0.00); $C^{\gamma}H$-$C^{\delta}H^{*L}$: 7.93 (0.00)
Pro^{5}		$\underline{3.73}$ (0.02) 8.30 (0.06)	β: −12.37 (0.06)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{4.37}$ (0.30); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: discarded $C^{\gamma}H^{*}$-$C^{\delta}H^{*}$: discarded
Gln^{6}	6.38 (0.01)	6.20 (0.72) 7.90 (0.66)	β: −14.00 (fixed)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: 4.37 (0.30); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 11.27 (0.34)
Cys^{7}	~~1.02~~ (0.00)	~~$\underline{5.72}$~~ (0.07) ~~1.15~~ (0.02)	β: −13.10 (0.13)	
Gln^{8}	~~7.34~~ (0.01)	9.07 (0.02) 4.06 (0.04)	β: −13.74 (0.01)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: 6.06 (0.01); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 8.24 (0.00)
Gly^{9}	~~1.57~~ (0.00) 7.95 (0.00)		α: −15.38 (0.00)	
Asp^{10}	6.45 (0.01)	6.71 (0.05) $\underline{4.88}$ (0.00)	β: −16.32 (0.03)	
Asp^{11}	4.09 (0.01)	$\underline{12.40}$ (0.00) 4.03 (0.00)	β: −12.79 (0.00)	
Gln^{12}	7.32 (0.01)	$\underline{11.09}$ (0.01) 4.55 (0.01)	β: −12.20 (0.02) γ: −14.00 (fixed)	$C^{\beta}H^{H}$-$C^{\gamma}H^{H}$: ~~$\underline{1.76}$~~ (0.07); $C^{\beta}H^{H}$-$C^{\gamma}H^{L}$: $\underline{5.09}$ (0.08) $C^{\beta}H^{L}$-$C^{\gamma}H^{H}$: discarded; $C^{\beta}H^{L}$-$C^{\gamma}H^{L}$: 5.39 (0.04)
Glu^{13}	~~5.66~~ (0.31)	~~$\underline{6.76}$~~ (0.91) ~~3.89~~ (0.05)	β: −14.64 (0.87)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{6.65}$ (0.55); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 8.16 (0.33)
Lys^{14}	4.10 (0.00)	$\underline{6.97}$ (0.03) 6.16 (0.04)	β: −14.00 (fixed)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{11.76}$ (0.03); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 18.10 (0.03) $C^{\gamma}H^{*}$-$C^{\delta}H^{*}$: 4.03 (0.00); $C^{\delta}H^{*}$-$C^{\epsilon}H^{*}$: 8.20 (0.00) $C^{\epsilon}H^{*}$-$N^{\zeta}H^{*}$: ~~3.17~~ (0.00)
Cys^{15}	8.19 (0.00)	4.54 (0.00) $\underline{2.55}$ (0.00)	β: −14.42 (0.00)	
Leu^{16}	6.68 (0.01)	7.52 (0.03)		$C^{\beta}H$-$C^{\gamma}H$: 4.35 (1.03) $C^{\gamma}H$-$C^{\delta}H^{*H}$: ~~7.56~~ (2.37); $C^{\gamma}H$-$C^{\delta}H^{*L}$: ~~8.98~~ (1.90)
Cys^{17}	7.66 (0.00)	$\underline{2.99}$ (0.00) 11.27 (0.00)	β: −13.78 (0.00)	
Asn^{18}	6.91 (0.00)	$\underline{3.87}$ (0.00) 11.99 (0.00)	β: −17.22 (0.00)	
Lys^{19}	~~9.68~~ (0.00)	~~8.38~~ (0.00)	γ: −14.00 (fixed)	$C^{\beta}H^{*}$-$C^{\gamma}H^{H}$: ~~7.71~~ (0.04); $C^{\beta}H^{*}$-$C^{\gamma}H^{L}$: ~~13.29~~ (0.15) $C^{\gamma}H^{H}$-$C^{\delta}H^{*}$: discarded; $C^{\gamma}H^{L}$-$C^{\delta}H^{*}$: 6.80 (0.08) $C^{\delta}H^{*}$-$C^{\epsilon}H^{*}$: 5.06 (0.08); $C^{\epsilon}H^{*}$-$N^{\zeta}H^{*}$: 6.45 (0.03)
Asp^{20}	7.67 (0.00)	$\underline{9.07}$ (0.00) 3.90 (0.00)	β: −15.49 (0.00)	
Glu^{21}	5.47 (0.01)	$\underline{6.26}$ (0.07) 8.30 (0.06)	β: −12.45 (0.14) γ: −17.54 (0.02)	$C^{\beta}H^{H}$-$C^{\gamma}H^{H}$: ~~$\underline{4.41}$~~ (0.04); $C^{\beta}H^{H}$-$C^{\gamma}H^{L}$: ~~$\underline{9.27}$~~ (0.44) $C^{\beta}H^{L}$-$C^{\gamma}H^{H}$: 12.13 (0.05); $C^{\beta}H^{L}$-$C^{\gamma}H^{L}$: 5.31 (0.08)
Cys^{22}	8.05 (0.00)	$\underline{3.27}$ (0.00) 10.84 (0.00)	β: −13.56 (0.00)	

TABLE 6 (continued)

Amino acid	NH-$C^{\alpha}H$	$C^{\alpha}H$-$C^{\beta}H$	Two-bond	Others
Pro^{23}		$\underline{7.55}$ (0.03)	β: −13.68 (0.03)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{7.51}$ (0.03); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 5.46 (0.02)
		7.82 (0.02)	δ: ~~−11.03~~ (0.01)	$C^{\gamma}H^{*}$-$C^{\delta}H^{H}$: 7.33 (0.02); $C^{\gamma}H^{*}$-$C^{\delta}H^{L}$: 5.78 (0.07)
Pro^{24}		6.84 (0.00)	γ: −12.59 (0.01)	$C^{\beta}H^{*}$-$C^{\gamma}H^{H}$: 5.73 (0.00); $C^{\beta}H^{*}$-$C^{\gamma}H^{L}$: 5.15 (0.01)
			δ: ~~−10.43~~ (0.00)	$C^{\gamma}H^{H}$-$C^{\delta}H^{H}$: 6.57 (0.01); $C^{\gamma}H^{H}$-$C^{\delta}H^{L}$: 6.73 (0.00)
				$C^{\gamma}H^{L}$-$C^{\delta}H^{H}$: 5.30 (0.02); $C^{\gamma}H^{L}$-$C^{\delta}H^{L}$: 7.54 (0.00)
Gly^{25}	4.27 (0.00)		α: −17.00 (0.00)	
	5.57 (0.00)			
Gln^{26}	~~10.08~~ (0.11)	~~8.68~~ (1.59)	β: −14.00 (fixed)	$C^{\beta}H^{H}$-$C^{\gamma}H^{H}$: ~~12.66~~ (6.76); $C^{\beta}H^{H}$-$C^{\gamma}H^{L}$: ~~5.69~~ (1.00)
		~~4.41~~ (2.25)	γ: ~~−13.94~~ (6.04)	$C^{\beta}H^{L}$-$C^{\gamma}H^{H}$: ~~4.77~~ (2.26); $C^{\beta}H^{L}$-$C^{\gamma}H^{L}$: ~~4.72~~ (0.66)
Cys^{27}	6.61 (0.00)	11.59 (0.00)	β: −14.12 (0.00)	
		$\underline{3.51}$ (0.00)		
Arg^{28}	8.96 (0.00)	$\underline{10.82}$ (0.00)	β: −13.21 (0.02)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: $\underline{3.87}$ (0.14); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 6.92 (0.00)
		4.49 (0.02)		$C^{\gamma}H^{*}$_$C^{\delta}H^{*}$: 8.34 (0.00); $C^{\delta}H^{*}$-$N^{\varepsilon}H$: 5.97 (0.00)
Phe^{29}	~~1.71~~ (0.08)	discarded		$C^{\delta}H^{*}$-$C^{\varepsilon}H^{*}$: 8.17 (1.54); $C^{\varepsilon}H^{*}$-$C^{\zeta}H$: 6.61 (1.09)
Pro^{30}		3.36 (0.00)	β: −11.98 (0.02)	$C^{\beta}H^{H}$-$C^{\gamma}H^{H}$: 6.60 (0.09); $C^{\beta}H^{H}$-$C^{\gamma}H^{L}$: 3.62 (0.02)
		$\underline{10.69}$ (0.01)	γ: −14.00 (fixed)	$C^{\beta}H^{L}$-$C^{\gamma}H^{H}$: $\underline{9.59}$ (0.07); $C^{\beta}H^{L}$-$C^{\gamma}H^{L}$: $\underline{8.64}$ (0.01)
				$C^{\gamma}H^{H}$-$C^{\delta}H^{*}$: 6.74 (0.05); $C^{\gamma}H^{L}$-$C^{\delta}H^{*}$: 6.06 (0.01)
Arg^{31}	7.88 (0.01)	8.36 (0.04)	β: −14.00 (fixed)	$C^{\beta}H^{H}$-$C^{\gamma}H^{*}$: 7.56 (0.53); $C^{\beta}H^{L}$-$C^{\gamma}H^{*}$: 7.33 (0.01)
		4.46 (0.03)		$C^{\gamma}H^{*}$-$C^{\delta}H^{*}$: 6.48 (0.00); $C^{\delta}H^{*}$-$N^{\varepsilon}H$: 6.18 (0.00)
Gly^{32}	4.88 (0.00)		α: −16.44 (0.00)	
	5.23 (0.00)			
Asp^{33}	~~6.12~~ (0.00)	5.70 (0.00)	β: −15.44 (0.00)	
		4.07 (0.00)		
Ala^{34}	~~4.49~~ (0.00)	7.72 (0.00)		
Asp^{35}	5.89 (0.00)	$\underline{10.58}$ (0.00)	β: −16.82 (0.00)	
		3.47 (0.00)		
Pro^{36}		discarded	δ: ~~−10.12~~ (0.00)	$C^{\beta}H^{*}$-$C^{\gamma}H^{*}$: 5.48 (0.00);
				$C^{\gamma}H^{*}$-$C^{\delta}H^{H}$: 6.26 (0.00); $C^{\gamma}H^{*}$-$C^{\delta}H^{L}$: 6.19 (0.00)
Tyr^{37}	~~6.01~~ (0.01)	2.63 (0.04)		$C^{\delta}H^{*}$-$C^{\varepsilon}H^{*}$: 6.03 (1.70);
Cys^{38}	6.69 (0.00)	$\underline{1.58}$ (0.00)	β: −14.21 (0.00)	
		11.35 (0.00)		
Glu^{39}	7.85 (0.00)	7.05 (0.00)		$C^{\beta}H^{*}$-$C^{\gamma}H^{*}$: 7.37 (0.00)

[a] For each coupling constant, the average and standard deviation (in parentheses) of its value over the six runs Dec(3, B; 0) through Dec(3, B; 5) are given. The first column specifies the residue type and sequence number. The second column contains the amide-alpha coupling constants, with those involving the high-field H^{α} above the low-field H^{α} in the case of nondegenerate glycine residues. The third column contains the alpha-beta coupling constants, with those involving the high-field $C^{\beta}H^{H}$ above the low-field $C^{\beta}H^{L}$; in those residues for which the β-methylene protons have been stereospecifically assigned, the coupling constant involving the $C^{\beta}H^{2}$ proton in the IUPAC nomenclature has been underlined. The fourth column contains the two-bond couplings between methylene protons at the side-chain positions indicated by Greek letters. The final column contains a semicolon-separated list of the pair of protons and values of all the other coupling constants in each amino acid, using the same conventions as in the previous columns, with an asterisk indicating degenerate protons. For each coupling constant, only those values greater than 1.0 Hz and not associated with line widths greater than 30 or less than 3 Hz were used to compute its average, as described in the text. Those coupling constants that did not pass this screen in any of the six runs are marked 'discarded'; those two-bond couplings whose values were held at −14 Hz during these calculations are marked '(fixed)'. Those average coupling constants whose accuracy is in doubt, either because they lie well outside the usual ranges, gave substantially different values in the six runs, or connected resonances that heavily overlapped the water signal, the diagonal peaks or other cross peaks, are indicated by having a horizontal line drawn through their values.

(Fig. 2c) that appears as a very broad and elongated tetraplet in the best-fit spectrum (Fig. 2d) is due to the $C^{\alpha}H$-$C^{\beta}H^{H}$ coupling constant in Cys[7], which was determined to be nearly 12 Hz in the best-fit spectrum, with $C^{\alpha}H$ and $C^{\beta}H^{H}$ line widths of 64 and 9 Hz, respectively. This constitutes an example of the 'line width versus coupling constant' local minimum discussed in the text (see

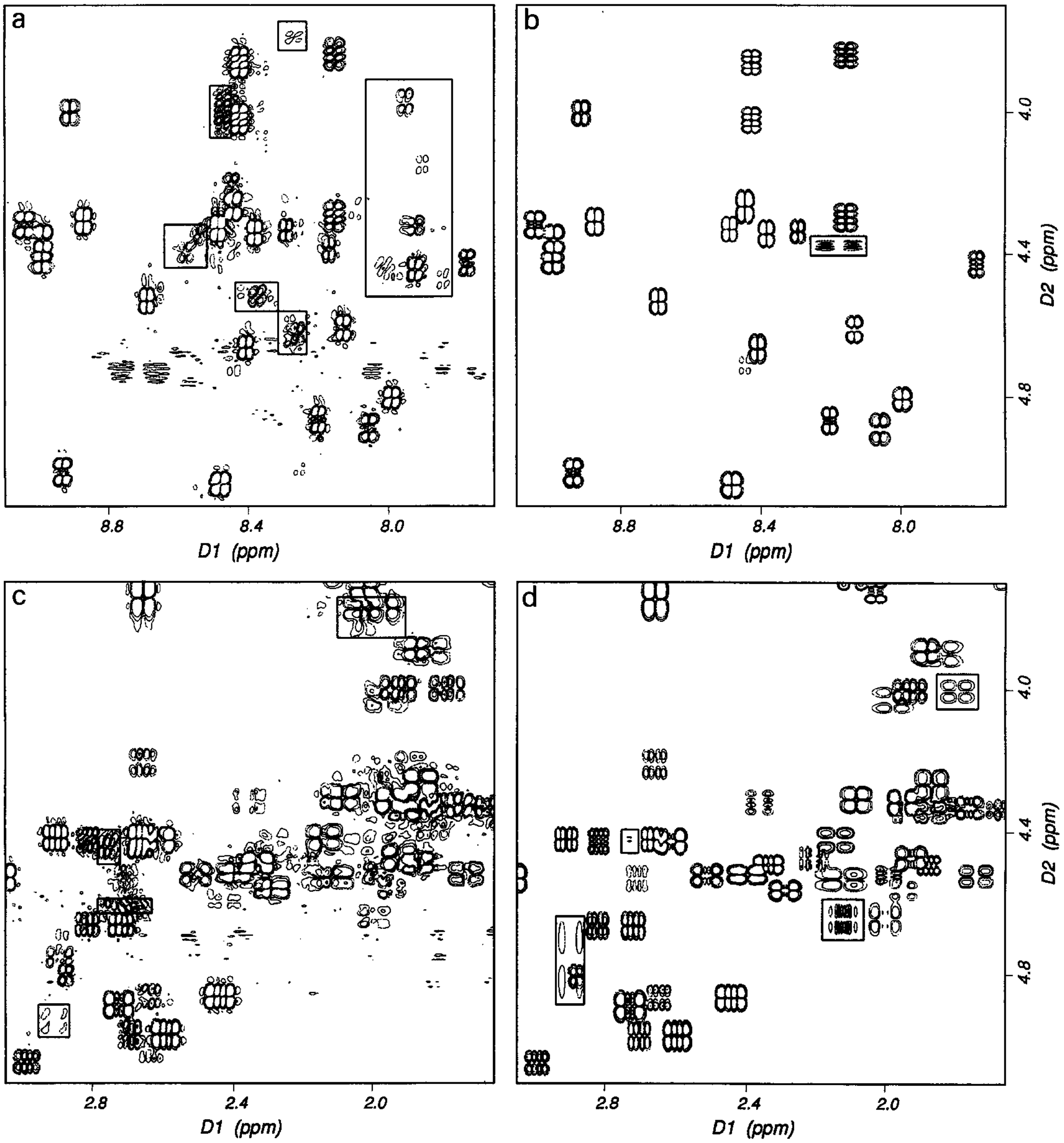

Fig. 2. Selected regions of the experimental and best-fit decorsin spectrum, denoted by Dec(3, B; 0) in the text. (a) A portion of the above-diagonal fingerprint region; (b) the same portion of the Dec(3, B; 0) best-fit spectrum; (c) a portion of the below-diagonal α-aliphatic region; (d) the same portion of the Dec(3, B; 0) best-fit spectrum. The cross peaks with boxes drawn around them that are present in the experimental spectra but not in the simulated spectra are due to the presence of impurities in the sample. The cross peaks with boxes drawn around them in the best-fit spectra are those whose basic shapes or sign patterns were not correctly reproduced (see text).

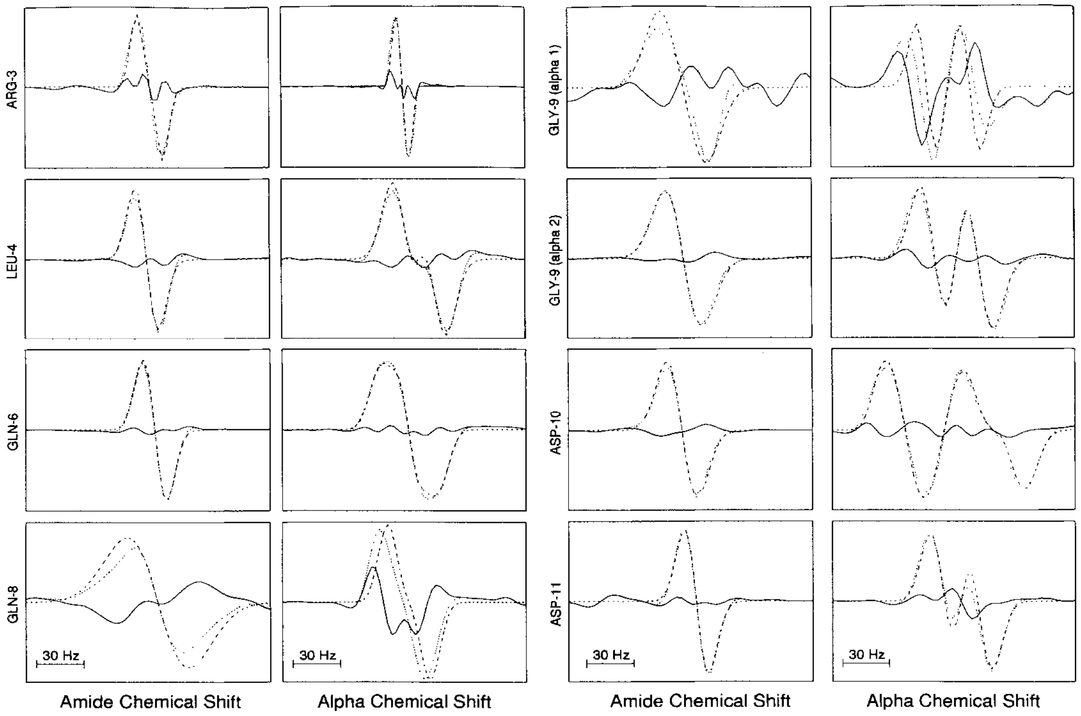

Fig. 3. Plots of cross sections along the amide (left) and alpha (right) chemical shift axes through the first eight nonterminal amide-alpha cross peaks in the experimental (dotted lines) and best-fit spectrum Dec(3, B; 0) (dashed lines), together with the difference between the two (solid lines). The scale of the chemical shift (horizontal) axis is the same for all these plots, while the intensity (vertical) scale is arbitrary and different for each.

also Yang and Havel (1994)), which succeeded in trapping all but two of the six runs Dec(3, B; 0–5). The boxed cross peak at ca. D_1 = 2.70 and D_2 = 4.40 ppm in Fig. 2c, that is also boxed and nearly missing in Fig. 2d, is due to the $C^{\alpha}H$-$C^{\beta}H^{L}$ coupling in Gln^{26}, which came out as 3.4 Hz in the simulated spectrum, with $C^{\alpha}H$ and $C^{\beta}H^{L}$ line widths of 8 and 25 Hz, respectively. It is heavily overlapped with neighboring decorsin cross peaks as well as (apparently) an impurity, and hence could not be determined reliably (the standard deviation reported in Table 6 is 2.25 Hz). The complex boxed cross peak at ca. D_1 = 2.10 and D_2 = 4.65 ppm in the best-fit spectrum (Fig. 2d) is from the $C^{\alpha}H$-$C^{\beta}H^{L}$ coupling in Glu^{13} of 3.83 Hz, and is completely missing from the experimental spectrum (Fig. 2c), probably because it is so close to the water signal. Finally, the octaplet at ca. D_1 = 1.80 and D_2 = 4.00 ppm in the experimental spectrum (Fig. 2c) is due to the $C^{\alpha}H$-$C^{\beta}H^{H}$ coupling in Glu^{21}, which gave rise to a tetraplet in the best-fit spectrum (Fig. 2d) with a coupling constant of 6.19 Hz. In this case the problem is probably due to strong coupling and diagonal overlap among the beta-gamma cross peaks, whose coupling constants passively split the $C^{\alpha}H$-$C^{\beta}H^{H}$ cross peak.

The fact that the fit was usually also quantitatively quite good is demonstrated in Fig. 3, which shows plots of the cross sections and their differences between rows and columns of the experimental and best-fit spectra through the amide-alpha cross peaks of the first eight nonterminal non-proline residues. The quantitative fit was noticeably less good for the glycine residues,

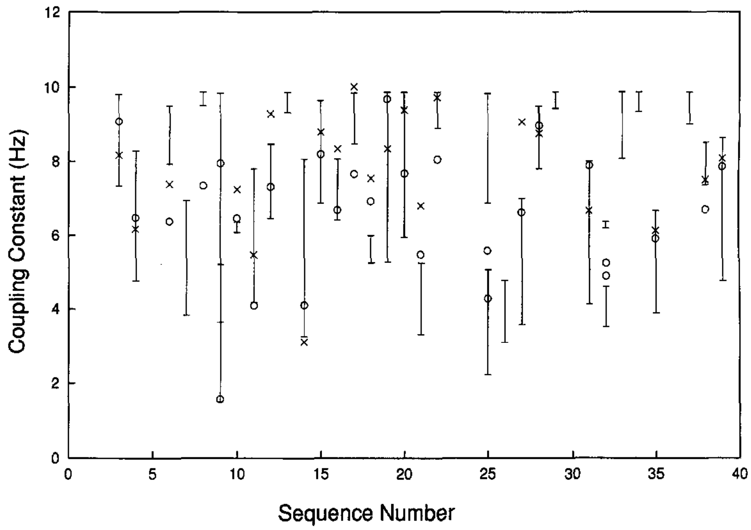

Fig. 4. A plot versus amino acid sequence number of the ranges in the amide-alpha coupling constants (in Hz) obtained on applying the Karplus relation to the distance geometry ensemble, as described in the text. Also shown are the 21 amide-alpha coupling constants that could be measured by applying the method of Kim and Prestegard (1989) to the Dec(1, B) spectra (crosses), as well as the average amide-alpha coupling constants from Table 6 for the 28 cases in which our method did not encounter any obvious problems (circles).

particularly along the α-chemical shift axis of the weaker cross peaks (e.g. the α-1 cross peak of Gly^9 in Fig. 3), possibly because of strong coupling or differential relaxation among the α-protons. The only other significant problems occurred at Gln^8 and Lys^{19}. Figure 3 shows that the former problem is due to a displacement in the α-carbon chemical shift, which in turn was found to be caused by the partial overlap of its $C^{\alpha}H$-$C^{\beta}H^{H}$ cross peak with a cross peak from Pro^2, pushing its apparent center upfield. A similar explanation also applies to Lys^{19}.

Figure 4 shows a plot of the amide-alpha coupling constants versus residue number that were obtained with our method and with the method of Kim and Prestegard (1989). Also shown in the plot are the ranges in the values of the amide-alpha coupling constants that were obtained on applying the Karplus relation to an ensemble of 25 structures, as described above. In five cases the method of Kim and Prestegard yielded clearly incorrect values because the corresponding cross peaks had been greatly distorted or eliminated by removal of the water line; our method also seriously underestimated these coupling constants, which are therefore not shown. In another pair of residues (Gln^{26} and Ala^{34}), the amide-alpha cross peaks overlapped perfectly so that the method of Kim and Prestegard could not be used at all; our largest remaining discrepancies with the ensemble ranges were likewise obtained with these residues. With Gln^8, the method of Kim and Prestegard yielded a value near 12 Hz, while our method yielded a value near 7 Hz and the values observed in the distance geometry ensemble were all between 9 and 10 Hz. In this case the dispersive component of the pseudo-1D spectrum for the cross peak exhibited very flat maxima, whose separation therefore could not be determined to better than ± 8 Hz. With an absorptive separation of 17.4 Hz, this led to an even larger uncertainty in the value of the coupling constant obtained by the method of Kim and Prestegard, which therefore cannot be regarded as significant and is also not shown in the figure. Finally, the method of Kim and Prestegard cannot be applied to the three glycines in decorsin.

Among the remaining 21 amide-alpha couplings for which reasonably reliable measurements could be obtained from the method of Kim and Prestegard, nine differed by 1.0 Hz or more from those obtained with our least-squares method, while four differed by less than 0.5 Hz. In only 6 of the 21 cases did the coupling constants calculated by our method exceed those obtained from the method of Kim and Prestegard. There were three cases in which our method produced values that disagreed with both the distance geometry ensemble and the method of Kim and Prestegard (Cys^{17}, Cys^{22} and Cys^{38}), and in all of these cases the discrepancies with the ensemble were less than 1.0 Hz. There were also three cases in which our method produced values which agreed with the ensemble, while those obtained from the method of Kim and Prestegard significantly disagreed (Asp^{10}, Gln^{12} and Cys^{27}). Finally, there were two cases in which the values obtained by both methods disagreed significantly with the values observed in the distance geometry ensemble, as well as with each other (Gln^{6} and Asn^{18}). The average difference between the 21 coupling constants calculated by our method and that of Kim and Prestegard was 1.11 Hz.

CONCLUSIONS

We have shown that least-squares fits to DQF-COSY spectra can yield reasonable estimates for most of the proton–proton coupling constants in peptides and small proteins. The exceptions are largely confined to those coupling constants whose active cross peaks strongly overlap either the diagonal of the spectrum, the water signal, or other cross peaks. In some cases, these unreliable coupling constants can be identified by the fact that their values depend strongly on the starting values used for the optimization, as do their exact chemical shifts and line widths. The comparisons made with those coupling constants that could be determined independently make us confident that all the coupling constants in VDA were correctly determined to within 0.5 Hz, with the possible exception of the $\mathrm{C}^{\delta}\mathrm{H}$-$\mathrm{N}^{\epsilon}\mathrm{H}$ coupling constant in the tryptophan and those involving the nearly degenerate γ- and δ-protons of the lysine residue. The accuracy of the coupling constants obtained for decorsin is less certain, but the available controls lead us to believe that at least two thirds have an accuracy of better than 1.0 Hz.

Overall, the least-squares method used here appears to be at least as reliable as other methods based on homonuclear DQF-COSY experiments when used with the same data, and exhibits the following additional advantages:

(1) Its ability to handle complicated cross peaks is limited only by the resolution of the spectra, and hence it can be applied to a far greater number of coupling constants.

(2) It performs these coupling constant determinations on all of the cross peaks together, largely automating the task.

(3) It utilizes the information contained in those cross peaks wherein the coupling constant plays only a passive role, which under favorable circumstances enables it to estimate coupling constants whose active cross peaks cannot be observed.

(4) It utilizes all the points of the cross peaks together, rather than just their largest and/or smallest points, which, providing that the correct line-shape model is used, should improve accuracy of the coupling constants, particularly when the peaks are broad or partially overlapped.

Even though considerably higher quality DQF-COSY spectra are needed to measure accurate coupling constants than those that are routinely collected for assignment purposes, the advent of pulsed gradient methods should make such spectra much easier to obtain than they have been in the past (see e.g. Davis et al., 1991). Work is also proceeding on extending the least-squares method described here to spectra with fewer overlapping cross peaks, e.g. carbon/nitrogen-dispersed and/or double-quantum spectra. Complete coupling constant information promises to be a valuable complement to NOE data for the purposes of structure determination (Wüthrich, 1986; Havel, 1991; Wagner et al., 1992). It could, for example, be used to refine structures obtained from distance geometry calculations, so that they also agree with the coupling constants. Before this is done, however, the coupling constant information should be employed as an independent check on the results of distance geometry calculations performed using the NOE data alone.

ACKNOWLEDGEMENTS

We thank Jeff Hoch and Alan Stern for useful discussions, and Dr. M. Kurz and Prof. Dr. H. Kessler for the VDA cyclic peptide sample. This work was supported by NIH Grants GM-38221 and GM-47467, and by a grant from BIOSYM Technologies, Inc. The computational facilities used were paid for in part by NIH Grant RR0-7452, and by a grant from the W.M. Keck Foundation.

REFERENCES

Bodenhausen, G., Freeman, R., Niedermeyer, R. and Turner, D.L. (1977) *J. Magn. Reson.*, **26**, 133–164.

Davis, A.L., Laue, E.D., Keeler, J., Moskau, D. and Lohman, J. (1991) *J. Magn. Reson.*, **94**, 637–644.

Derome, A.E. and Williamson, M.P. (1990) *J. Magn. Reson.*, **88**, 177–185.

De Marco, A., Llinás, M. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2727–2742.

Havel, T.F. (1991) *Prog. Biophys. Mol. Biol.*, **56**, 43–78.

Kessler, H., Müller, A. and Oschkinat, H. (1985) *Magn. Reson. Chem.*, **23**, 844–852.

Kessler, H., Haupt, A., Schudok, M., Ziegler, K. and Frimmer, M. (1988) *Int. J. Pept. Protein Res.*, **32**, 183–193.

Kim, Y. and Prestegard, J.H. (1989) *J. Magn. Reson.*, **84**, 9–13.

Krezel, A.M., Wagner, G., Symour-Ulmer, J. and Lazarus, R.A. (1994) *Science*, **264**, 1944–1947.

Pearson, G.A. (1987) *J. Magn. Reson.*, **74**, 541–545.

Piantini, U., Sørensen, O. and Ernst, R.R. (1982) *J. Am. Chem. Soc.*, **104**, 6800–6801.

Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Seymour, J.L., Henzel, W.J., Nevins, B., Stults, J.T. and Lazarus, R.A. (1990) *J. Biol. Chem.*, **265**, 10143–10147.

Shaka, A.J. and Freeman, R. (1983) *J. Magn. Reson.*, **51**, 169–173.

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 7772–7777.

Wagner, G., Hyberts, S. and Havel, T.F. (1992) *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 167–198.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Yang, J. and Havel, T.F. (1994) *J. Biomol. NMR*, **4**, 807–826.