Journal of Magnetic Resonance **148,** 35–46 (2001)
doi:10.1006/jmre.2000.2220, available online at http://www.idealibrary.com on IDEAL®

Automatic Assignment of NOESY Cross Peaks and Determination of the Protein Structure of a New World Scorpion Neurotoxin Using NOAH/DIAMOD

Yuan Xu,* Michael J. Jablonsky,†,[1] Patricia L. Jackson,†,[1] Werner Braun,*,[2] and N. Rama Krishna†

**Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, Texas 77555-1157; and †NMR Core Facility, Comprehensive Cancer Center and Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294-2041*

Received April 6, 2000; revised August 24, 2000

The 3D NMR structures of the scorpion neurotoxin, CsE-v5, were determined from the same NOESY spectra with NOAH/DIAMOD, an automated assignment and 3D structure calculation software package, and with a conventional manual assignment combined with a distance geometry/simulated annealing (X-PLOR) refinement method. The NOESY assignments and the 3D structures obtained from the two independent methods were compared in detail. The NOAH/DIAMOD program suite uses feedback filtering and self-correcting distance geometry methods to automatically assign NOESY spectra and to calculate the 3D structure of a protein. NOESY cross peaks were automatically picked using a standard software package and combined with 74 manually assigned NOESY peaks to start the NOAH/DIAMOD calculations. After 63 NOAH/DIAMOD cycles, using REDAC procedures in the last 8 cycles, and final FANTOM constrained energy minimization, a bundle of 20 structures with the smallest target functions has a RMSD of 0.81 Å for backbone atoms and 1.11 Å for all heavy atoms to the mean structure. Despite some missing chemical shifts of side chain protons, 776 (including 74 manually assigned) of 1130 NOE peaks were unambiguously assigned, 150 peaks have more than one possible assignment compatible with the bundle structures, and only 30 peaks could not be assigned within the given chemical shift tolerance ranges in either the D1 or the D2 dimension. The remaining 174, mainly weak NOE peaks were not compatible with the final 20 best bundle structures at the last NOAH/DIAMOD cycle. The automatically determined structures agree well with the structures determined independently using the conventional method and the same NMR spectra, with the mean RMSD in well-defined regions of 0.84 Å for bb and 1.48 Å for all heavy atoms from residues 2–5, 18–26, 32–36, and 39–45. This study demonstrates the potential of the NOAH/DIAMOD program suite to automatically assign NMR data for proteins and determine their structure.

Key Words: **automated NMR spectra assignment; self-correcting distance geometry; NOAH; neurotoxins; DIAMOD; XPLOR.**

INTRODUCTION

The assignment of cross peaks in NOESY spectra is a crucial step in protein structure determination by NMR. As manual interpretation of NMR spectra is time consuming, tedious, and error-prone, advanced iterative approaches have been suggested to automate the assignment of NOESY peaks and 3D structure calculation (*1–18*). We have developed the NOAH/DIAMOD program suite (*8, 12, 15, 16*) based on feedback filtering and self-correcting distance geometry (SECODG) (*8, 19–21*). In previous tests, the suite automatically assigned simulated and experimental protein NOESY spectra and determined 3D structures from high-quality NOESY peak lists. More than 80% of the NOESY peaks could be automatically assigned within the given chemical shift tolerance and 95–99% of those assigned peaks agreed with assignments made using conventional methods (*8, 12*). We recently used NOAH/DIAMOD to determine the NMR structure of a 46-residue protein, Crambin (Ser22, Ile25), in a completely automatic fashion (*15*). After we completed the work in 1997, the X-ray structure of crambin was published (*22*). Our automatically determined structure agreed well with the X-ray structure in all well-defined regions. In this paper, we report on a detailed comparison of the 3D NMR structures of the CsE-v5 neurotoxin that was determined from the same spectral data with our NOAH/DIAMOD package and with a conventional manual assignment and X-PLOR refinement method in an independent way. The CsE-v5 neurotoxin (6.3 kDa) isolated from the venom of the New World scorpion *Centruroides sculpturatus Ewing,* is an α-neurotoxin that is specific to insect sodium channels (*23*). Its 3D structure has not been previously determined, although it shares some sequence similarity to other neurotoxins such as CsE-v3 and CsE-v1 toxins. However, it exhibits amino acid deletions in the J- and M-loops, a characteristic shared by the Old World scorpion toxins (*24*). Thus its structure is expected to show some significant differences from those of CsE-v3 and CsE-v1.

We want to assess the quality of the automated assignment

[1] Present address: Department of Chemistry, University of Alabama at Birmingham.

[2] To whom correspondence should be addressed.

1090-7807/01 $35.00

procedure and the accuracy of the determined 3D structure in a realistic situation. We used exactly the same time domain NMR spectra and the same manual sequential assignments at the outset of this test. However, we performed spectral data processing, NOESY cross-peak identification, cross-peak volume integration, distance calibration, and structure calculation in a completely independent way at the two laboratories. We therefore believe that the results of this study also bear on the general accuracy of NMR solution structures, as we determined the NMR solution structure by two independent methods from the same NMR data. The automatically determined structures agree within 1 Å in well-defined regions for the backbone fold. As our SECODG-based automatic assignment method requires much less time than conventional methods and generates structures of a similar quality, we believe it will eventually replace manual methods as the first approach to NOE spectral interpretation, such that the experimentalist can concentrate on few critical cases.

MATERIALS AND METHODS

Variant-5 Purification and Sample Preparation

The variant-5 neurotoxin was isolated from scorpion venom with a two-step column chromatography procedure (*24*). First, the crude venom was separated into 12 toxin-containing fractions on an ion-exchange carboxymethylcellullose column with a pH and ammonium acetate gradient. The variant-5 was purified from fractions 7 and 8 using two additional columns eluted with ammonium acetate gradients: CM–Sephadex at pH 6.0 and DEAE–Sephadex A-25 at pH 8.5. The purified toxin was lyophilized three times to remove all ammonium acetate. The NMR sample (250 μl in a Shigemi microcell) was made to a concentration of 1.0 mm and pH 4.0 with 10% D_2O. After completion of the experiments in 90% H_2O, the sample was lyophilized and redissolved in 100% D_2O for identifying slowly exchanging amide protons and for additional NMR measurements.

Experiments

The NMR measurements were performed on a Bruker AM-600 spectrometer equipped with an Aspect 3000 computer. Data were collected at 293 and 303 K. NOESY, DQF-COSY, and TOCSY measurements were performed in pure absorption mode using time proportional phase increment and presaturation for water suppression. Mixing times of 100 and 200 ms for NOESY in H_2O and 200 ms in D_2O were employed. In addition, a jump–return 200-ms NOESY in H_2O was collected. The TOCSY experiment was performed with a mixing time of 70 ms in both H_2O and D_2O. For the NOESY and TOCSY experiments, 128 scans of 2 K complex data points were collected for each of the 512 serial files. For the COSY experiments, the data size was 4 K complex with 1024 *t*1 points. COSY data were zero filled to 8 K. Data were processed with FELIX using various window functions.

Spin System and Sequence Specific Assignments

The spin system assignments were made using standard procedures for nonlabeled proteins (*25*). Using TOCSY and NOESY data at 293 and 303 K, all protons were uniquely (but not necessarily stereospecifically) assigned except for cases of overlap of similar residues, e.g., the protons for Lys9, Lys12, and Lys50. The ϕ angle constraints ($-60 \pm 40°$ for $J < 7$ Hz and $-120 \pm 40°$ for $J > 8$ Hz) were obtained from the DQF-COSY in H_2O. The stereo-specific assignments for the AMX spin systems and the valines were made from coupling constants and the 100-ms NOESY.

Data Processing for NOAH/DIAMOD Structure Determination

All NOESY spectra were processed using Felix95's "E-Z 2D Transform" protocol. A 2 K × 2 K point matrix was rephased several times with baseline correction at the final stage. The jump–return NOESY was processed such that either the aliphatic region or the amide region shows positive intensities. The Felix automatic peak-picking routine was used to pick all peaks at a contour level of 0.035 for the 200-ms NOESY spectrum with water presaturation. The filter function "remove diagonal peaks" with a tolerance of 0.02 ppm was first used to remove diagonal peaks. Then the filter function "symmetrize spectrum" with tolerance of 0.02 ppm was applied to remove unsymmetrical, water, and other artifact peaks in the presaturation NOESY spectrum. Volume integration and optimization with the Lorentzian lineshapes algorithm were used to obtain cross-peak intensities. After applying these filter functions, 2674 NOE cross peaks were saved in "Felix peak file" format. The corresponding 2674 peak intensities were optimized via Felix peak volume optimization routines, "optimize peak centers," "optimize peak widths," and "optimize peak volumes," in consecutive order and saved as a peak volume file. In all these processes we did not manually edit any peak volumes. An in-house FORTRAN90 program was used to find symmetric cross-peak pairs at a tolerance of 0.02 ppm for both the D1 and the D2 dimensions and to calculate the average peak volumes. This program removed peaks with negative volumes and saved 1080 symmetric NOE peaks in an output peak file with corresponding chemical shifts in a format suitable for input to the NOAH program.

In addition, 6 manually assigned sequential peaks were added to the list. These peaks were present only on one side of the diagonal of the water presaturation spectrum and had been eliminated by the filter function "symmetrize spectrum." Another 24 unassigned peaks manually picked from the D_2O spectrum and 20 unassigned peaks picked from a jump–return spectrum were also added to obtain a total of 1130 NOE peaks for use in the automatic assignment and structure calculations.

All those manually picked peaks were not seen in water-presaturated NOESY spectrum because of interference of the strong water line but could be identified in D_2O and jump–return spectra. Although we manually picked 44 extra peaks, they were not manually assigned.

Additional Constraints Used for the NOAH/DIAMOD Calculation

A total of 74 manually assigned sequential (backbone) NOE peaks (6.5% of 1130 and provided by the UAB group) were kept fixed throughout the NOAH/DIAMOD cycles. The chemical shifts of 320 protons were manually derived from COSY/TOCSY spectra. We obtained 27 ϕ and 27 χ_1 angular constraints from the corresponding experimental coupling constants, 24 stereo-specific assignments for β protons, and 8 constraints for 4 disulfide bridges. In addition to these experimental data, very loose dihedral angle constraints were used to restrict the range of the backbone and χ_1 angles of the 20 individual amino acid types to those found in a statistical analysis of these angles in proteins (*26*). We used 85 such angular constraints. The ranges of angular constraint from statistics were quite different for a given residue type and were listed in the table of that reference. For instance, for the lower and upper limits of ϕ angular constraint of residue alanine we used a range from -150 to 80°, which covers more than 98% of the 2232 alanine residues studied in different proteins in the reference. We have used such angular constraints in our several studies and it did improve the convergence and reduce the optimization times (*15, 27, 28*). In this study we have not used any hydrogen bond constraints for our automatic assignment and structure calculation. Pseudo-atoms were used for the assignments of $CH_2\beta$ or $CH_2\gamma$ methylene protons and aromatic ring protons in case of overlapping chemical shifts of the protons.

NOAH/DIAMOD Assignments and Structure Calculations

The NOAH/DIAMOD structure calculations were run as previously described (*15*). The NOAH/DIAMOD package uses as input the NOE cross-peak list with the cross-peak intensities and chemical shifts of the NOE intensities. Cross-peak intensities were converted to upper distance constraints by the usual relationship: $I_{i,j} = Ar_{i,j}^{-6}$. An upper distance limit of $r_{i,j} = 2.2$ Å was assigned to the strongest NOESY cross-peak intensity $I_{i,j}$ to determine the constant A. The rest of the upper distance limits were then calculated by the inverse sixth power law; the maximum NOE distance is taken as 6.0 Å for non-pseudo proton pairs. The van der Waals distance was used as lower distance limits. The manual, unambiguous, and ambiguous upper distance constraints were given relative weights of 10, 5, and 1 in the DIAMOD target function (*15*). Angular constraints obtained from coupling constants were weighted five times as those from statistically derived angular constraints. The high weight factor of 10 for the 74 manually assigned distance constraints was especially needed at the initial NOAH/DIAMOD cycles with a high number of ambiguous constraints. For the same reason the chemical shift tolerance was set, at the initial cycles, rather restrictive to 0.015 ppm, and was then gradually increased to 0.03 ppm toward later NOAH/DIAMOD cycles. In every NOAH/DIAMOD cycle, the 10 structures (of 50 total) with the lowest DIAMOD target functions were used in NOAH for the NOESY peak assignments. In the last 4 cycles, the 20 best structures were used. The REDAC procedure (*29*) was applied toward the final cycles to increase the convergence of the bundle of structures. Floating assignments (*1*) were used for diastereotopic methylene protons that were not stereospecifically assigned manually. The CPU time for 63 NOAH/DIAMOD cycles was about 140 h on a SGI workstation with a R10K CPU.

Manual Assignment and X-PLOR Structure Calculation

The solution structures were generated using a hybrid protocol involving distance geometry (DG) and dynamical simulated annealing (SA) (*30, 31*) with minor modifications (*32*) using the X-PLOR/QUANTA/CHARMM modeling package (Molecular Simulations Inc., San Diego, CA) on an Indigo2 Workstation (Silicon Graphics Inc.). The distance geometry substructures were first generated using only medium and long-range (residues i, $i + 2$ or greater) distance and torsion angle constraints. Unlike the NOAH/DIAMOD procedure, the manual assignment/X-PLOR refinement procedure *did not* use the angular constraints from statistical distribution data. Two hundred distance geometry structures were generated and 68 structures with no distance constraint violations larger than 0.3 Å and no dihedral angle violations larger than 5° were selected for simulated annealing calculations. Dynamic simulated annealing calculations followed by energy minimization were then performed using a four-step protocol (*32*). Only the resulting structures consistent with the same criteria for the residual constraints as mentioned above were selected for final analysis and designated the simulated annealing ensemble. An average structure was calculated from these selected structures and then subjected to a restrained energy minimization to remove bond angle and bond length distortions. This was followed by an unrestrained energy minimization to obtain the final energy minimized average structure ⟨SA⟩. This structure had an RMSD value of 1.09 Å for backbone atoms and 1.71 Å for all atoms, with respect to the simulated annealing ensemble.

RESULTS

Convergence and NOE Peak Assignments of the NOAH/DIAMOD Calculations

At the end of 63 iterative cycles, NOAH assigned 702 NOE peaks unambiguously and 150 ambiguously (i.e., each peak has two possible assignments, neither of which violated the bundle

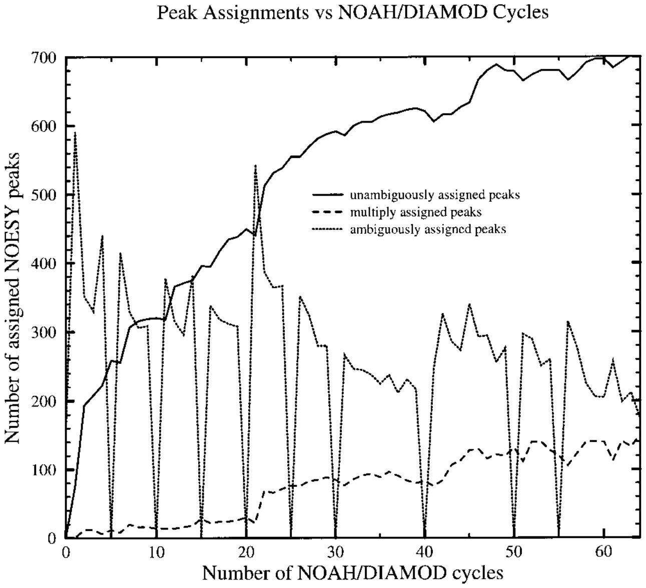

FIG. 1. Plot of NOE peak assignments against NOAH/DIAMOD cycles. The solid line shows unambiguously assigned NOE peaks, the dotted line shows the number of ambiguous NOE peak assignments, and the dashed line represents the number of NOE peaks that have been test assigned.

structure according to the NOAH criteria). Overall 776 cross peaks were finally unambiguously assigned, including the 74 manually assigned peaks used at the initial phase and kept fixed during the calculations. The final unambiguously assigned cross peaks represent 69% of the 1130 peaks in the NOE cross-peak list. From the remaining cross peaks, 174 peaks could not be assigned due to incompatibility with the final bundle of structures and 30 NOE cross peaks were outside any possible combinations of chemical shifts in the proton list within the given chemical shift tolerance of ±0.03 ppm. The latter peaks were most likely arising from side chain protons whose chemical shifts were not available from manual resonance assignment. The number of peaks assigned by NOAH as a function of the NOAH/DIAMOD cycles is shown in Fig. 1. The number of unambiguously assigned peaks rapidly increases to 600 within 30 cycles and then levels off at around 45 to 60 cycles. The spread in the ensemble of structures as measured by the average of the pairwise distance root mean square deviations (DRMSD) within the bundle of the 10 best structures is shown in Fig. 2. There is a significant reduction in the precision of structures seen in the last 20 cycles. A profound effect on precision can be observed from the REDAC procedure (*29*), applied during cycle 56 and beyond. The average RMSDs of the final 20 best structures to the mean are 0.64 ± 0.12 Å for backbone atoms and 0.93 ± 0.11 Å for all heavy atoms before energy minimization. The improvement in the precision of the structures compared to our previous study on Crambin (*15*) are due to the complete assignment of backbone proton chemical shifts and the REDAC procedure (*29*).

Energy Minimization of NOAH/DIAMOD Structures

The final bundle of 20 structures obtained from NOAH/DIAMOD were then energy refined using the FANTOM program (*33–35*). The average RMSD of the bundle structures to their mean increased slightly after the energy minimization from 0.64 to 0.80 ± 0.13 Å for backbone and from 0.93 to 1.11 ± 0.14 Å for all heavy atoms, as we used only the 776 unambiguously assigned distance constraints during the energy minimization. The constrained energy minimization reduced significantly all the individual energy terms, such as hydrogen bond, Lennard–Jones, torsional angle, and disulfide bond energy terms in the initial structures, with an overall decrease for the average total energies from 881 to −270 kcal/mol (Table 1). The violations of the experimental constraints only resulted in a slight increase in the NOE distance and dihedral angles pseudo-energy terms. The unambiguously assigned 776 distance constraints are satisfied in the final 20 minimized structures with no constraint violations >0.5 Å except one constraint. There were no angular constraints violated by >20° in all 20 structures. All the disulfide distance constraints were satisfied to <0.01 Å. The FANTOM refined structures have good stereogeometry and minimal constraint violations.

Secondary and Tertiary Structure of CsE-v5

A superposition of the 20 final structures of CsE-v5 after FANTOM energy minimization is shown in Fig. 3a (stereo view). The overall topology of the NOAH/DIAMOD CsE-v5 structure is similar to other scorpion toxin structures, especially the homolo-

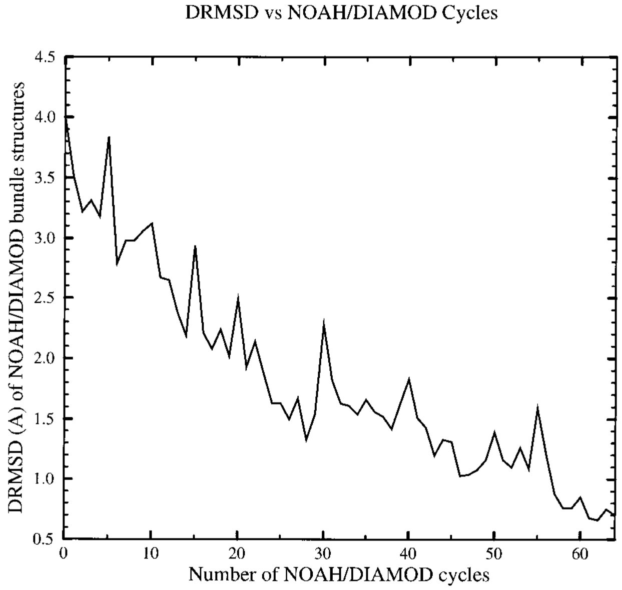

FIG. 2. Plot of distance root mean square deviations (DRMSD) against NOAH/DIAMOD cycles. Although the plot fluctuates vigorously when the number of unambiguous assignments was small at earlier cycles, it converges to a compact set of structures after cycle 55.

TABLE 1

Average Energy Reductions before and after the Energy Minimization Using the FANTOM Program with Unambiguously Assigned 776 Experimental NOE Distance Constraints, Dihedral Angle Constraints, and 8 Disulfide Bridge Constraints

Electrostatic	Hydrogen bonds	Lennard–Jones	Torsion angles	Disulfide bonds	NOE distance	Dihedral angles	Total energies
113	−25	521	154	107	2	9	881
59	−78	−362	80	7	12	12	−270

Note. The force field used is ECEPP/2. The average energies are calculated based on final 20 NOAH/DIAMOD structures with lowest DIAMOD target function values at the 63rd NOAH/DIAMOD cycle. Row 2 lists the ECEPP/2 energies before the FANTOM minimization and row 3 is the corresponding ECEPP/2 energies after the FANTOM minimization.

gous neurotoxin, CsE-V, from the same scorpion (*36*), that were determined via conventional manual NOE peak assignment methods (*32, 36–38*). The backbone conformation of CsE-v5 contains an α-helix (residues 18–26), an antiparallel β-sheet with three strands (residues 1–4, 32–36, and 39–45), a β-bulge between residues 44 and 45, and several loops. The spatial orientation of the α-helix with respect to the β-sheet is stabilized by two disulfide bridges (Cys21–Cys40 and Cys25–Cys42) that connect the α-helix and the second β-strand to form a $\alpha\beta$DB motif (*37*).

Figure 3b shows the conformational change of the average structures before and after the energy refinement (stereo view). The secondary structures are well defined and the backbone and heavy atom RMSDs are, respectively, 0.26 and 0.86 Å for α-helix (18–26); 0.33 and 0.73 Å for β-sheets (32–36 and 39–45), and 0.45 and 0.89 Å for all secondary structures. Overall, the conformational deviations are well within the bundle of structures with a backbone RMSD of the two mean structures of 0.95 and 1.26 Å for all heavy atoms.

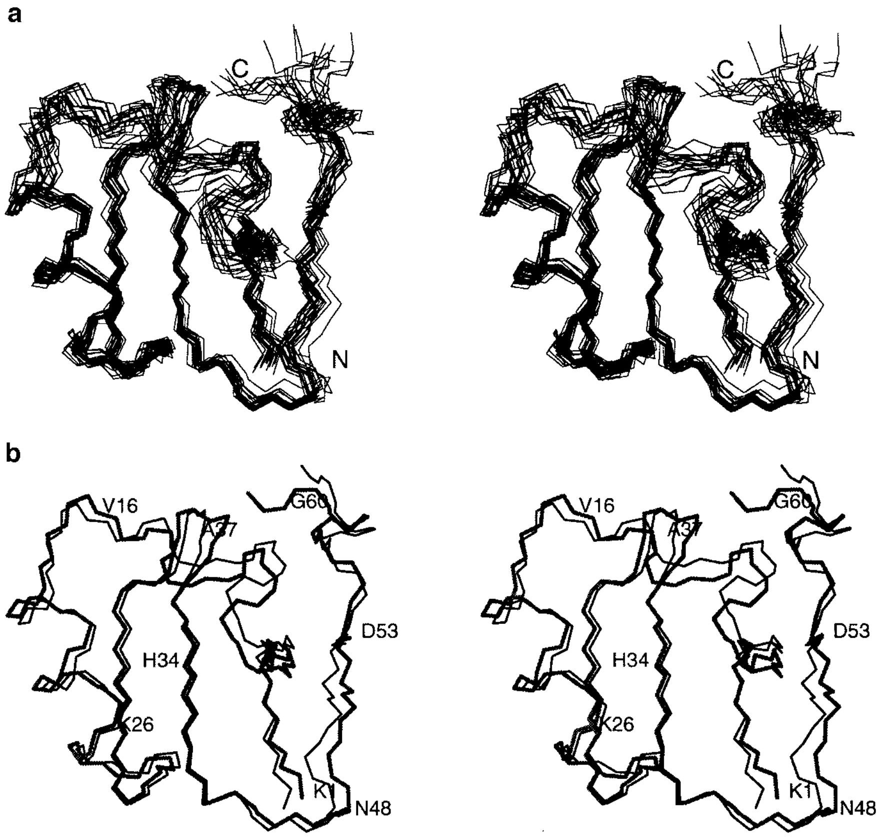

FIG. 3. (a) Superposition of 20 NOAH/DIAMOD bundle structures to their mean after the FANTOM energy minimization using 776 unambiguously assigned NOE distance constraints and dihedral angle constraints (side-by-side stereo view). (b) Superposition of the two mean structures (backbone) before and after energy minimization. The thick line is the mean structure before the minimization.

TABLE 2

Number of Compatible NOE Distance Constraints Assigned Automatically via NOAH/DIAMOD and Assigned Manually

	Intraresidue assignment	Sequential assignment	Medium-range assignment	Long-range assignment	Total number of assignment
Manually assigned	245	212	94	144	695
Auto assigned	332	208	78	158	776
Comparible[a]	331 (100%)[c]	196 (94%)[d]	74 (95%)[d]	145 (92%)[e]	746 (96%)
Compatible[b]	236 (96%)[c]	198 (93%)[d]	94 (100%)[d]	136 (94%)[e]	664 (96%)

Note. Manually assigned: manually assigned number of NOE distance constraints. Auto assigned: automatically assigned number of unambiguous NOE distance constraints by NOAH/DIAMOD programs.

[a] Compatible: number of compatible distance constraints given by NOAH/DIAMOD when the constraints were fitted to 20 final X-PLOR structures with violations less than, respectively, 0.2, 0.5, and 1.0 Å.

[b] Compatible: number of compatible distance constraints given by manual assignment when the constraints were fitted to 20 final NOAH/DIAMOD structures with violations less than, respectively, 0.2, 0.5, and 1.0 Å. The numbers in parentheses are in terms of the percentage (compatible[a or b]/(auto or manual assign)). The last column indicates the total number of distance constraints that can be fitted to either X-PLOR or NOAH/DIAMOD structures.

[c] No constraint violations are larger than 0.2 Å and violated more than 10 structures.

[d] No constraint violations are larger than 0.5 Å and violated more than 10 structures.

[e] No constraint violations are larger than 1.0 Å and violated more than 10 structures.

Comparison of Distance Constraints Derived from the Manual Assignment and NOAH/DIAMOD Procedure

The distance constraints derived in two independent ways, from the manual assignment and from the unambiguous assignments of the NOAH/DIAMOD procedure, were compared in detail (Table 2). We checked the consistency of the constraints derived manually with the 3D structures calculated by NOAH/DIAMOD and vice versa (cross checking). We used cutoffs of 0.2, 0.5, and 1.0 Å to identify consistently occurring (i.e., in more than 10 structures) intraresidual, sequential, and medium- and long-range constraints violations. We used a cutoff distance of 1.0 Å for long-range assignments to account for the larger uncertainty of these constraints, as many of these constraints have pseudo-atom corrections and are derived from relatively weak NOE peaks. Among the 695 manually derived distance constraints, 664 constraints are compatible with the NOAH/DIAMOD structures in this cross check. The compatible constraints represent a large majority (96%) of the total constraints. We obtained the same ratio (746 versus 776) in the comparison of the NOAH/DIAMOD constraints to the final X-PLOR structures. Compatibility among the intraresidual constraints is highest and for long-range constraints is lowest, but the compatibility in all four classes of constraints is above 90%.

Table 3 lists individually all severe differences in the distance constraint lists of both procedures as found by our cross-examine procedure. We analyzed the NOESY spectra and cross-peak lists used in both procedures to elucidate the reason for the major differences in the constraints set. The differences are in most cases not related to the assignment problem per se. For example, peak 198 at (1.97 ppm, 8.53 ppm) is a very weak peak, automatically picked by FELIX and finally unambiguously assigned to Pro5_Hβ_2-Asp7_HN by NOAH. However, the same cross peak was not found in the NOESY spectrum used for the manual assignment. The difference is therefore due to different parameter settings used for the two independent NOESY spectral processings. Peak 865 (2.93 ppm, 1.98 ppm) was assigned as an interresidue contact of Lys12_QE-Ile58_Hβ_2 by NOAH/DIAMOD and was manually assigned as an intraresidue contact to Lys9_QE-Hβ_2. This intraresidue contact was assigned by NOAH/DIAMOD to another cross peak (847) nearby, located at 2.98 and 1.93 ppm.

TABLE 3

The Largest Distance Violations of Either the Automatically Assigned or the Manually Assigned Distance Constraints When Fitted against X-PLOR or NOAH/DIAMOD Structures

	NOE peak	Chem. shifts	UPL	Viol. ranges (Å)
NOAH/DIAMOD assignment				
Pro5_Hβ_2-Asp7_HN	198	1.97, 8.53	5.70	1.81–2.07
Lys12_QE-Ile58_Hβ_2	865	2.93, 1.98	4.99	2.35–6.97
Lys28_HN-Lys29_Hβ_2	476	7.22, 2.08	3.83	1.06–2.22
Manual assignment				
Tyr4_Hβ_3-Ser52_Hβ_2		3.07, 3.91	3.50	1.32–4.73
Asp7_Hβ_3-Lys9_HN		2.68, 8.15	3.50	1.85–2.95
Lys12_HN-Asn57_HD21		8.44, 6.98	6.10	1.57–5.27
Lys12_HN-Asn57_HD22		8.44, 7.60	6.10	1.60–6.08
Ser31_Hα-Gly33_HN		4.36, 8.62	5.00	2.09–2.14
Ser31_HN-Cys42_Hβ_2		8.50, 2.82	5.00	1.21–1.95

Note. Rows 3 to 5 shows the distance constraints assigned by the NOAH/DIAMOD program that violated more than 1.0 Å and all 20 X-PLOR structures. The section entries under "Manual assignment" show the distance constraints assigned manually that violated more than 1.0 Å and all 20 NOAH/DIAMOD structures. Column 1: assignment, column 2: peak number used in automatic assignment, column 3: the chemical shifts in ppm of corresponding protons in column 1, column 4: the upper distance limit of the constraints; and column 5: the range of the distance violations from smallest to largest among 20 structures.

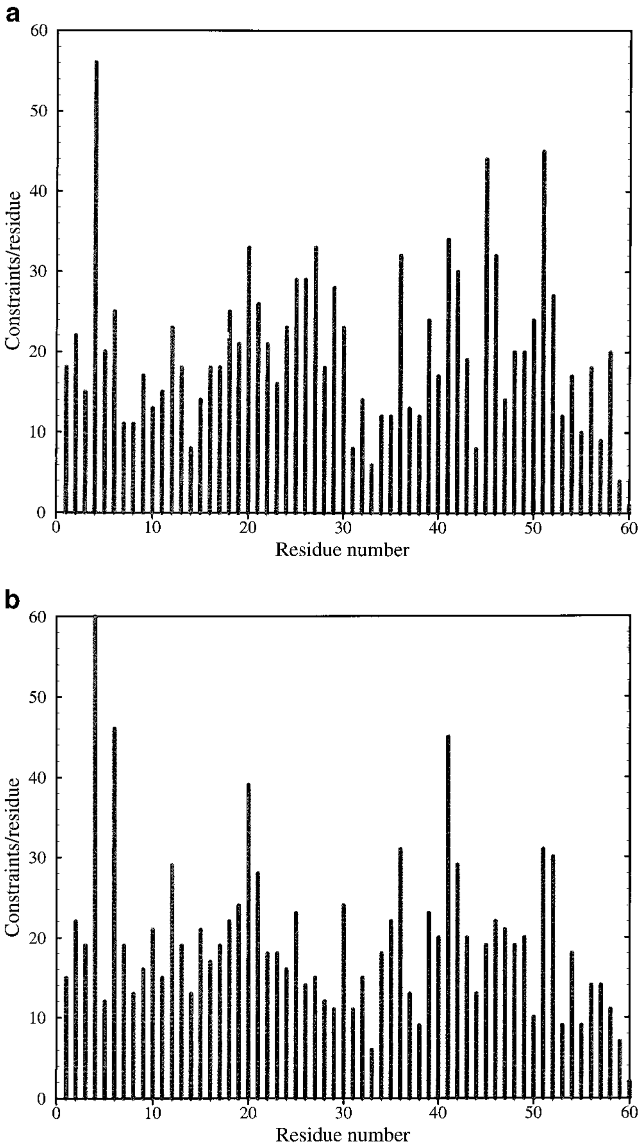

FIG. 4. (a) Plot of unambiguous constraints assigned by NOAH/DIAMOD per residue along the protein sequence at end of the NOAH/DIAMOD cycles. (b) Number of constraints per residue assigned manually used for the X-PLOR calculation.

Peak 847 was assigned to Lys9_QE-Hβ_2 by NOAH/DIAMOD because the peak position matches the chemical shifts of Lys9 Hβ_2 and QE (2.97 ppm, 1.95 ppm) within the chemical shift tolerance used in the NOAH/DIAMOD, whereas peak 865 with chemical shifts of (2.93 ppm, 1.98 ppm) is slightly outside this range. The chemical shifts of peak 865 (2.93 ppm, 1.98 ppm) match well with the long-range proton pair Lys12_QE-Ile58_Hβ_2 (2.94, 1.95), explaining the choice of NOAH/DIAMOD as an alternative possible assignment to the manual assignment. However, we do not have at present all chemical shifts of the side chain protons, so we cannot exclude other assignment possibilities for peak 865. Peak 476 was assigned by NOAH/DIAMOD to the proton pair Lys28_HN-Lys29_Hβ_2, the same distance constraint was not found in the manual assignment. However, the distance violation of about 1 to 2 Å for this constraint is still within the error range of using a different calibration scheme in the two independent approaches. In the second part of Table 3 the largest violations of the manually assigned distance constraints are given when cross checked against the NOAH/DIAMOD bundle structures. These six deviating constraints are again a small fraction (less than 1%) of the total number of manually derived distance constraints. They arise in most cases from weak long-range cross peaks with upper limit constraints of 5 to 6 Å, which were either not present in the automatically picked peak list or occurred in regions with severe overlap.

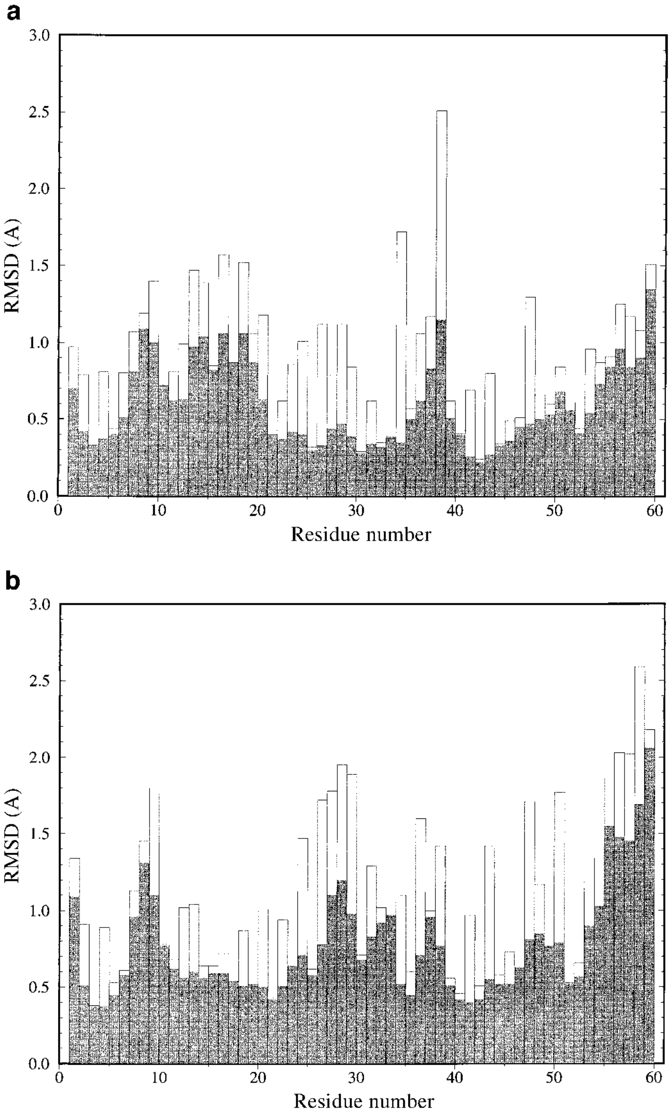

FIG. 5. (a) Pairwise RMSD of the final 20 bundle structures determined automatically. The gray bar shows the backbone RMSD changes for each residue and open bar for all heavy atoms. (b) Pair wise RMSD changes of 20 final X-PLOR structures calculated using manually assigned distance constraints.

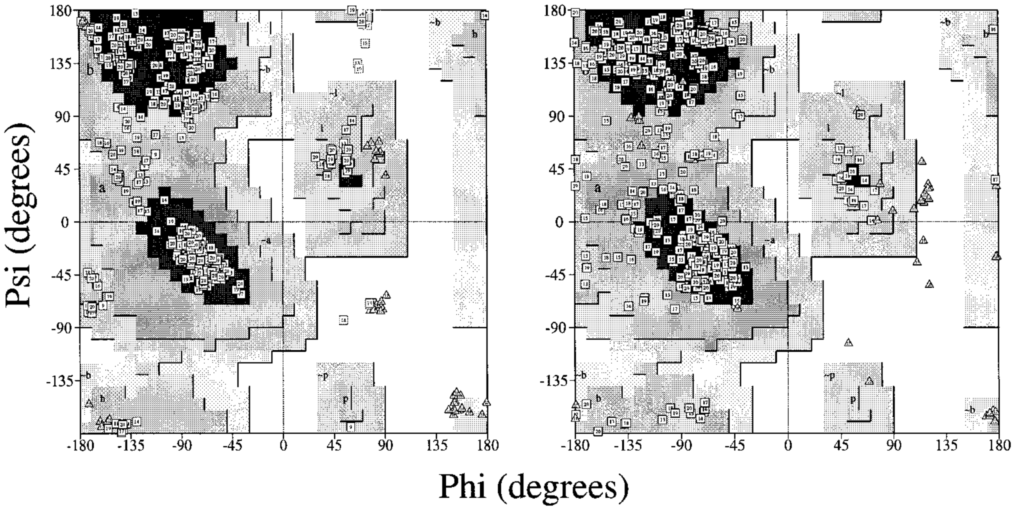

FIG. 6. Ramachandran plot of final 20 energy refined NOAH/DIAMOD structures (left) and the Ramachandran plot of final 20 X-PLOR structures using manually assigned NOE distance constraints (right).

The numbers of distance constraints per residue determined by each method are shown graphically in Figs. 4a and 4b. The distribution of the distance constraints per residue is not identical but highly similar in both methods, e.g., a high number of constraints at residues 4, 6, 20, 41, and 51 and fewer constraints at around residues 14, 31 to 33, and towards the C-terminal region.

Comparison of the 3D Structures Determined by the Two Methods

The sampling of the bundle of the 20 final energy refined structures, calculated by NOAH/DIAMOD/FANTOM and X-PLOR, is shown in Figs. 5a and 5b. The pairwise RMSD for backbone and all heavy atoms of the final structures to their mean structure at each residue is shown in Fig. 5a for the 20 FANTOM structures and in Fig. 5b for the 20 best X-PLOR structures. The RMSD distribution of the two methods is the same, particularly in the well-defined regions 4 to 6, at around residue 20 and 40 to 45. These regions correlate quite well with the highly constrained regions as shown in Fig. 4. The sampling of the backbone dihedral angles ϕ and φ is shown in the Ramachandran plots (Fig. 6) of the 20 final structures. Most residues of all structures determined by both methods are within the low-energy regions, except for a few residues in the loop areas. The slightly larger scatter in the X-PLOR structures (Fig. 6, right) is due to the nonuse of angular constraints from statistical distribution data. The backbone folds of the two mean structures are compared in Fig. 7 in stereo view. Both structures agree quite well, especially in the regions of regular secondary structures of CsE-v5, i.e., at residues 2–5, 18–26, 32–36, and 39–45. The backbone RMSD values of these well-defined regions between the two mean structures are 0.84 Å for backbone and 1.48 Å for all heavy atoms. Major deviations between the two structures are seen in the loop region

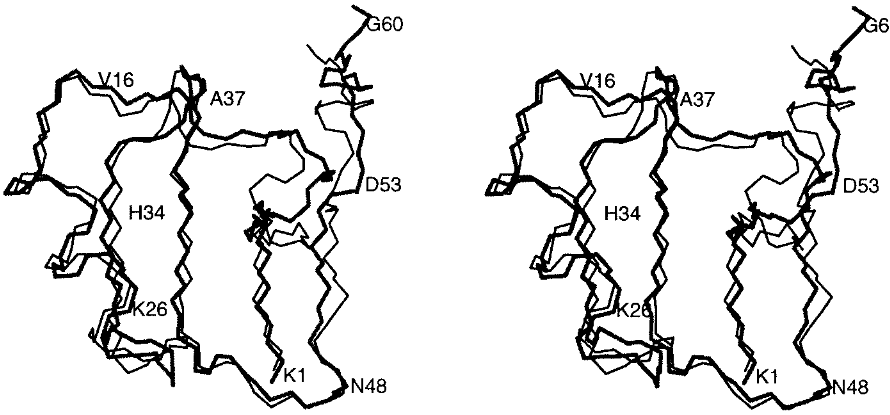

FIG. 7. Superposition of the energy minimized X-PLOR (thick line) and NOAH/DIAMOD mean structures.

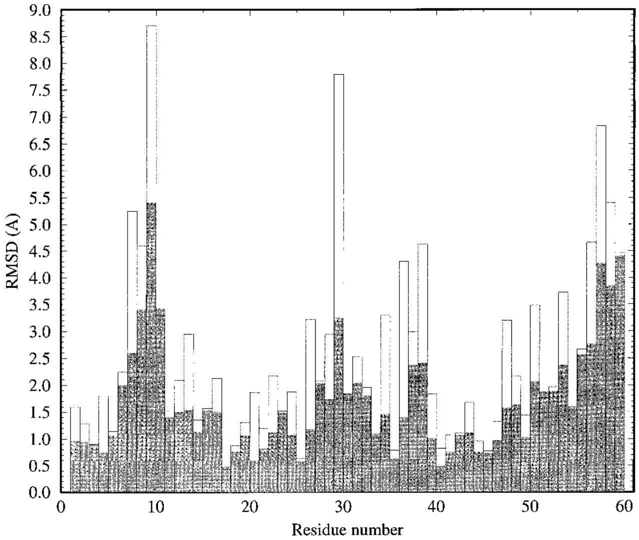

FIG. 8. Plot of residue-by-residue RMSD between the energy minimized X-PLOR and NOAH/DIAMOD mean structures. The gray bar is for the backbone and the open bar is for all heavy atoms.

near residue Lys 9 and toward the C-terminal region (Fig. 8), which have a small number of constraints (Fig. 4) in both methods and also show the largest deviations within the bundle of structures (Figs. 3a and 5).

The residue-by-residue contact maps for both final bundles of structures (Fig. 9) show the striking similarity of the α-helical region and the three stranded β-sheet topology in the independently determined structures. The maps also identify similar side chain contacts, as shown in the lower triangle. The patterns of these contacts overlap precisely with only a few exceptions. A minor deviation is the presence of the contacts of residues 7–9 to 29 in the NOAH/DIAMOD structures compared to the X-PLOR structures.

Functional motifs in the 3D structures of CsE-v5 are consistently shown in both structures. Two hydrophobic patches are observed in all structures, one hydrophobic patch is formed on one side of the protein by tyrosine 4, 36, and 41 and several other nonpolar residues nearby, and a second hydrophobic patch is observed on the other side of the protein by Cys15, Val16, Ala17, and Tyr20. Several scorpion toxins exhibit a "herringbone motif," consisting of orthogonally aligned aromatic side chains (*32, 36, 37*). This motif is also present in CsE-v5, as seen in Fig. 10, which shows the side chain conformations of tyrosines 4, 20, 36, and 41.

DISCUSSION

Quality of the Automated Assignments and 3D Structures of NOAH/DIAMOD/FANTOM

We compared the assignments and 3D structures determined by our automated assignment procedure with a completely independent conventional assignment and 3D structure determination. We used exactly the same time domain NMR spectra and the same manual sequential assignments for both procedures, but performed spectral data processing, NOESY cross-peak identification, cross-peak volume integration, distance calibration, and structure calculation completely independent in this test. The automatically determined structures agree within 1 Å in well-defined regions for the backbone fold. The extent of the assignment of the NOESY spectrum is similar by both methods, and the quality of the automated assignment is comparable to that of the manual procedure, even though the manual and automatic peak assignments did not always coincide. However, 96% of the automatically assigned peaks were consistent with the structures calculated from the manual assignment procedure.

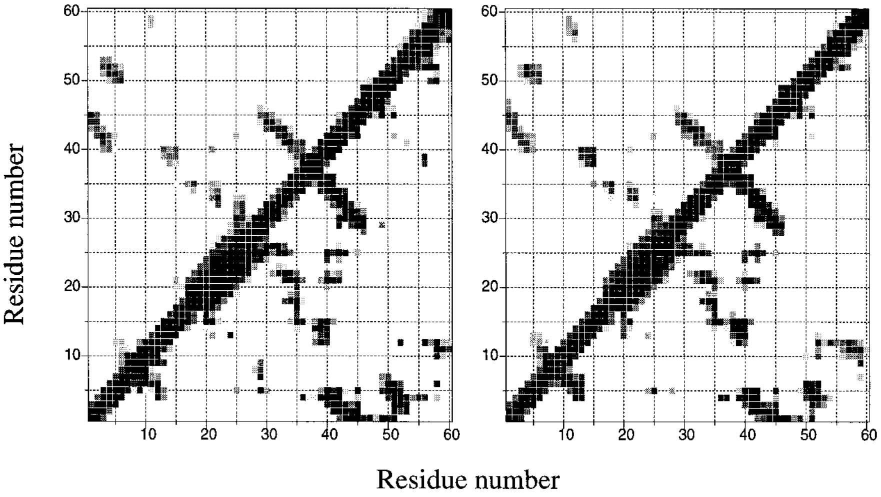

FIG. 9. Residue–residue contact map of final 20 energy minimized NOAH/DIAMOD structures (left) and X-PLOR structures (right). Black squares show the contact distances that are less than 3.0 Å, gray squares show the distances larger than 3.0 Å but less than 5.0 Å, and light gray squares show distances over 5.0 Å. The upper triangle shows the backbone contacts and the lower triangle shows all other contacts.

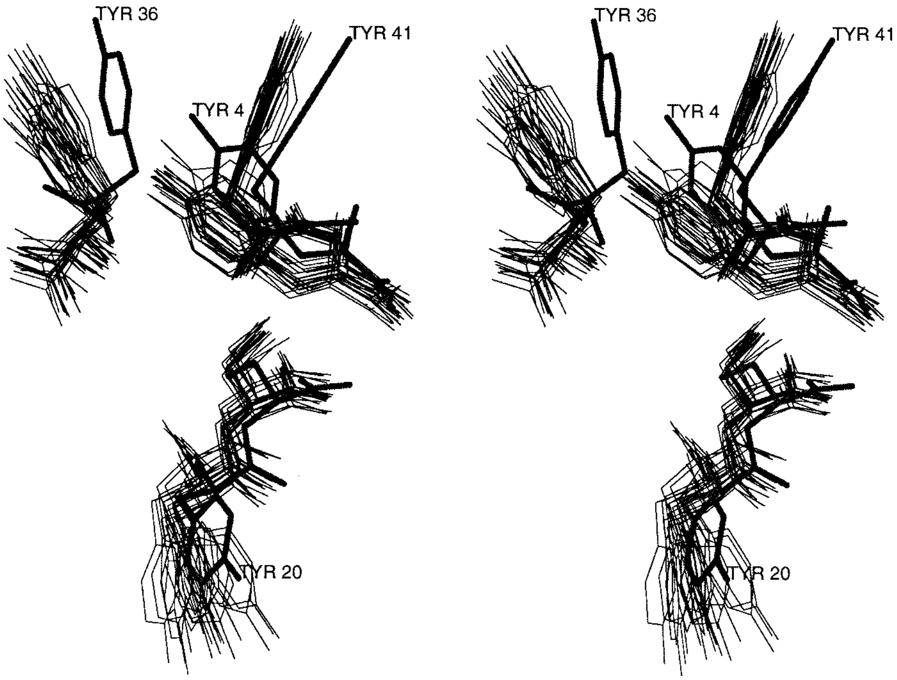

FIG. 10. The side-by-side stereo view of the TYR ring configurations between X-PLOR mean (thick line) and final 20 energy minimized NOAH/DIAMOD structures.

We have mentioned in Results that 174 NOESY peaks could not be assigned although their chemical shifts can be found in the proton's chemical shift lists because of incompatibility with the final 20 bundle structures. On the other hand, NOAH also identified 30 peaks that have no corresponding chemical shifts to all resolved protons in either the D1 or the D2 dimension and so cannot be assigned. Because the NOESY peaks were automatically picked, they do contain many noise peaks and other artifact peaks. Especially, we have used a relatively low contour level to pick all the cross peaks in order not to miss any possible very weak cross peak. It is our hope that those noise peaks could be gradually eliminated from the real peaks during the iteration assignment and structure-based filtering. In fact, our automatic approach assigned about 80 more peaks than the manual assignment. We have noticed, while visually examining the peaks on the spectra, that many of those incompatible peaks have very small peak intensities and bad lineshapes and would have not been picked using manual peak picking. It is quite possible that most of these peaks are just noise peaks or peaks contributed from impurities in the sample that happen to have overlapping chemical shifts with the protons. Some of those peaks are possibly real NOE peaks, which could not be assigned, as they arise from minor conformations in a flexible part of the proteins such as loop regions or surface-accessible side chains. Missing chemical shifts can also contribute to incompatibility of the peak assignment with the structure bundles, as they can lead to wrong assignments in regions of chemical shift overlaps. The advantage of using the NOAH/DIAMOD procedure is that one can use many more cross peaks and does not have to be concerned about noise peaks too much, as the algorithm will filter them out automatically during the iteration process.

Other possible sources of errors in the automated method are that the method is sensitive to the quality of the peak-picking procedure, the personal preference used for processing the NOESY spectra, estimating the cross peak intensities, etc. The differences between the manually and automatically calculated structures could also come from the different force fields used in X-PLOR and DIAMOD/FANTOM.

Overall, it is impossible to have an identical match between the manual and automatic assignment of the NOESY peaks and the calculated structures especially at the protein's loop regions. At present, we could not make any statement that manual assignment is more accurate than automatic assignment or vice versa. In fact the manual and automated structures were refined with different energy functions (X-PLOR vs FANTOM), which would result in some structural differences even if the constraint lists were identical. There is an intrinsic uncertainty in protein structures determined from NMR data arising from the specific rules used by a person or program in assigning NOESY spectra, as well as from details of the structure calculation and features of the particular software used for structure generation.

Final Remarks

We have assigned the 2D NOESY spectrum of the scorpion neurotoxin CsE-v5 and determined its three-dimensional structure using the NOAH/DIAMOD/FANTOM suite. The comparison of the assignments and the 3D structures by an independent manual assignment and 3D structure calculation shows that the automated procedure can determine the NOE peak assignments and 3D structures comparable in quality to the manual procedure. The automatically determined structures agree within 1 Å in well-defined regions. The NOAH/DIAMOD procedure saves time in the interpretation of NOESY spectra if used in combination with manual assignments. We plan to further improve the method and expand its scope toward a completely automated assignment procedure.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants MCB-9630775 (NRK) and DBI-9714937 (WB), NCI Grant CA-13148 (UAB NMR Facility), DOE Grant DE-FG03-96ER62267 (WB), and the Sealy and Smith Foundation (WB).

REFERENCES

1. P. Güntert, W. Braun, and K. Wüthrich, Efficient computation of three-dimensional protein structures in solution from NMR data using the program DIANA and the supporting programs CALIBA, HABAS, and GLOMSA, *J. Mol. Biol.* **217,** 517–530 (1991).
2. P. Güntert, K. Berndt, and K. Wüthrich, The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination, *J. Biomol. NMR* **3,** 601–606 (1993).
3. B. J. Hare and J. H. Prestegard, Application of neural networks to automated assignment of NMR spectra of proteins, *J. Biomol. NMR* **4,** 35–46 (1994).
4. R. Meadows, E. Olejniczak, and S. Fesik, A computer-based protocol for semiautomated assignments and 3D structure determination of proteins, *J. Biomol. NMR* **4,** 79–96 (1994).
5. D. Zimmermann, C. Kulikowski, L. Wang, B. Lyons, and G. Montelione, Automated sequencing of amino acid spin systems in proteins using multidimensional HCC(CO)NH-TOCSY spectroscopy and constraint propagation methods from artificial intelligence, *J. Biomol. NMR* **4,** (1994).
6. C. Antz, K. Neidig, and H. Kalbitzer, A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis, *J. Biomol. NMR* **5,** 287–296 (1995).
7. N. Morelle, B. Brutscher, J. Simorre, and D. Marion, Computer assignment of the backbone resonances of labelled proteins using two-dimensional correlation experiments, *J. Biomol. NMR* **5,** 154–160 (1995).
8. C. Mumenthaler and W. Braun, Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry, *J. Mol. Biol.* **254,** 465–480 (1995).
9. M. Nilges, Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities, *J. Mol. Biol.* **245,** 645–660 (1995).
10. D. Zimmerman and G. Montelione, Automated analysis of nuclear magnetic resonance assignments for proteins, *Curr. Opin. Struct. Biol.* **5,** 664–673 (1995).
11. M. Nilges, Structure calculation from NMR data, *Curr. Opin. Struct. Biol.* **6,** 617–623 (1996).
12. C. Mumenthaler, P. Guntert, W. Braun, and K. Wuthrich, Automated combined assignment of NOESY spectra and three-dimensional protein structure determination, *J. Biomol. NMR* **10,** 351–362 (1997).
13. M. Nilges, J. Macias, S. O'Donoghue, and H. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from beta-spectrin, *J. Mol. Biol.* **269,** 408–422 (1997).
14. D. Zimmerman, C. Kulikowski, Y. Huang, W. Feng, M. Tashiro, S. Shimotakahara, C. Chien, R. Powers, and G. Montelione, Automated analysis of protein NMR assignments using methods from artificial intelligence, *J. Mol. Biol.* **269,** 592–610 (1997).
15. Y. Xu, J. Wu, D. Gorenstein, and W. Braun, Automated 2D NOESY assignment and structure calculation of Crambin (S22/I25) with the self-correcting distance geometry based NOAH/DIAMOD programs, *J. Magn. Reson.* **136,** 76–85 (1999).
16. Y. Xu, C. H. Schein, and W. Braun, in "Biological Magnetic Resonance" (N. R. Krishna and L. J. Berliner, Eds.), Vol. 17, pp. 37–76, Kluwer Academic/Plenum, New York, 1999.
17. H. N. B. Moseley and G. T. Montelione, Automated analysis of NMR assignments and structures for proteins, *Curr. Opin. Struct. Biol.* **9,** 635–642 (1999).
18. G. T. Montelione, C. B. Rios, G. V. T. Swapna, and D. E. Zimmerman, in "Biological Magnetic Resonance" (N. R. Krishna and L. J. Berliner, Eds.), Vol. 17, pp. 81–130, Kluwer Academic/Plenum, New York, 1999.
19. W. Braun and N. Go, Calculation of protein conformations by proton–proton distance constraints, *J. Mol. Biol.* **186,** 611–626 (1985).
20. W. Braun, Distance geometry and related methods for protein structure determination from NMR data, *Q. Rev. Biophys.* **19,** 115–157 (1987).
21. G. Hänggi and W. Braun, Pattern recognition and self-correcting distance geometry calculations applied to myohemerythrin, *FEBS Lett.* **344,** (1994).
22. A. Yamano, N. Heo, and M. Teeter, Crystal structure of Ser-22/Ile-25 form crambin confirms solvent, side chain substate correlations, *J. Biol. Chem.* **272,** 9597–9600 (1997).
23. J. M. Simard, H. Meves, and D. D. Watt, in "Natural Toxins: Toxicology, Chemistry, and Safety" (R. F. Keeler, N. B. Mandava, and A. T. Tu, Eds.), pp. 236–263, Alaken, Fort Collins, CO, 1992.
24. D. Babin, D. Watt, S. Goos, and R. Mlejnek, Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing, *Arch. Biochem. Biophys.* **164,** 694–706 (1974).
25. K. Wuthrich, "NMR of Proteins and Nucleic Acids," Wiley, New York, 1986.
26. R. Abagyan and M. Totrov, Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins, *J. Mol. Biol.* **235,** 983–1002 (1994).
27. Y. Xu, I. P. Sugar, and N. R. Krishna, VARIABLE target intensity-restrained global optimization (VARTIGO) procedure for determining three dimensional structures of polypeptide from NOESY data: Application to gramicidin S, *J. Biomol. NMR* **5,** 39 (1995).
28. Y. Xu and N. R. Krishna, Structure determination from NOESY intensities using a metropolis simulated annealing (MSA) refinement of dihedral angles, *J. Magn. Reson. B* **108,** 192–196 (1995).
29. P. Guntert and K. Wuthrich, Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral constraints, *J. Biomol. NMR* **1,** 447–456 (1991).

30. P. C. Driscoll, A. M. Gronenborn, L. Beress, and G. M. Clore, Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata:* A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, *Biochemistry* **28,** 2188–2198 (1989).

31. M. Nilges, A. M. Gronenborn, A. T. Brunger, and G. M. Clore, Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2, *Protein Engin.* **2,** 27–38 (1988).

32. W. Lee, M. J. Jablonsky, D. D. Watt, and N. R. Krishna, Proton nuclear magnetic resonance and distance geometry/simulated annealing studies on the variant-1 neurotoxin from the new world scorpion *Centruroides* sculpturatus Ewing, *Biochemistry* **33,** 2468–2475 (1994).

33. T. Schaumann, W. Braun, and K. Wuthrich, A program, FANTOM, for energy refinement of polypeptides and proteins using a Newton–Raphson minimizer in the torsion angle space, *Biopolymers* **29,** 679–694 (1990).

34. B. Freyberg and W. Braun, Efficient search for all low energy conformations of Metenkephalin by Monte Carlo methods, *J. Comp. Chem.* **12,** 1065 (1991).

35. R. Fraczkiewicz and W. Braun, Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules, *J. Comp. Chem.* **19,** 319–333 (1998).

36. M. J. Jablonsky, D. D. Watt, and N. R. Krishna, Solution structure of an old world-like neurotoxin from the venom of the new world scorpion *Centruroides sculpturatus* Ewing, *J. Mol. Biol.* **248,** 449–458 (1995).

37. W. Lee, C. H. Moore, D. D. Watt, and N. R. Krishna, Solution structure of the variant-3 neurotoxin from *Centruroides sculpturatus* Ewing, *Eur. J. Biochem.* **219,** 89–95 (1994).

38. M. J. Jablonsky, P. L. Jackson, J. O. Trent, D. D. Watt, and N. R. Krishna, Solution structure of a beta-neurotoxin from the New World scorpion Centruroides sculpturatus Ewing, *Biochem. Biophys. Res. Commun.* **254,** 406–412 (1999).