Journal of Biomolecular NMR, **15:** 103–113, 1999.
KLUWER/ESCOM

Application of automated NOE assignment to three-dimensional structure refinement of a 28 kDa single-chain T cell receptor

Brian J. Hare & Gerhard Wagner
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, U.S.A.

Received 26 May 1999; Accepted 2 August 1999

Key words: chemical shifts, NOESY, protein NMR

Abstract

An automated procedure for NOE assignment and three-dimensional structure refinement is presented. The input to the procedure consists of (1) an ensemble of preliminary protein NMR structures, (2) partial sequence-specific assignments for the protein and (3) the positions and volumes of unassigned NOESY cross peaks. Chemical shifts for unassigned side chain protons are predicted from the preliminary structures. The chemical shifts and unassigned NOESY cross peaks are input to an automated procedure for NOE assignment and structure calculation (ARIA) [Nilges et al. (1997) *J. Mol. Biol.*, **269**, 408–422]. ARIA is optimized for the task of structure refinement of larger proteins. Errors are filtered to ensure that sequence-specific assignments are reliable. The procedure is applied to the 27.8 kDa single-chain T cell receptor (scTCR). Preliminary NMR structures, nearly complete backbone assignments, partial assignments of side chain protons and more than 1300 unassigned NOESY cross peaks are input. Using the procedure, the resonant frequencies of more than 40 additional side chain protons are assigned. Over 400 new NOE cross peaks are assigned unambiguously. Distances derived from the automatically assigned NOEs improve the precision and quality of calculated scTCR structures. In the refined structures, a hydrophobic cluster of side chains on the scTCR surface that binds major histocompatibility complex (MHC)/antigen is revealed. It is composed of the side chains of residues from three loops and stabilizes the conformation of residues that interact with MHC.

Introduction

The assignment of NOESY spectra is a limiting step in the determination of three-dimensional structures of proteins by NMR. For large proteins, ^{13}C-dispersed, ^{15}N-dispersed and ^{1}H-^{1}H NOESY spectra are often used to supplement other methods for determining sequence-specific assignment of side chains. Moreover, NOESY data provide the majority of restraints used to calculate high-resolution NMR structures. Therefore, methods for rapid and accurate assignment of NOESY spectra are extremely important.

Since manual assignment of NOESY data is tedious and prone to error, significant effort has been directed toward automating NOE assignment. Attempts have been made to automatically assign NOESY data without knowledge of sequence-specific assignments or three-dimensional structure (Oshiro and Kuntz, 1993; Kraulis, 1994). Methods for simultaneous NOESY peak list assignment and structure calculation include NOAH (Mumenthaler and Braun, 1995; Mumenthaler et al., 1997) and ARIA (Nilges et al., 1997). These routines are interfaced with NMR structure determination software and have been applied successfully to a number of proteins. In these methods, NOESY cross peaks are assigned using chemical shift lists from proteins with complete or nearly complete sequence-specific assignments. ARIA and NOAH are well suited for structure determination of smaller proteins, since nearly complete side chain assignments can be readily obtained for these proteins from HCCH-TOCSY and ^{15}N-edited proton TOCSY data. For larger proteins, however, ambiguity in side

chain assignments is a barrier to the application of these methods.

Methods for obtaining approximate chemical shifts of unassigned nuclei may overcome this limitation. Random coil chemical shifts determined for small, unstructured peptides (Wishart et al., 1995) provide a crude estimate for the chemical shifts of unassigned nuclei. More accurate estimates can be obtained for ^{1}H and ^{13}C chemical shifts from the chemical shift assignments of a close homologue of the target protein (Wishart et al., 1997). If the three-dimensional structure of the target protein is known, semi-empirical methods predict ^{1}H chemical shifts of non-exchangeable protons with reasonable accuracy (Ösapay and Case, 1991; Williamson and Asakura, 1993; Sitkoff and Case, 1997) and ab initio methods can be used to predict ^{13}C and ^{15}N chemical shifts (Oldfield, 1995).

We are interested in using automated NOE assignment to collect distance restraints in order to refine the three-dimensional structure of the D10 single-chain T cell receptor (scTCR). The scTCR is a 27.8 kDa protein containing 255 amino acid residues. We previously reported the solution structure of the scTCR (Hare et al., 1999). The protein consists of the two variable domains (V_α and V_β) from the murine D10 TCR joined by a hydrophilic peptide linker. Each of the domains belongs to the immunoglobulin superfamily. The domains are closely associated in a pseudodyad arrangement in the structure.

Here, we show that the method of iterative NOE assignment and structure calculation can be extended in order to make sequence-specific assignment of side chain protons. Because a preliminary three-dimensional structure for the D10 scTCR protein is available, but homologous proteins have not yet been assigned, we chose to use semi-empirical methods to predict the chemical shifts for unassigned protons. The side chain assignments facilitate automated assignment of NOESY peak lists, thereby providing distance restraints used to calculate refined structures. Side chain conformations are determined more precisely in the refined structures than in the preliminary structures. A hydrophobic cluster of side chains is revealed on the surface of V_α. It is composed of the residues from three loops that bind major histocompatibility complex (MHC)/antigen and stabilizes the conformation of residues that interact with MHC.

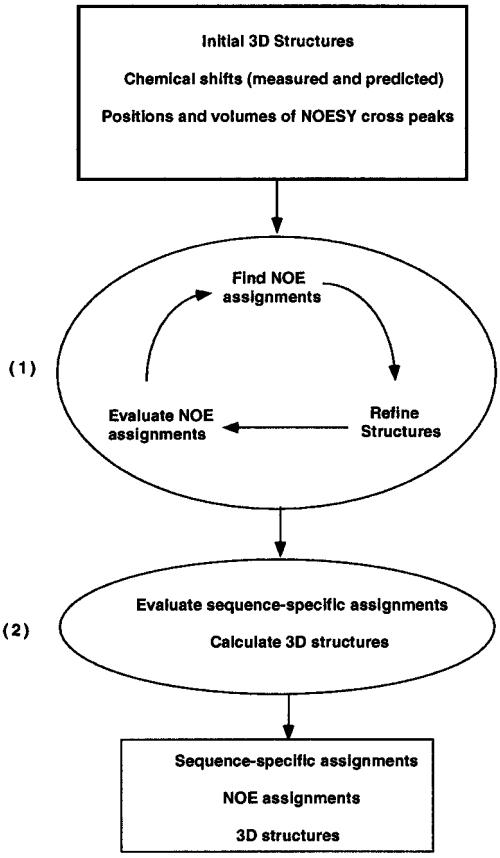

Figure 1. Flow diagram for the structure refinement procedure.

Methods

Overview of the automated assignment method

The main steps of the procedure are illustrated in Figure 1. Inputs to the automated assignment procedure include initial 3D structures, measured and predicted sequence-specific assignments and the positions and volumes of NOESY cross peaks. We used ARIA (Nilges et al., 1997) to assign NOESY cross peaks and refine 3D structures in step (1). Briefly, ARIA consists of a series of routines to assign and calibrate NOE data, analyze distance violations and merge the data into a table of distance restraints read by XPLOR. NOESY cross peaks are assigned by ARIA using peak lists generated from raw NOESY spectra together with chemical shift lists for the target protein. A distance criterion is employed to discard assignment possibilities that correspond to large distances in the structure. ARIA iteratively assigns NOEs by using structures calculated in one step to assign and calibrate NOEs in the next step. ARIA output includes a list of assigned

distance constraints. In step (2) of the automated assignment procedure, the sequence-specific assignments included among the distance constraints are evaluated, as described below, and new 3D structures are calculated.

Input data

The ensemble of preliminary D10 scTCR structures was calculated with NOE distance, dihedral angle and hydrogen bond restraints (Table 2). These restraints are a subset of those used to calculate our original structures (Hare et al., 1999). They include only those NOE distance restraints that could be manually assigned without ambiguity. In the course of our original structure determination (Hare et al., 1999), backbone $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{1}H^{N}$ and ^{15}N were assigned nearly completely. A total of 348 sequence-specific side chain proton assignments were obtained. Numerous unassigned NOESY cross peaks remained in a ^{15}N-edited NOESY-HSQC spectrum (90 ms mixing time). Manual assignment of these peaks is hampered by uncertainty in the sequence-specific assignments of many side chain protons.

A total of 1327 unassigned cross peaks were picked manually from a ^{15}N-edited NOESY-HSQC spectrum (90 ms mixing time) acquired on a 1 mM sample of uniformly ^{15}N-labeled D10 scTCR. Peaks were picked along ^{1}H and ^{15}N strips with chemical shifts assigned to the backbones of residues in either the V_{α} or V_{β} domains. Volumes of picked peaks were calculated using the program peakint (Schäfer, 1992).

The chemical shifts for unassigned side chain protons were predicted using the program SHIFTS version 3.0b2 (Ösapay and Case, 1991; Sitkoff and Case, 1997). Structures from the ensemble were input to the program one at a time. Predicted chemical shifts from all of the structures in the ensemble were averaged. The predicted chemical shifts for unassigned side chain protons were used to complete a proton chemical shift table for the D10 scTCR.

The chemical shifts are input as upper and lower limits in order to allow for uncertainty in the values of predicted chemical shifts. The upper and lower limits for predicted chemical shifts are calculated as follows:

$$\begin{aligned} \delta^{p}_{upper} &= \delta_{predict} + \Delta_{predict} \\ \delta^{p}_{lower} &= \delta_{predict} - \Delta_{predict} \end{aligned} \quad (1)$$

where $\delta_{predict}$ is the predicted chemical shift from SHIFTS and δ^{p}_{upper} and δ^{p}_{lower} are the upper and lower limits, respectively, for predicted chemical shifts input to the automated assignment procedure. The uncertainty in predicted chemical shift, $\Delta_{predict}$, used in the assignment procedure is chosen to be large enough that most chemical shifts will be within $+/-\ \Delta_{predict}$ of the predicted value, but small enough to allow the types of most protons within a side chain to be distinguishable. We chose to use the value $\Delta_{predict} = 0.3$ ppm based on the accuracy of chemical shift prediction for the D10 scTCR (see Results).

For measured proton and nitrogen chemical shifts, the uncertainty in measured chemical shift, $\Delta_{measure}$, is used to calculate upper (δ^{m}_{upper}) and lower (δ^{m}_{lower}) chemical shift bounds.

$$\begin{aligned} \delta^{m}_{upper} &= \delta_{measure} + \Delta_{measure} \\ \delta^{m}_{lower} &= \delta_{measure} - \Delta_{measure} \end{aligned} \quad (2)$$

For measured amide nitrogen chemical shifts, $\Delta_{measure} = 0.2$ ppm. For proton chemical shifts measured in the direct and indirect dimensions of 3D datasets, $\Delta_{measure} = 0.02$ and 0.03 ppm, respectively. These values are larger than the spectral resolution of the data because chemical shifts are affected by small differences in sample conditions.

Assignment using ARIA

The ARIA protocol for NOE assignment, calibration and structure calculation used in step (1) of the assignment procedure (Figure 1) is similar to the protocol reported previously (Nilges et al., 1997). Only differences from this protocol will be discussed. To make the procedure computationally more efficient, ARIA is used to refine existing structures rather than to calculate structures starting from a random polypeptide chain. Fifteen structures are refined in each iteration. In the first iteration, each of the input structures is refined. In subsequent iterations, the ten lowest-energy structures from the previous iteration are refined. Five additional structures are calculated in each iteration by refining the five lowest-energy input structures. As suggested previously (Goll et al., 1998), the XPLOR parameters RSWITch and ASYMptote are reduced from 1.0 to 0.5 and 1.0 to 0.1, respectively, in the distance restraint potential, thereby minimizing NOE assignment errors. An ARIA run consisting of eight iterations is completed in 48 h on an SGI workstation with an R10000 processor.

Evaluating sequence-specific assignments and calculating 3D structures

In step (2) of the procedure, unambiguously assigned NOESY cross peak assignments are evaluated. All of the NOESY cross peaks unambiguously assigned

to protons with previously determined chemical shifts are accepted. NOESY cross peaks unambiguously assigned to protons with predicted chemical shifts are accepted only if the same proton was assigned by two or more NOESY cross peaks within a small chemical shift range ($2 * \Delta_{assign} = 0.050$ ppm). The considerations leading to this method for filtering assignments are described in the Discussion section. Distances derived using the calibration by ARIA of all accepted NOESY cross peaks are used to calculate structures using a simulated annealing protocol (Nilges et al., 1988) in XPLOR version 3.851 (Brünger, 1993). The stereochemical quality of the structures was assessed using PROCHECK_nmr (Laskowski et al., 1993). A total of 87 structures were calculated using the preliminary data together with the automatically assigned NOEs and 15 were accepted based on the criteria of low total energy, no distance violations greater than 0.5 Å and no dihedral angle violations greater than 0.5°.

Results

Chemical shift prediction

Chemical shifts for side chain protons in the D10 scTCR were predicted from an ensemble of preliminary NMR structures. To assess the accuracy of chemical shift prediction for side chain protons, predicted and measured chemical shifts for the 348 manually assigned side chain protons are compared in Figure 2. The rmsd between all measured and predicted side chain proton chemical shifts (0.34 ppm) is better than the rmsd between measured and random coil shifts (0.42 ppm). The correlation between measured and predicted chemical shifts is best for protons with chemical shifts between 2 ppm and 4 ppm (Figure 2) The majority of protons in this chemical shift range are side chain methylene protons. The dotted lines in Figure 2 bracket the predicted values that are in the range $+/-$ 0.3 ppm from the measured values. Greater than 70% of the predicted values are within this range.

NOE assignment

The automated procedure (Figure 1) assigns 405 of the 1327 unassigned input NOE cross peaks. Because the chemical shifts of nearly all backbone nuclei are known, assignments in the $^{1}H^{N}$ and ^{15}N dimensions are exclusively to nuclei with experimentally determined chemical shifts. The distribution of NOE assignments in the indirect proton dimension between protons with experimentally determined and predicted chemical shifts is shown in Figure 3. The majority

Table 1. New sequence-specific ^{1}H assignments for the D10 scTCR from the automated procedure

Residue	Atom type	Chemical shift (ppm)
$Gln^{6\beta}$	H^{β}	1.68
$Cys^{23\beta}$	H^{β}	3.68
$Gln^{25\beta}$	H^{γ}	2.74
$Asn^{27\beta}$	H^{β}	2.05
$Met^{32\beta}$	H^{β}	2.22
$Trp^{34\beta}$	H^{β}	3.54
$Tyr^{35\beta}$	H^{β}	2.83
$Arg^{36\beta}$	H^{γ}	1.22
$Gln^{37\beta}$	H^{β} or H^{γ}	2.38
$Gln^{37\beta}$	H^{β} or H^{γ}	2.55
$Asp^{38\beta}$	H^{β}	2.86
$Arg^{44\beta}$	H^{δ}	3.55
$Lys^{57\beta}$	H^{γ}	1.19
$Lys^{57\beta}$	H^{δ}	1.76
$Asp^{59\beta}$	H^{β}	3.20
$Lys^{66\beta}$	H^{β}	1.72
$Arg^{69\beta}$	H^{β}	1.88
$Arg^{69\beta}$	H^{γ}	1.28
$Pro^{70\beta}$	H^{β}	1.98
$Gln^{72\beta}$	H^{γ}	2.09
$Gln^{86\beta}$	H^{β}	2.02
$Ser^{88\beta}$	H^{β}	4.36
$Phe^{91\beta}$	H^{β}	3.27
$Arg^{113\beta}$	H^{δ}	3.18
$Gln^{5\alpha}$	H^{β} or H^{γ}	1.91
$Gln^{5\alpha}$	H^{β} or H^{γ}	2.22
$Glu^{14\alpha}$	H^{β} or H^{γ}	2.00
$Glu^{14\alpha}$	H^{β} or H^{γ}	2.29
$Leu^{20\alpha}$	H^{β}	2.08
$Asp^{26\alpha}$	H^{β}	1.92
$Asp^{30\alpha}$	H^{β}	2.49
$Pro^{39\alpha}$	H^{δ}	4.06
$Lys^{41\alpha}$	H^{β} or H^{γ}	1.66
$Lys^{41\alpha}$	H^{δ}	0.92
$Val^{52\alpha}$	H^{γ}	0.60
$Lys^{55\alpha}$	H^{γ}	0.60
$Lys^{56\alpha}$	H^{β}	1.82
$Arg^{61\alpha}$	H^{β}	2.04
$Glu^{70\alpha}$	H^{β}	2.14
$Leu^{75\alpha}$	H^{β}	0.48
$Tyr^{88\alpha}$	H^{ϵ}	6.62
$Lys^{103\alpha}$	H^{β}	0.99
$Leu^{104\alpha}$	H^{δ}	0.28

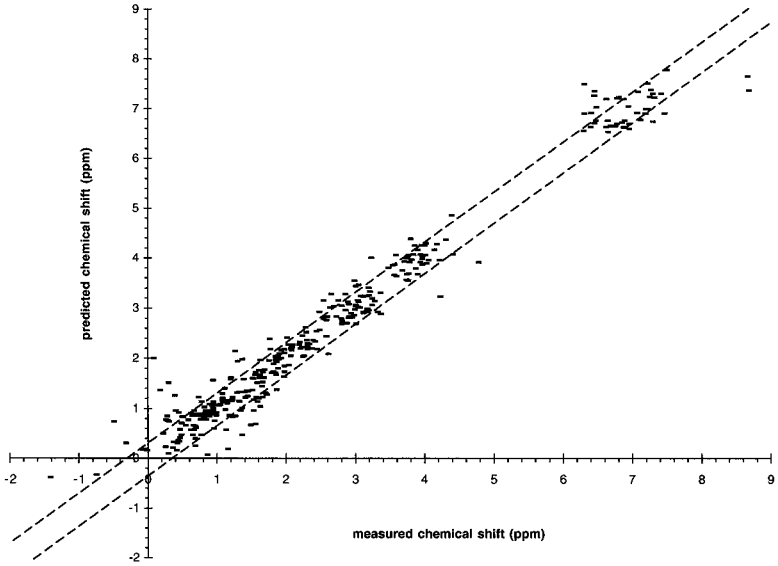

Figure 2. Correlation between measured chemical shift values and chemical shifts predicted by the program SHIFTS (Sitkoff and Case, 1997) for side chain protons in the D10 scTCR. Only side chain protons with manually assigned chemical shifts are shown. The dotted lines bracket the range $+/-$ 0.3 ppm between measured and predicted values.

of NOEs are assigned to protons with experimentally determined chemical shifts.

Among NOEs assigned to protons with experimentally determined chemical shifts, over 50% are assigned to $^1H^N$ or $^1H^\alpha$ protons. Many of the others are assigned to methyl protons and aromatic protons. A new interdomain NOE between the $^1H^N$ of Gln$^{106\beta}$ and a $^1H^\delta$ of Leu$^{45\alpha}$ is also assigned.

The NOEs to protons with predicted chemical shifts produce chemical shift assignments for 43 side chain protons (Table 1). Most of the chemical shift assignments are for protons in methylene groups from aliphatic side chains, but chemical shift assignments for a Tyr $^1H^\epsilon$ and a Leu $^1H^\delta$ are also determined. Two NOEs are assigned to most of the protons in Table 1. Three NOEs are assigned to each of 10 protons. For example, the chemical shift of $^1H^\beta$ of Asp$^{26\alpha}$, located in the CDR1β loop, is assigned by NOEs to $^1H^N$ of Ser$^{27\alpha}$, Thr$^{28\alpha}$ and Phe$^{29\alpha}$ (Figure 4A). Most of the NOEs that determine chemical shift assignments are intra-residue or sequential rather than medium or long range (Figure 3). In some cases, however, longer-range NOEs determine chemical shift assignments. For example, the $^1H^\epsilon$ chemical shift of Tyr$^{88\alpha}$ is determined by NOEs from $^1H^N$ protons of Ala$^{86\alpha}$ and Leu$^{112\alpha}$ (Figure 4B).

D10 scTCR structures

Statistics for D10 scTCR structures calculated with and without the automatically assigned NOEs are compared in Tables 2 and 3. The addition of the automatically assigned NOE data causes only a slight increase in the mean total energy of the final ensemble of structures (Table 2). The ranges in energy among structures in the preliminary (477.5–519.3 kcal/mol) and final (481.6–533.4 kcal/mol) ensembles are also similar.

The additional distance restraints result in more precise structures. The largest improvement is noted in the definition of the side chains (Table 2). The largest decrease in average rmsd from the mean structure occurs for residues in loops connecting the β strands (Table 3). An overall improvement in the quality of the structures calculated with the automatically assigned NOE data is shown by an increase in the percentage of residues with backbone ϕ and ψ angles in most favored regions of the Ramachandran plot from 59% to 63% (Table 2).

Discussion

NOE assignment

Overall, 40% of the NOEs assigned in the D10 scTCR by the automated procedure are long range. The percentage of long-range NOEs is highest (45%) among

Table 2. Statistics for the ensembles of preliminary and refined D10 scTCR structures[a]

	Preliminary	Refined
NOE distance restraints		
Total	1615	2020
Intraresidue	630	680
Sequential ($\|i-j\|=1$)	500	648
Medium range ($\|i-j\|<5$)	77	123
Long range	384	544
Interdomain	24	25
Dihedral angle restraints (ϕ,ψ)	277	277
Hydrogen bonds	60	60
Average energies (kcal mol^{-1})		
E_{total}	501.4 ± 13	510.1 ± 15
E_{repel}[b]	129.8 ± 5.6	134.6 ± 7.7
Ramachandran plot		
Most favorable regions	58.8 ± 3.2	63.2 ± 2.2
Additionally allowed regions	33.7 ± 3.5	31.2 ± 2.1
Generously allowed regions	6.1 ± 1.7	4.6 ± 1.1
Disallowed regions	1.4 ± 0.7	1.0 ± 0.6
Average rms deviations		
Ideal bonds (Å)	0.002 ± 0.000	0.002 ± 0.000
Ideal angles (°)	0.528 ± 0.004	0.531 ± 0.004
Distance restraints (Å)	0.019 ± 0.002	0.017 ± 0.001
Dihedral restraints (°)	0.230 ± 0.048	0.250 ± 0.039
Average rms deviations from the mean structure (Å)		
Secondary structures[c]	0.99 ± 0.24	0.82 ± 0.20
Backbone heavy atoms[d]	1.52 ± 0.23	1.28 ± 0.32
All heavy atoms[d]	2.15 ± 0.18	1.86 ± 0.30

[a] The reported statistics are averages over ensembles of 15 preliminary structures and 15 refined structures. The ensembles of preliminary and final structures were selected from 80 and 87 starting structures, respectively. The structures in the ensembles have low energy, no distance violations greater than 0.5 Å and no dihedral angle violations greater than 5°.
[b] The XPLOR repel energy $E_{repel} = k_{repel}\ (\max(0, S*r_{min})^2 - r^2))^2$ was evaluated with $k_{repel} = 1$, $S = 0.8$.
[c] Backbone heavy atoms (C^{α}, C, N^{H}) of β-domain residues: 3–6,11–13,20–25,32–36,44–48,56–57,66–69,75–79,88–93, 107–114; α-domain residues: 3–6,9–10,18–23,32–37,44–47,56–59,63–66,73–77,86–92,106–112.
[d] β-domain residues: 3–116; α-domain residues: 1–115.

NOEs to protons with assigned chemical shifts (Figure 3). The large number of long-range NOEs results from (1) the long mixing time (90 ms) of the NOESY-HSQC data set that was assigned here, (2) the β-strand secondary structure of the scTCR protein and (3) the prior manual assignment of many intra-residue and sequential NOEs from datasets with shorter mixing times (Hare et al., 1999). Facility in assigning long-range NOEs is an advantage of using automated NOE assignment for structure refinement, since the structural information used to evaluate assignment options is difficult to incorporate into manual assignment procedures.

Most of the chemical shift assignments determined using the automated procedure are methylene protons from aliphatic side chains. More than 50% are methylene protons from Gln, Glu, Lys or Arg residues. Many of the assigned methylene protons are in residues that are exposed to the surface. As a result, most of the NOEs that determine chemical shift assignments are intra-residue or sequential rather than medium or long range (Figure 3). Overrepresen-

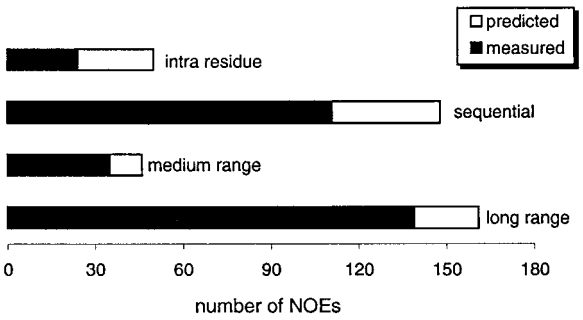

Figure 3. Bar graph showing the distribution of NOEs assigned by the automated procedure. NOEs to protons with measured (i.e. experimentally determined) and predicted chemical shifts are shown as filled and open bars, respectively. The numbers of intra-residue, sequential ($|i - j| = 1$), medium range ($|i - j| < 5$) and long range ($|i - j| \geq 5$) NOEs are shown.

tation of methylene protons among protons assigned automatically results from their underrepresentation among the protons assigned manually. Furthermore, chemical shift prediction for side chain methylene protons is quite good (Figure 2).

Predicting chemical shifts for side chain protons appears to improve the performance of the automated assignment procedure. More than half of the chemical shifts assigned to protons by the automated procedure (Table 1) differ from random coil chemical shifts by more than $\pm$ 0.3 ppm. 20% differ by more than 0.5 ppm. For example, the chemical shift of an $^1H^{\beta}$ of Asp$^{26\alpha}$ (1.92 ppm), assigned by the NOE cross peaks shown in Figure 4A, is shifted upfield by more than 0.7 ppm from the random coil value. The unusual chemical shift for $^1H^{\beta}$ of Asp$^{26\alpha}$ results from a ring-current effect of the nearby aromatic residue Phe$^{29\alpha}$. Protons with unusual chemical shifts are often difficult to assign manually. In contrast, chemical shift effects predicted from the three-dimensional structure are easily incorporated into the automated assignment procedure. More accurate chemical shift prediction will improve the performance of the automated assignment procedure by increasing the percentage of chemical shifts predicted within a small tolerance. For example, 'homologous assignment' (Wishart et al., 1997) should become more useful as the number of proteins with chemical shift assignments increases.

Accuracy of assignment

Prediction of proton chemical shift from three-dimensional structures is not sufficiently accurate that structural information is sufficient to ensure accurate assignments in spectral regions with high peak density.

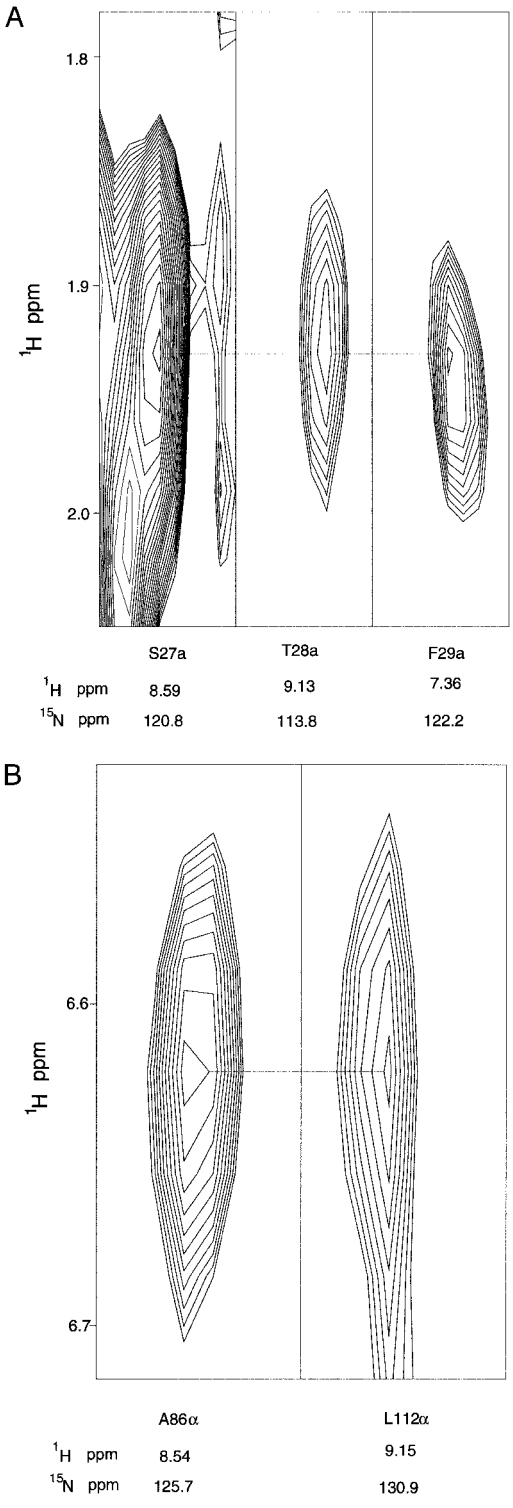

Figure 4. Two spectral regions from the 3D ^{15}N-edited NOESY-HSQC spectrum (90 ms mixing time). Strips along the indirect ^{1}H dimension are shown. (A) The NOE cross peaks used by the automated assignment procedure to assign the $^1H^{\beta}$ of Asp$^{26\alpha}$. (B) NOE cross peaks used by the automated procedure to assign $^1H^{\epsilon}$ of Tyr$^{88\alpha}$.

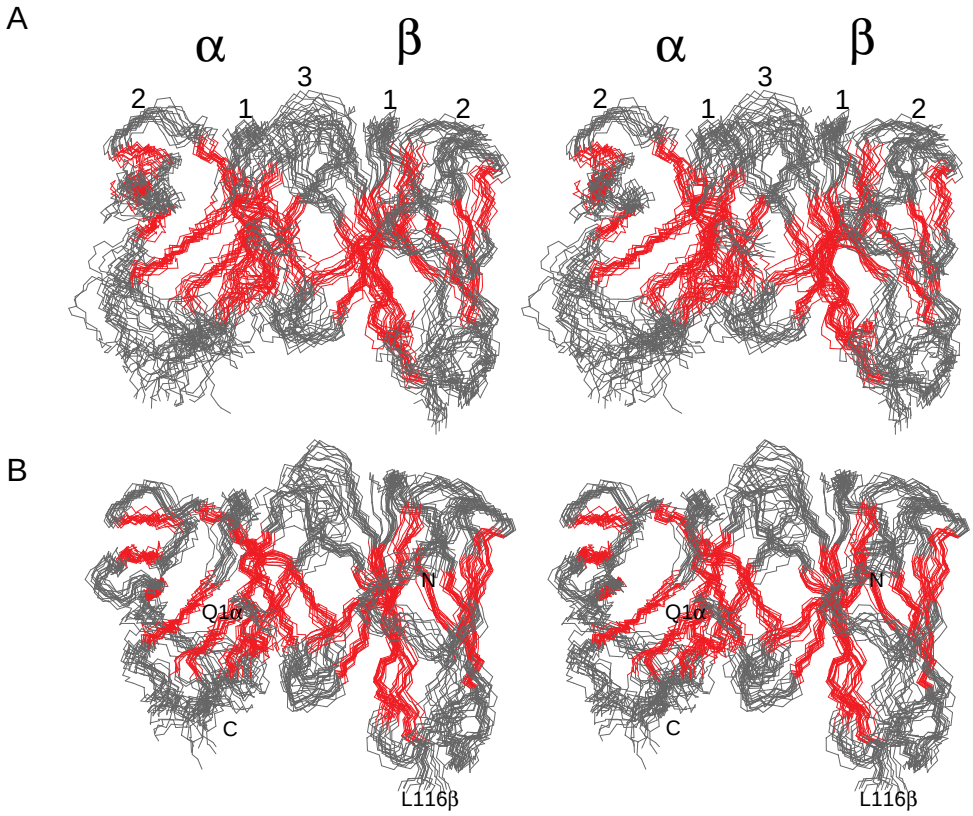

Figure 5. Stereoviews of the ensembles of structures calculated without (A) and with (B) distance restraints derived from the automatically assigned NOEs. Backbone residues (V_α: 3–115, V_β: 5–116) are superimposed. Residues in β strands are colored red and residues in loops are colored gray. The V_α and V_β domains are labeled α and β in (A). CDR loops are labeled 1, 2 and 3 in (A). N and C termini are labeled in (B). The polypeptide linker connecting L116β and Q1α is not shown.

This is demonstrated by extending a model introduced by Mumenthaler et al. (1997). The chance of random assignment of a cross peak is related to both the probability that randomly selected atoms are near one another in space and the number of cross peaks within the chemical shift tolerance range $2*\Delta_{predict}$. Assuming protons are evenly distributed within a sphere of radius R that represents the protein, the probability, q, that two randomly selected atoms are closer to each other than d_{max} is approximately given by:

$$q = (d_{max}/R)^3 \tag{3}$$

From the initial structures, we approximate the D10 scTCR protein by a sphere of radius 35 Å. If the maximal distance for which an NOE may be observed, d_{max}, is 5 Å then $q = 0.003$. Approximately 60% of the NOE cross peaks input to the automated assignment procedure are in the region between 1 and 4 ppm. Assuming a uniform density of 800 NOE cross peaks between 1 and 4 ppm, 160 cross peaks will be in the chemical shift range $2*\Delta_{predict} = 0.6$ ppm. The probability that a proton will be assigned a chemical shift randomly is:

$$1 - (1 - q)^{160} = 0.38 \tag{4}$$

To reduce the rate of random assignment to an acceptable level, we only accept assignments that are duplicated within a small chemical shift tolerance ($2*\Delta_{assign} = 0.05$ ppm). The probability that two randomly selected protons are within 5 Å of another proton is approximately

$$q = ((d_{max}/R)^3)^2 = 8.5 \times 10^{-6} \tag{5}$$

Defining m as the average number of cross peaks within each 0.05 ppm interval and $n = \Delta_{predict}/\Delta_{assign}$, the probability that a proton will be assigned a chemical shift randomly is

$$[1 - ((1 - q)^{m*(m-1)/2})] * n = 0.008 \tag{6}$$

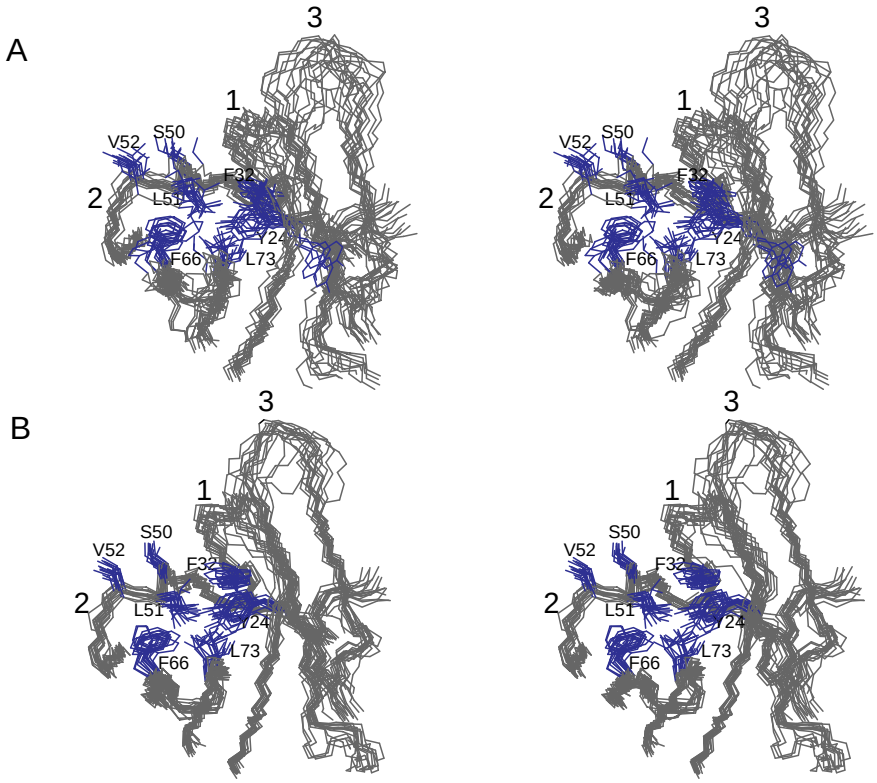

Figure 6. Stereoviews of the MHC/antigen-binding surface of the V_α domain. Ensembles of structures calculated without (A) and with (B) distance restraints derived from the automatically assigned NOEs are shown. Heavy atoms of V_α residues 3–91 are superimposed. Backbone heavy atoms are shown in gray and side chain heavy atoms of residues discussed in the text are shown in blue. The loop labeled DE connects the D and E β strands in the V_α domain.

The probability of random assignment for the automated procedure compares favorably to the error rate expected for manual assignment.

Manual assignment of the NOE data used by the automated assignment procedure would be difficult because of extensive spectral overlap and uncertainty in the sequence-specific assignments of side chains. Manual inspection of automatically assigned NOEs indicates that nearly all appear correct. Occasionally, misidentification of peak centers in overlapped spectral regions causes the automated procedure to make NOE assignments that appear inconsistent with previous NOE assignments. However, we are confident that the overall rate of error is no higher than if the NOEs had been assigned manually. Moreover, the use of over 1600 ‘hard’ NOEs during the refinement procedure and reduction in the value for the RSWitch and ASYMptote parameters in the structure calculations for the NOEs assigned by the automated procedure (see Methods) should prevent a single misassigned NOE from significantly affecting the calculated structures.

D10 scTCR structure refinement

Overlay of the backbones of low-energy structures calculated using the complete set of restraints (Figure 5B) and the original set of manually assigned NOEs (Figure 5A) reveals greater precision in backbone conformations in the structures calculated using the complete set of restraints. The MHC/antigen binding surface of the TCR is composed of loop regions (CDRs) in V_α and V_β, which are variable in sequence among different TCRs. The conformations of CDRs in both domains are better defined in structures calculated with the automatically assigned NOEs (Figure 5B and Table 3). A total of 76 NOEs, many of them long range, are assigned to CDR1 and CDR2 residues by the automated procedure. In contrast, long-range NOEs are only observed to residues near the ends of the CDR3 loops. As a result, the conformations of CDRs 1 and 2 are better defined than CDR3 in the refined structures.

Using the scTCR structure nearest the mean in the original ensemble to model the D10 TCR complex

Table 3. Average root mean square deviations (rmsd) from the mean structure for all heavy atoms of secondary structural elements[a]

	Preliminary structures		Refined structures	
	rmsd (Å)	Range (Å)	rmsd (Å)	Range (Å)
α domain				
A strand	1.81	1.18–2.43	1.53	0.99–2.91
AB loop	2.92	1.61–4.84	1.71	1.16–2.54
B strand	1.14	0.83–1.70	0.96	0.58–1.56
BC loop (CDR1)	2.80	2.11–3.83	1.68	1.20–2.35
C strand	1.72	1.02–3.27	1.23	0.76–3.41
CC′ loop	2.83	1.83–4.29	2.08	1.33–3.48
C′ strand	0.99	0.62–1.42	0.79	0.47–1.46
C′C″ loop (CDR2)	1.49	0.89–2.06	1.09	0.71–1.68
C″ strand	1.47	1.01–2.11	1.35	0.97–1.69
C″D loop	1.85	1.25–2.70	1.49	0.96–1.98
D strand	1.46	0.73–2.69	1.33	0.92–1.97
DE loop	2.68	1.60–4.77	1.60	1.15–2.37
E strand	1.45	0.90–2.10	0.93	0.64–1.57
EF loop	2.50	1.81–3.62	2.38	1.44–3.51
F strand	1.41	0.93–2.18	1.00	0.69–2.19
FG strand (CDR3)	2.71	1.76–4.76	2.08	1.11–4.17
G strand	1.78	1.18–2.80	1.67	1.05–2.84
β domain				
A strand	1.56	0.89–2.93	1.30	0.80–2.54
AB loop	1.45	0.76–2.42	1.19	0.49–2.34
B strand	1.13	0.75–1.98	0.95	0.59–1.67
BC loop (CDR1)	1.69	1.13–2.07	1.52	1.01–2.37
C strand	1.46	1.00–2.24	1.32	0.84–2.42
CC′ loop	2.69	1.89–3.81	2.37	1.54–3.70
C′ strand	1.36	0.75–2.03	1.22	0.76–1.93
C′C″ loop (CDR2)	1.61	1.07–2.21	1.53	1.09–2.38
C″ strand	3.03	2.42–3.71	2.89	2.25–3.45
C″D loop	2.09	1.50–2.81	1.99	1.37–2.95
D strand	1.87	1.17–3.87	1.22	0.85–2.04
DE loop	1.39	0.84–2.09	1.24	0.85–2.12
E strand	1.14	0.56–2.15	1.00	0.63–1.68
EF loop	1.87	1.16–3.35	1.44	0.90–2.32
F strand	1.21	0.78–1.68	1.09	0.71–1.53
FG loop (CDR3)	2.88	2.02–4.64	2.29	1.42–3.11
G strand	1.41	1.02–2.11	1.20	0.73–1.69

[a]Residues in secondary structural elements are defined in Table 2.

with MHC/antigen, CDR2α residues Ser$^{50\alpha}$, Leu$^{51\alpha}$ and Val$^{52\alpha}$ appear positioned to interact with mostly hydrophobic residues in an MHC helix (Hare et al., 1999). The average conformations of CDR2α in the original and refined structures are similar, but the side chains of CDR2α residues are better defined in the refined (Figure 6B) than in the original (Figure 6A) ensemble of structures. Although Val$^{52\alpha}$ is at the apex of a loop on the periphery of the TCR protein, its conformation is well defined in the refined structure.

Hydrophobic interactions that stabilize the conformation of CDR2α are revealed in the refined structures. A hydrophobic cluster is formed by the side chains of Tyr$^{24\alpha}$ and Phe$^{32\alpha}$ in CDR1, Leu$^{51\alpha}$ in CDR2 and Phe$^{66\alpha}$ and Leu$^{73\alpha}$ in the loop connecting the D and E β strands. Many NOEs in this region are assigned by the automated procedure, including NOEs between the $^{1}H^{\delta}$ of Phe$^{32\alpha}$ in CDR1 and $^{1}H^{N}$ of Ile$^{49\alpha}$ (CDR2), Leu$^{51\alpha}$ (CDR2) and Ala$^{92\alpha}$.

The well-defined CDR2α conformation is consistent with data indicating low backbone mobility on picosecond timescales for CDR2α (Hare et al., 1999) and suggests a 'preformed' binding site on CDR2α for MHC. This is consistent with interaction between CDR2α and mostly conserved MHC residues in a TCR/MHC/antigen complex (Garboczi et al., 1996). Furthermore, it is consistent with the absence of conformational change observed in CDR2α of a TCR upon binding an MHC/antigen complex (Garcia et al., 1998).

Conclusions

An automated method for assigning NOESY cross peaks in a protein with known preliminary structures and partial chemical shift assignments is described and applied to the 27.8 kDa D10 scTCR protein. The method uses predicted chemical shifts to complete a chemical shift assignment table that is input to an automated routine for combined NOESY cross peak assignment and structure calculation. The method facilitates rapid determination of the chemical shifts of side chains and assignments of NOESY cross peaks. We automatically assigned over 30% of 1327 previously unassigned NOESY cross peaks from the scTCR protein. The new NOEs are particularly helpful in defining side chains in the NMR structures. Further work is required to determine how dependent the method is on the completeness of sequence-specific assignments and to assess the accuracy of the automatically assigned NOEs. We believe the method will be applicable to a variety of NOESY spectra, thereby speeding up the structure determination process by NMR.

Data deposition

Molecular coordinates for the refined D10 scTCR structures have been deposited with the Protein Data Bank (accession number 1bwm). New ^{1}H assignments are contained in a deposition with the BioMagResBank (accession number 4330).

Acknowledgements

This work was supported by the Medical Foundation and by the Myasthenia Gravis Foundation of America (fellowships to B.J.H.) and grants from the National Institutes of Health (AI/CA 37581) and the National Science Foundation (MCB 9527181) to G.W. We thank Dr. S. Hyberts for assistance with computations and Dr. Y. Aubin for helpful discussions.

References

Brünger, A.T. (1993) *XPLOR version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. and Wiley, D.C. (1996) *Nature*, **384**, 134–141.

Garcia, K.G., Degano, M., Pease, L.R., Huang, M., Peterson, P.A., Teyton, L. and Wilson, I.A. (1998) *Science*, **279**, 1166–1172.

Goll, C.M., Pastore, A. and Nilges, M. (1998) *Structure*, **6**, 1291–1302.

Hare, B.J., Wyss, D.F., Osburne, M.S., Kern, P.S., Reinherz, E.L. and Wagner, G. (1999) *Nat. Struct. Biol.*, **6**, 574–581.

Kraulis, P.J. (1994) *J. Mol. Biol.*, **243**, 696–718.

Laskowski, R.A., MacArthur, M.W., Moss, D.W. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Mumenthaler, Ch. and Braun, W. (1995) *J. Mol. Biol.*, **254**, 465–480.

Mumenthaler, Ch., Güntert, P., Braun, W. and Wüthrich, K. (1997) *J. Biomol. NMR*, **10**, 351–362.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **239**, 129–136.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

Oldfield, E. (1995) *J. Biomol. NMR*, **5**, 217–225.

Ösapay, K. and Case, D.A. (1991) *J. Am. Chem. Soc.*, **113**, 9436–9444.

Oshiro, C.M. and Kuntz, I.D. (1993) *Biopolymers*, **33**, 107–115.

Schäfer, N. (1992) Diploma Thesis, ETH, Zürich.

Sitkoff, D.F. and Case, D.A. (1997) *J. Am. Chem. Soc.*, **119**, 12262–12273.

Williamson, M.P. and Asakura, T. (1993) *J. Magn. Reson.*, **B101**, 63–71.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.

Wishart, D.S., Watson, M.S., Boyko, R.F. and Sykes, B.D. (1997) *J. Biomol. NMR*, **10**, 329–336.