Journal of Biomolecular NMR, **15:** 315–330, 1999.
KLUWER/ESCOM

Maximum entropy approach to the determination of solution conformation of flexible polypeptides by global conformational analysis and NMR spectroscopy – Application to DNS^1-c-[D-A$_2$bu^2, Trp^4,Leu^5]-enkephalin and DNS^1-c-[D-A$_2$bu^2, Trp^4, D-Leu^5]enkephalin

M. Groth, J. Malicka, C. Czaplewski, S. Ołdziej, L. Łankiewicz, W. Wiczk & A. Liwo*
Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

Received 7 June 1999; Accepted 1 November 1999

Key words: conformational equilibrium, empirical force fields, maximum entropy method, molecular dynamics, Monte Carlo methods, opioid peptides

Abstract

A method is proposed to determine the conformational equilibrium of flexible polypeptides in solution, using the data provided by NMR spectroscopy and theoretical conformational calculations. The algorithm consists of the following three steps: (i) search of the conformational space in order to find conformations with reasonably low energy; (ii) simulation of the NOE spectrum and vicinal coupling constants for each of the low energy conformations; and (iii) determining the statistical weights of the conformations, by means of the maximum-entropy method, in order to obtain the best fit of the averaged NOE intensities and coupling constants to the experimental quantities. The method has been applied to two cyclic enkephalin analogs: DNS^1-c-[D-A$_2$bu^2,Trp^4,Leu^5]enkephalin (ENKL) and DNS^1-c-[D-A$_2$bu^2,Trp^4,D-Leu^5]enkephalin (ENKD). NMR measurements were carried out in deuterated dimethyl sulfoxide. Two techniques were used in conformational search: the electrostatically driven Monte Carlo method (EDMC), which results in extensive search of the conformational space, but gives only energy minima, and the molecular dynamics method (MD), which results in a more accurate, but also more confined search. In the case of EDMC calculations, conformational energy was evaluated using the ECEPP/3 force field augmented with the SRFOPT solvation-shell model, while in the case of MD the AMBER force field was used with explicit solvent molecules. Both searches and subsequent fitting of conformational weights to NMR data resulted in similar conformations of the cyclic part of the peptides studied. For both ENKL and ENKD a common feature of the low-energy solution conformations is the presence of a type II′ or type IV β-turn at residues 3 and 4; the ECEPP/3 force field also gives a remarkable content of type III β-turn. These β-turns are tighter in the case of ENKL, which is reflected in different distributions of the D-A$_2$bu(N$^\gamma$H)$\cdots$D-A$_2$bu(CO) and D-A$_2$bu(N$^\gamma$H)$\cdots$Gly^3(CO) hydrogen-bonding distances, indicating that the D-A$_2$bu(N$^\gamma$H) amide proton is more shielded from the solvent than in the case of ENKD. This finding conforms with the results of temperature coefficient data of the D-A$_2$bu(N$^\gamma$H) proton. It has also been found that direct (MD) or Boltzmann (EDMC) averages of the observables do not exactly conform with the measured values, even when explicit solvent molecules are included. This suggests that improving force-field parameters might be necessary in order to obtain reliable conformational ensembles in computer simulations, without the aid of experimental data.

Abbreviations: A$_2$bu, α,γ-diaminobutyric acid; DNS, 5-dimethylamino-naphthalene-1-sulfonyl (dansyl); ECEPP, Empirical Conformational Energy Program for Peptides and Proteins; EDMC, Electrostatically Driven Monte Carlo Method; MD, Molecular Dynamics; NOE, Nuclear Overhauser Effect.

Introduction

Nuclear magnetic resonance and particularly nuclear Overhauser effect (NOE) spectroscopy are useful tools for conformational studies of peptides and proteins (Wüthrich, 1986; Wagner, 1990; Yang et al., 1993). A usual procedure is to convert the NOE intensities into interproton distances and implement the latter in molecular dynamics (MD) simulations as distance restraints (Wüthrich, 1986; Wagner, 1990; Yang et al.,

*To whom correspondence should be addressed. E-mail: adam@chemik.chem.univ.gda.pl

1993). While such a procedure is justifiable in the case of proteins, which occur in a well-defined conformation, its application to flexible polypeptides that occur in a multiplicity of conformations is not straightforward. In the last case the experimental observables should rather be regarded as conformational averages (Brüschweiler et al., 1991; Blackledge et al., 1993). According to this principle, a number of procedures have been proposed. Most of them implement the experimental information already during conformational search as time-averaged (Torda et al., 1989, 1990, 1993; Bonvin et al., 1991, 1994) or ensemble-averaged (Brüschweiler et al., 1991; Scheek et al., 1991; Blackledge et al., 1993; Kemmink et al., 1993; Gippert et al., 1998) restraints. Other procedures use the experimental information after the search, by determining the statistical weights of the calculated low-energy conformations, in order to obtain average interproton distances, vicinal coupling constants, and temperature coefficients of amide protons that fit best to the experimental data (Shenderovich et al., 1988; Nikiforovich et al., 1993; Pearlman, 1996). In this work we have followed and further developed the second approach. We have implemented the electrostatically driven Monte Carlo (EDMC) (Ripoll and Scheraga, 1988, 1989, 1990, 1998) and MD methods in the conformational search and devised a fitting procedure that is based on the maximum entropy principle; this gives a reasonable fit to the data and, at the same time, enables as many conformations as possible to enter the statistical ensemble. The approach has been tested on two cyclic enkephalin analogs of the following amino acid sequence:

DNS–c–[D-A_2bu–Gly–Trp–Leu–] (ENKL)
DNS–c–[D-A_2bu–Gly–Trp–D-Leu–] (ENKD)

where DNS denotes the 5-dimethylamino-naphthalene-1-sulfonyl (dansyl) end group and A_2bu denotes the α,γ-diaminobutyric acid residue.

Methods

Synthesis

ENKL and ENKD were synthesized using the solid-phase method on p-alcoxybenzyl alcohol resin. The substrates were fluorene-9-yl-methoxycarbonyl (Fmoc) derivatives of amino acids, while diisopropylocarbodiimide (DIPCI) and 1-hydroxybenzotriazole (HOBt), or 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were used as coupling reagents (Fields and Noble, 1990).

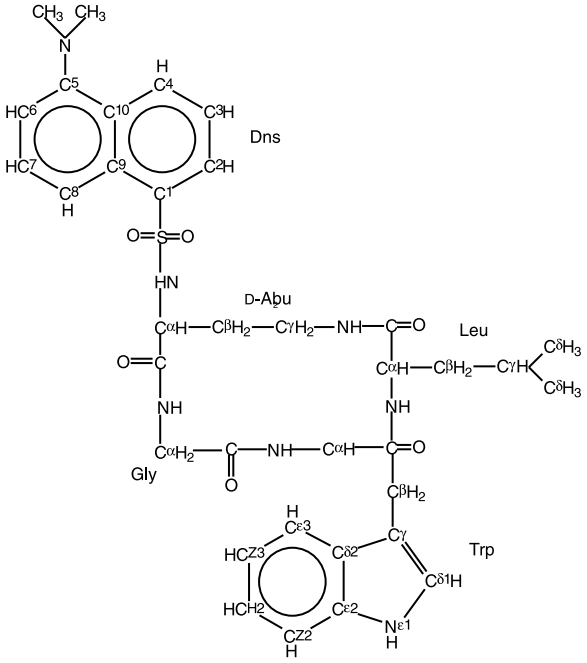

Figure 1. Structural diagram of ENKL and ENKD with proton labels.

The coupling reagents were purchased from Fluka AG. Dansyl hydrochloride was purchased from Aldrich. The reagents were used without additional purification. Fmoc-Leu, Fmoc-Trp, and Fmoc-Gly were purchased from Novabiochem and used without purification. Fmoc-D-Leu was synthesized following the procedure of Bodanszky and Bodanszky (1984). Fmoc-D-A_2bu(Boc)-OH was synthesized as described below. Cyclization was performed using 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethylammonium tetrafluoroborate (TBTU) (Schmidt and Neubert, 1991).

Preparation of Fmoc-D-A_2bu(Boc)-OH

Fmoc-D-Gln-OH was synthesized first from D-Gln and FmocCl, using a standard procedure (Bodanszky and Bodanszky, 1984). Then the γ-carboxyamido group of glutamine was converted into the amino group using the procedure of Waki et al. (1981). The γ-amino group was subsequently protected with Boc using an appropriate procedure (Bodanszky and Bodanszky, 1984).

Purification of the peptides

The obtained peptides were purified by means of the preparative RP-HPLC method. The moving phase consisted of an isocrat comprising 31% and 33% CH_3CN in water for ENKL and ENKD, respectively, and 0.1% TFA. The purity of the peptides was assessed by means of the analytical RP-HPLC method. The molecular constitution of the peptides was confirmed using FAB-MS and ^{1}H NMR COSY spectroscopy.

NMR measurements

Proton NMR spectra were recorded on a 499.89 MHz VARIAN spectrometer at the Interuniversity Nuclear Magnetic Resonance Laboratory at the Technical University of Gdańsk. The experiments were carried out in deuterated dimethyl sulfoxide (DMSO-d_6). Temperature coefficients of the amide resonances were calculated from the spectra obtained at four temperatures: 22, 30, 40, and 50 °C. Two-dimensional ^{1}H–^{1}H COSY and ^{1}H–^{1}H NOESY spectra were recorded at 22 °C, the mixing time τ_m being 0.30 s.

The spin systems of the amino acids were identified based on the position and shape of the signals of leucine methyl protons ($C^{\delta}H_3$) and glycine methylene protons ($C^{\alpha}H_2$). The latter were coupled with the glycine amide proton only, while the characteristic signals of leucine methyl protons served to determine the positions of the remaining aliphatic protons and the amide protons of this residue. The resonance peaks of α,γ-diaminobutyric acid were unambiguously assigned by analysis of the magnetization transfer from the amide ($N^{\alpha}H$ and $N^{\gamma}H$) through the $C^{\alpha}H$ and $C^{\gamma}H_2$ protons to the $C^{\beta}H_2$ protons. The remaining signals within the absorption range of aliphatic protons were assigned to the $C^{\alpha}H$ and $C^{\beta}H_2$ protons of tryptophan. The part of the NMR spectrum that contained tryptophan and dansyl proton signals required especially careful analysis, because of their partial overlap. The analysis of the coupling of the indole moiety was started from the low-field ($\delta \approx 10.8$ ppm) signal of the indole imide proton, which was coupled to the $C^{\delta 1}H$ proton only. The remaining peaks were assigned based on the well-known order of the signals of the $C^{Z1}H$, $C^{H2}H$, $C^{Z3}H$, and $C^{\epsilon 3}H$ indole-ring protons of tryptophan (see Figure 1 for proton labels). The ^{1}H NMR spectrum of dansyloglycine was used to assign the peaks coming from the dansyl protons. The ^{1}H NMR spectra of ENKL and ENKD, together with signal assignments are shown in Figure 2.

The $^3J_{C^{\alpha}HN}$ coupling constants were determined from ^{1}H NMR spectra by measuring the distance between the multiplet lines of the H^{α} protons. The estimated experimental error was 0.1 Hz.

Conformational calculations

Force fields

In the present study the ECEPP/3 (Nemethy et al., 1992) and AMBER (Weiner et al., 1986) force fields were used. In both cases the total conformational energy, E_{tot}, can generally be expressed as the sum of the bond energy, E_b, bond angle energy, E_θ, electrostatic energy, E_{es}, nonbonded energy, E_{nb}, and torsional energy, E_{tor}, as given by Equation 1:

$$\begin{aligned} E_{tot} = & \sum_{\text{bonds}} k_i^d (d_i - d_i^\circ)^2 + \sum_{\text{bond angles}} k_i^\theta (\theta_i - \theta_i^\circ)^2 \\ & + \sum_{i<j} \left[\frac{q_i q_j}{\epsilon r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \\ & + \sum_{\text{H-bonded pairs}} \left[\frac{q_i q_j}{\epsilon r_{ij}} + \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] \\ & + \sum_{\text{torsional angles}} \sum_n V_{ni} [1 + \cos(n\phi_i - \gamma_{ni})] \end{aligned} \quad (1)$$

where d_i, d_i°, and k_i^d are the length of the ith bond, the 'strainless' length, and the force constant, respectively, θ_i, θ_i°, and k_i^θ are the value of the ith bond angle, the 'strainless' value, and the force constants, respectively, r_{ij} is the distance between atoms i and j, A_{ij}, B_{ij}, C_{ij}, and D_{ij} are pair-specific constants in the nonbonded potentials, ϵ is the relative dielectric permittivity, q_i is the partial charge of atom i, ϕ_i is the ith torsional angle, n is the multiplicity of a torsional energy term, V_{ni} is the torsional constant of multiplicity n characteristic of the ith angle and γ_{ni} is the respective phase angle.

In the case of the ECEPP/3 force field, which assumes rigid valence geometry, the first two terms are not present except in the cases of cyclic peptides, where they have to be considered when closing intrachain loops [e.g. the N^{γ}(D-A_2bu)–C′(Leu) amide bridge in this study].

In the case of the calculations with the AMBER force field the solvent was considered at microscopic level (see below), while in the case of the ECEPP/3 force field the solvation energy (E_{solv}) was evaluated in the SRFOPT solvent-accessible surface model (Vila et al., 1991), whose parameters pertain to solvation by water. However, because of the comparatively high

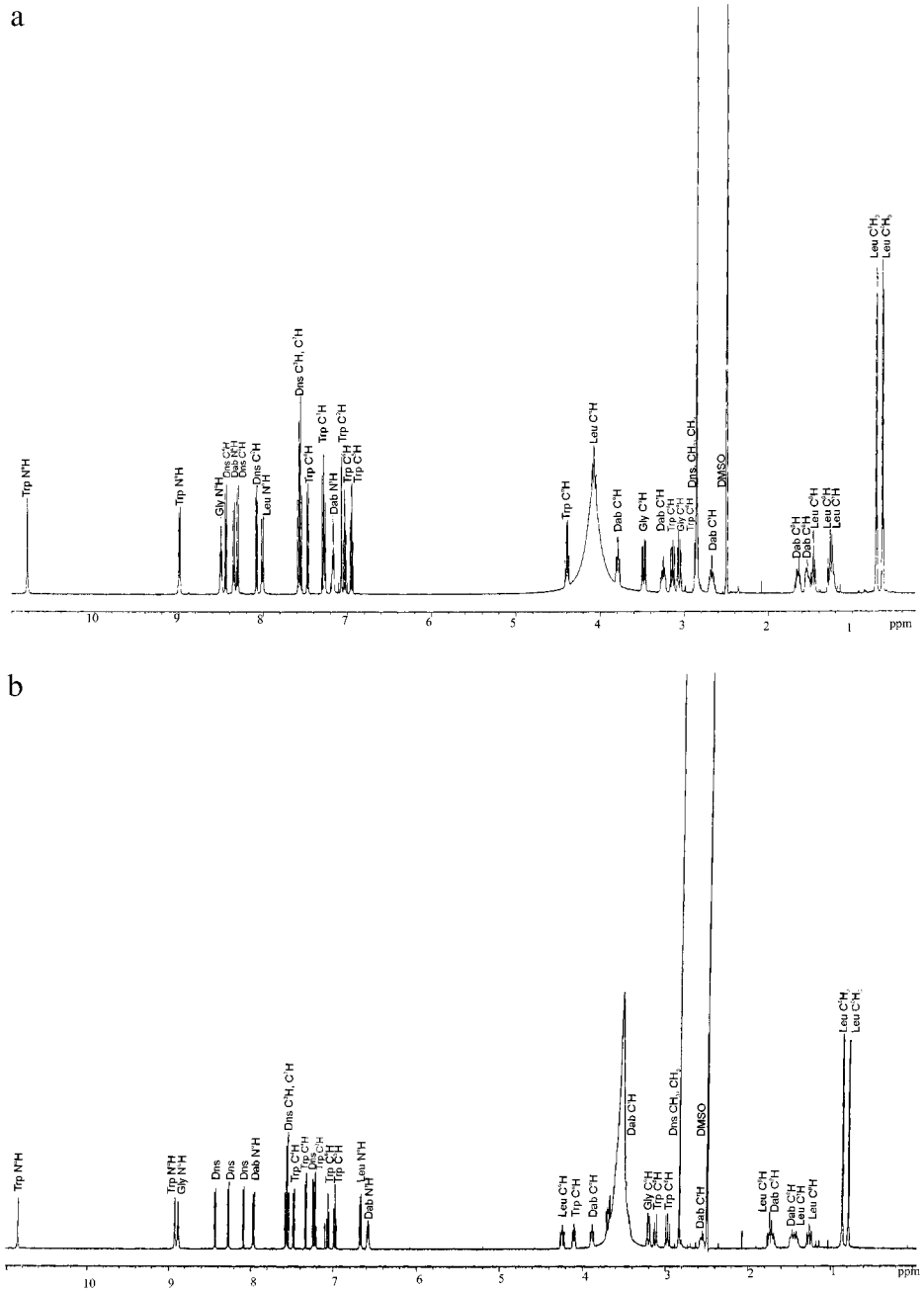

Figure 2. The ^{1}H NMR spectra of ENKL (a) and ENKD (b). See Figure 1 for the proton labels.

dielectric constant and basicity of DMSO (Reichardt, 1988), we considered it more appropriate to use this solvation model than not introducing solvation at all. In our earlier study of oxytocin and vasopressin we found that using SRFOPT results in much better agreement of the conformationally averaged interproton distances with those obtained from NOE data than in the case of in vacuo calculations (Liwo et al., 1996). In the SRFOPT model the solvation energy is expressed by Equation 2:

$$E_{solv} = \sum_i \sigma_i A_i \tag{2}$$

where σ_i is the solvation energy of the ith atom per unit surface area and A_i is the solvent-accessible surface area of the ith atom exposed to the solvent.

Force field parameters for the DNS end group

The valence geometry of the DNS N-terminal end group, not present in the standard ECEPP/3 database, was taken from appropriate data from the Cambridge database of crystal structures of small organic molecules (Allen et al., 1983). Because the ECEPP/3 force field uses rigid valence geometry, there was no need to determine bond or bond angle constants. For the AMBER force field, the bond and bond angle constants were assigned by extrapolating data from related functional groups present either in the AMBER (Weiner et al., 1986) or the MMX (Gajewski et al., 1990) force fields. The partial atomic charges were assigned the Mulliken population charges calculated with the PM3 semiempirical method (Steward, 1989) (ECEPP/3) or fitted using the RESP algorithm (Bayly et al., 1993) to reproduce the 6-31G** molecular electrostatic potential of the DNS–NH–CH_3 model compound (AMBER). The charges are shown in Figure 3. Torsional constants were calculated from the 6-31G** energy curves of the rotation about appropriate bonds of DNS–NH–CH_3. All non-standard parameters of the DNS end group are summarized in Table 1. Ab initio calculations were carried out using the program GAMESS (Schmidt et al., 1993).

EDMC calculations

Global conformational search of the peptides studied was carried out using the electrostatically driven Monte Carlo method (EDMC) (Ripoll and Scheraga, 1988, 1989, 1990, 1998) with the ECEPP/3+SRFOPT force field. A total of about 2000 energy-minimized conformations were generated for ENKL and ENKD, respectively. The working temperature[1] was 1000 K. The resulting conformations were subsequently subjected to a cluster analysis, using the minimum-variance algorithm (Späth, 1980). The root mean square (rms) deviation between heavy atoms at optimal superposition was taken as a measure of the distance between conformations, and a cut-off value of 0.2 Å was used to separate the families. For ENKL and ENKD, 397 and 526 families of conformations, respectively, were obtained.

[1] As opposed to canonical Monte Carlo sampling of the energy landscape, in the EDMC method the temperature is a purely abstract parameter which only defines the probability of accepting energy minima with a higher energy than the current one.

Table 1. Non-standard bond, bond angle (AMBER), and torsional (ECEPP/3 and AMBER) parameters of the DNS end group. See Figure 3 for atom types

Bond	k_d (kcal/mol × Å^2)	d° (Å)	
CA–S* [a]	250.0	1.789	
S*–O[a]	525.0	1.435	
S*–N[a]	230.0	1.618	
Bond angle	k_θ (kcal/mol × rad^2)	θ° (deg)	
N2–CA–CA[b]	70.0	120.0	
CT–N2–CT[b]	50.0	118.0	
CA–CA–S*[b]	60.0	120.0	
CA-S*–O[a]	60.0	109.5	
O–S*–O[a]	140.0	109.5	
CA–S*–N[a]	45.0	109.5	
O–S*–N[a]	45.0	109.5	
S*–N–H[a]	130.0	114.0	
S*–N–CT[a]	50.0	119.0	
Dihedral angle	V_n (kcal/mol)	γ	n
ECEPP/3			
CA–S*–N–CT[c]	2.5	0.0	2
CA–CA–S*–N[c]	0.0	180.0	2
CA–CA–N–CT[b]	6.8	180.0	2
AMBER			
CA–CA–S*–O[c]	0.0	180.0	2
CA–CA–S*–N[c]	0.0	180.0	2
CA–S*–N–CT[c]	1.3	0.0	1
CA–S*–N–CT[c]	2.3	0.0	2
CA–S*–N–CT[c]	1.0	0.0	3
O–S*–N–CT[c]	1.0	180.0	2
CA–S*–N–H[c]	0.0	180.0	2
O–S*–N–H[c]	0.0	180.0	2
CA–CA–N–CT[b]	6.8	180.0	2
Improper dihedral angle	V_n (kcal/mol)	γ	n
AMBER			
S*–CT–N–H[b]	0.0	180.0	2
CT–CT–N2–CA[b]	1.0	180.0	2
CA–CA–CA–CA[b]	1.0	180.0	2
CA–CA–CA–S*[b]	1.0	180.0	2
N2–CA–CA–CA[b]	1.0	180.0	2

[a] Adapted from the MMX force field (Gajewski et al., 1990).
[b] Extrapolated from AMBER parameters pertaining to similar atom types.
[c] Determined in this work.

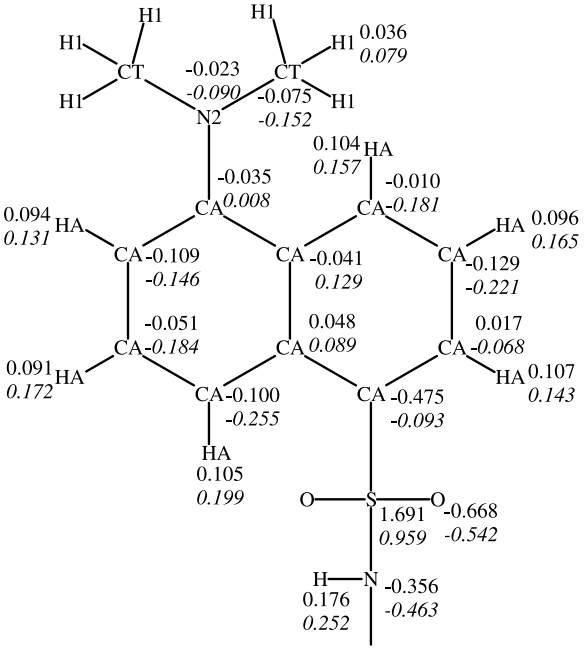

Figure 3. ECEPP (upper values) and AMBER (lower values in italics) charges of the DNS end group.

Molecular dynamics simulations

Molecular dynamics simulations were carried out with the AMBER force field (Weiner et al., 1986) using the AMBER 4.1 package (Pearlman et al., 1995). Explicit DMSO molecules with methyl groups considered as extended atoms were used. The corresponding bond, bond angle and vdW parameters, as well as partial atomic charges were taken from the literature (Liu et al., 1995). The initial solvent configuration around the peptide was obtained by filling a cubic box with DMSO molecules subject to the condition of non-overlap. The shortest distance of peptide atoms from the box boundary was 15 Å, corresponding to about 2500 DMSO molecules in a box. The simulations were carried out at 298 K, in a periodic box, with applying the minimum-image convention. A spherical cut-off with 9.0 Å radius was applied to nonbonded and electrostatic interactions. The integration step was 2 fs. The initial 500 ps were carried out in the N, P, T scheme, until solvent density was close to 1.095 g/cm^3, a value characteristic of liquid DMSO (Riddick et al., 1986). Then the simulations were carried out in the N, V, T scheme. The total duration of each run was 7.5 ns. The conformations obtained in the last 4 ns of simulations (after this period of time the energy of the system was reasonably established) were collected each 200th step and the resulting set of conformations was subjected to a minimum-variance cluster analysis, as described in the preceding section. For ENKL and ENKD, 600 and 900 families of conformations, respectively, were obtained.

Calculation of the statistical weights of the conformations by fitting the theoretical to the experimental NMR data

In this study we used the MORASS program (Post et al., 1990; Meadows et al., 1994) to compute theoretical NOE integral intensities. This program solves the system of Bloch differential equations for the cross-relaxation of a system of interacting proton spins. The correlation time τ_c was set at 0.45 ms, based on data for other small cyclic peptides (Bhaskaran et al., 1992; Yu et al., 1992).

The theoretical NOE intensities are averages over all conformations of the ensemble:

$$\overline{v_{kl}} = V_\circ \sum_{i=1}^{NC} x_i v_{ikl} \quad k, l = 1, 2, \ldots NP \tag{3}$$

$$x_i \geq 0, \quad i = 1, 2, \ldots NC \tag{4}$$

$$\sum_{i=1}^{NC} x_i = 1 \tag{5}$$

where $\overline{v_{kl}}$ is the integral intensity of the NOE between protons k and l averaged over all conformations, v_{ikl} is this intensity for conformation i, x_i is the statistical weight (fraction) of the ith conformation, $V_\circ$ is a scaling factor, and NP and NC are the number of protons and the number of conformations, respectively.

The vicinal NH–C$^\alpha$H coupling constants corresponding to the ith conformation can be calculated from the empirical Karplus relationship [Equation 6].

$$J_{ik} = a_{\circ k} + a_{1k} \cos\theta_{ik} + a_{2k} \cos^2\theta_{ik} \tag{6}$$

where J_{ik} is the coupling constant of the kth angle and the ith conformation and θ_{ik} is the corresponding angle. For the C$^\alpha$H–NH coupling constants of non-glycine residues $\theta = \phi \mp 60°$ for the L- and D-residues, respectively and for the C$^\alpha$H$_2$–NH coupling constant of glycine and the C$^\gamma$H$_2$–N$^\gamma$H coupling constant of D-A$_2$bu, $\theta = \phi$. The values of the constants $a_\circ$, a_1, and a_2 were taken from the literature (Bystrov, 1976) and were 0.40, −1.1, and 9.4 for non-glycine residues and 14.9, −1.1, and −9.4 for glycine and the D-A$_2$bu side chain, respectively. We chose to use these constants and not those determined later by Pardi et al. (1984), because they were determined using peptide data, while the latter were determined using protein

data and are therefore less relevant to the case studied in our work. In fact, trial runs with the constants determined by Pardi et al. showed a much poorer fit of the coupling constants to the experimental data.

As in the case of NOE intensities, the coupling constants must be averaged over conformations:

$$\overline{J_k} = \sum_{i=1}^{NC} x_i J_{ik} \tag{7}$$

Thus, the average NOE intensities and the average coupling constants are functions of the weights $x_1, x_2, \ldots, x_{NC}$. The weights could therefore be determined by least-squares fitting of the calculated NOE intensities and coupling constants to the corresponding experimental quantities, as given by Equation 8:

$$\begin{aligned}
&\min \Phi(V_\circ, x_1, x_2, \ldots, x_{NC}, a_{\circ 1}, a_{11}, a_{12}, \ldots, a_{NJ}) = \\
&\sum_{(k,l)\in\mathcal{K}} w_{kl}[v_{kl}^{exp} - \overline{v_{kl}}(V_\circ, x_1, x_2, \ldots x_{NC})]^2 \\
&+ w_J \sum_{k=1}^{N\theta}[J_k^{exp} - \overline{J_k}(x_1, x_2, \ldots x_{NC})]^2 \\
&+ \sum_{I=1}^{NJ} \frac{1}{\sigma_{a_{\circ I}}^2}(a_{\circ I} - a_{\circ I}^\circ)^2 + \frac{1}{\sigma_{a_{1I}}^2}(a_{1I} - a_{1I}^\circ)^2 \\
&+ \frac{1}{\sigma_{a_{2I}}^2}(a_{2I} - a_{2I}^\circ)^2
\end{aligned} \tag{8}$$

where $\mathcal{K}$ is the set of all signals considered, w_{kl} is the weight of the intensity of the NOE between protons k and l, w_J is the weight of the coupling-constant term, $N\theta$ is the number of angles for which the coupling constants were determined, NJ is the number of the sets of the constants in the Karplus equation, a_{kI}° denotes the 'standard' value of a_{kI} in the Karplus equation, $\sigma_{a_{kI}}$ is its estimated standard deviation. Including the last sum accounts for the fact that the values of the coefficients in Equation 6 are uncertain within the limits determined by their standard deviations. This does not increase the number of degrees of freedom, because the increase of the number of parameters is accompanied by the same increase of the number of terms in the minimized sum. In this study the last sum consisted of six terms: three for non-glycine-type and three for glycine-type coupling constants (cf. Equation 6). We assumed $\sigma_{a_\circ} = \sigma_{a_1} = \sigma_{a_2} = 2$ Hz and $w_J = 0.1$ Hz^{-1}; the latter value provided comparable magnitude of the NOE and coupling-constant term. The values of $\sigma_{a_\circ}$, σ_{a_1}, and σ_{a_2} were assigned based on the differences of the parameters of the Karplus equation reported by Bystrov for amino acids and small peptides (Bystrov, 1976) and by Pardi for BPTI (Pardi et al., 1984).

The sets $\mathcal{K}$ of interproton NOEs consisted of all observed off-diagonal signals except those coming from geminal protons (e.g. the protons of methylene groups). Equivalent protons (such as the protons of the methyl groups) were grouped together and the sums of NOEs coming from the whole groups were considered as single signals. Signals not observed in the spectra and resulting from pairs of protons belonging to different residues were considered as anti-NOEs and the experimental intensities in Equation 8 were set to zero. This implied that the conformations with close contacts between such protons had small statistical weights. The weights of the differences in the NOE intensities were calculated as follows:

$$w_{kl} = \frac{1}{v_{kl}^{exp} + a} \tag{9}$$

where a is a constant. A similar formula for weights was applied by Bonvin et al. (1991, 1994). When $a \gg v_{kl}$ all weights are nearly equal and therefore signal intensities are fitted. For small a, the NOE part of Φ becomes the sum of the squares of the relative errors in signal intensities and therefore the spectrum to be fitted gradually becomes a binary one, each entry indicating whether there is a NOE or not between a given pair of protons (depending on whether it is a NOE or an anti-NOE). According to the least-squares method w_{kl} should be proportional to the inverse of the square of the estimated standard deviation of the intensity of the corresponding signal: $w_{kl} = b/\sigma_{kl}^2$. Therefore, in order to estimate a we registered nine spectra of ENKD and calculated the average values and variances of the intensities of 27 representative signals (chosen to cover the whole intensity range). Then we fitted the variances to Equation 10, obtaining a equal to $2\overline{v}$, $\overline{v}$ being the average NOE intensity, with the relative standard deviation of 1.

$$\sigma_{kl}^2 = b(v_{kl} + a) \tag{10}$$

Based on this, a was set at the average NOE intensity, which provided a reasonable compromise between exact and binary-pattern fitting. This value can be considered as the lower bound of a determined by fitting the variance to Equation 10.

The measurements carried out to estimate a in Equation 9 were also used to estimate the weighted standard deviation in NOE intensity, which was found to be 0.019.

The Φ of Equation 8 arises from the maximum-likelihood principle of maximizing the joint probability distribution function of all observables (Carroll and Ruppert, 1988). If the errors obey Gaussian distributions, this results in minimization of a weighted sum of squares, the weights being equal to the inverses of the squares of the estimated standard deviations of the measured quantities (Carroll and Ruppert, 1988).

Minimization of Φ of Equation 8 usually results in the predominance of only a few conformations, while the weights of the remaining ones are close to zero. This cannot be considered reasonable in the case of ensembles obtained from MD simulations, which contain many very similar conformations. However, such a result is understandable in terms of the principles of least-squares fitting. Assume that there are two conformations, *a* and *b*, and conformation *a* fits slightly better to the experimental data than does conformation *b*. Even if the difference in fitting is very small, the least-squares procedure will result effectively in a weight of 1 for conformation *a* and 0 for conformation *b*. If the difference in fitting is not significant in terms of the experimental error, it is, however, reasonable to consider that the weights of both conformations are nearly equal. In order to prevent overfitting, we have implemented the maximum entropy approach (Gull, 1988; Daniell, 1991). There are a number of maximum-entropy algorithms (Livesey and Brochon, 1987; Gull, 1988; Brochon et al., 1990; Lieu and Hicks, 1994; Lyon et al., 1997) and in this study we have applied the simplest one, in which the 'entropy' term, $-\alpha \sum_{i=1}^{NC} x_i \log x_i$ is subtracted from the minimized sum of squares [Equation 8]. The resulting functional is expressed by Equation 11:

$$\Psi(V_\circ, x_1, x_2, \ldots, x_{NC}) = \Phi(V_\circ, x_1, x_2, \ldots, x_{NC}) + \alpha \sum_{i=1}^{NC} x_i \log x_i \qquad (11)$$

The entropy term reaches its global minimum if the statistical weights of all conformations are equal. This can be regarded as the reference state, in which no information about the preference of individual conformations is provided. Weight differentiating comes only from the Φ term that includes experimental information. Therefore a common procedure is to choose the coefficient at the entropy term, α, so that the weighted χ^2 value be equal to the number of observations (Livesey and Brochon, 1987; Brochon et al., 1990; Daniell, 1991), which is equivalent to the requirement that the mean errors in the fitted quantities

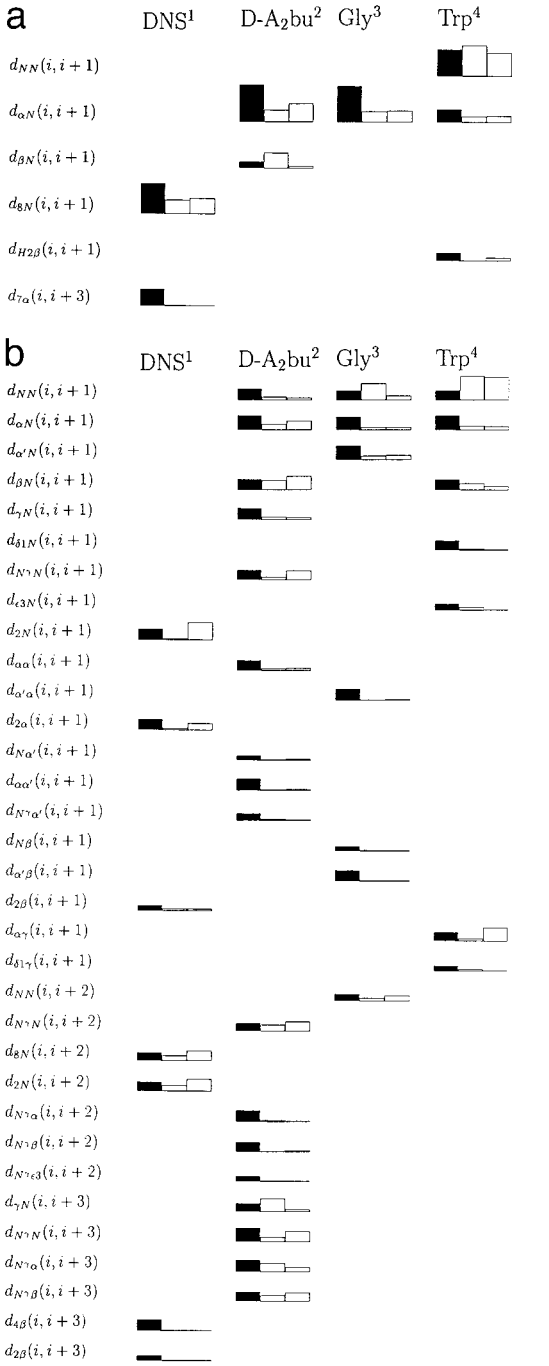

Figure 4. Weighted [with weights of Equation 9] volumes of observed signals of pairs of protons belonging to different amino acid residues in 2D NOE spectra of ENKL (a) and ENKD (b) (represented as the heights of the leftmost black bars), together with values obtained by fitting of EDMC (middle bars) and MD (rightmost bars) ensemble. See Figure 1 for proton labels. Each row corresponds to a fixed relative position of protons in the amino acid sequence [$d(i, i + 1)$ for protons of neighboring residues, etc.] and each column corresponds to the residue containing the first proton of a pair.

be comparable with the error estimates. In data analysis it is a natural approach, because the expected 'misfit' measures should be equal to the estimated experimental inaccuracy; if the agreement between theory and experiment is better, one starts fitting the noise.

Another approach to avoid overfitting was proposed earlier by Nikiforovich et al. (1993). Instead of determining one set of conformational weights, they used a Monte Carlo procedure to generate a *distribution* of weights, so that the calculated distribution of observables (interproton distances and coupling constant) approached the distribution of experimental quantities. However, this approach can be applied only when the number of conformations in the ensemble is limited (14 in their case); otherwise a reasonably dense distribution of weights is effectively impossible to generate (Nikiforovich et al., 1993). In our view, given the uncertainties inherent in empirical force fields it is advisable to use as large an ensemble as possible and therefore the maximum-entropy approach is preferred. Pearlman (1996) minimized a target function composed of appropriately weighted potential energy, distance-restraint violation, coupling-constant violation, and a potential energy dispersion term. The energy dispersion term should to some extent work towards equating the weights.

Minimization of Ψ was carried out using the Secant Unconstrained Minimization Solver (SUMSL) routine (Gay, 1983). Minimization of Φ (which is a sum of squares) was carried out using the Marquardt method (Marquardt, 1963). The variables were common logarithms of unnormalized weights, which also comprised the factor $V_\circ$ of Equation 3. Although both the Marquardt method and SUMSL are local minimizers, we found the same minima, independent of starting points, which suggests that Ψ is either unimodal or the alternative minima are very shallow and robust local minimizers can jump over them to find the global minimum.

Results and discussion

Determination of conformational equilibrium

Because EDMC is a fast and thorough method to search the conformational space, these calculations were carried out and their results processed first. With $\alpha = 0$ [no entropy term in Equation 11], the average deviation in the coupling constant for ENKD was only 0.88 Hz (Table 2), which is lower than the about 1 Hz estimate of this error that can be inferred from Figures 7 and 9 in (Bystrov, 1976). It should be noted that the main portion of the error comes from the approximate form of the Karplus equation and not from the error inherent in the experimental determination of the coupling constants which in our case was 0.1 Hz (cf. the Methods section). With $\alpha = 0.2$, the standard deviations increased beyond estimated errors, which removed the danger of overfitting and the ensembles became reasonably rich; therefore we decided to use this value. The experimental and ensemble-averaged NOE intensities are compared in Figure 4. It should be noted that the signals corresponding to protons of the same residues and anti-NOEs that are not shown in the diagrams were also included in the fitting. It can be noted that the entropy and number of conformations with significant weights increase much faster than the sum of the squares of the errors Φ or the standard deviation of the coupling constant and NOE intensity (Table 2). The conformations of ENKL and ENKD constituting 70% of the ensemble obtained with this value of α are shown in Figures 5a and b.

Two and three long MD runs were started for ENKL and ENKD, respectively, using the conformations with the top statistical weights, obtained by fitting EDMC ensembles with $\alpha = 0.0$ (these weights were 0.45 and 0.44 for ENKL and 0.28, 0.25, and 0.18 for ENKD, respectively). These runs were aimed at obtaining more representative ensembles, because EDMC produces only energy minima. The model used in MD simulations was also more appropriate, because it explicitly accounted for the solvation by DMSO molecules, while parameters pertaining to hydration and not DMSO solvation were used in EDMC calculations (no parameters of the SRFOPT model for DMSO solvation to be used with the ECEPP/3 force field are available). As in the case of EDMC ensembles, only a very limited number of conformations was obtained with $\alpha = 0$ (Table 2) and the coupling constants were clearly overfitted. With $\alpha = 2$, the standard deviations in coupling constants increased to error limits and a significantly large number of conformations was required to comprise 70% of the ensemble (however, the weights of this 70% fraction still differ significantly, so the respective conformations are not equally probable). The conformations that constitute about 70% of the statistical ensemble are shown in Figures 6a and 6b.

For ENKL the weighted deviation in NOE intensity ranges from 0.015 to 0.018, which is on the order of the experimental error of 0.019 (cf. the Methods

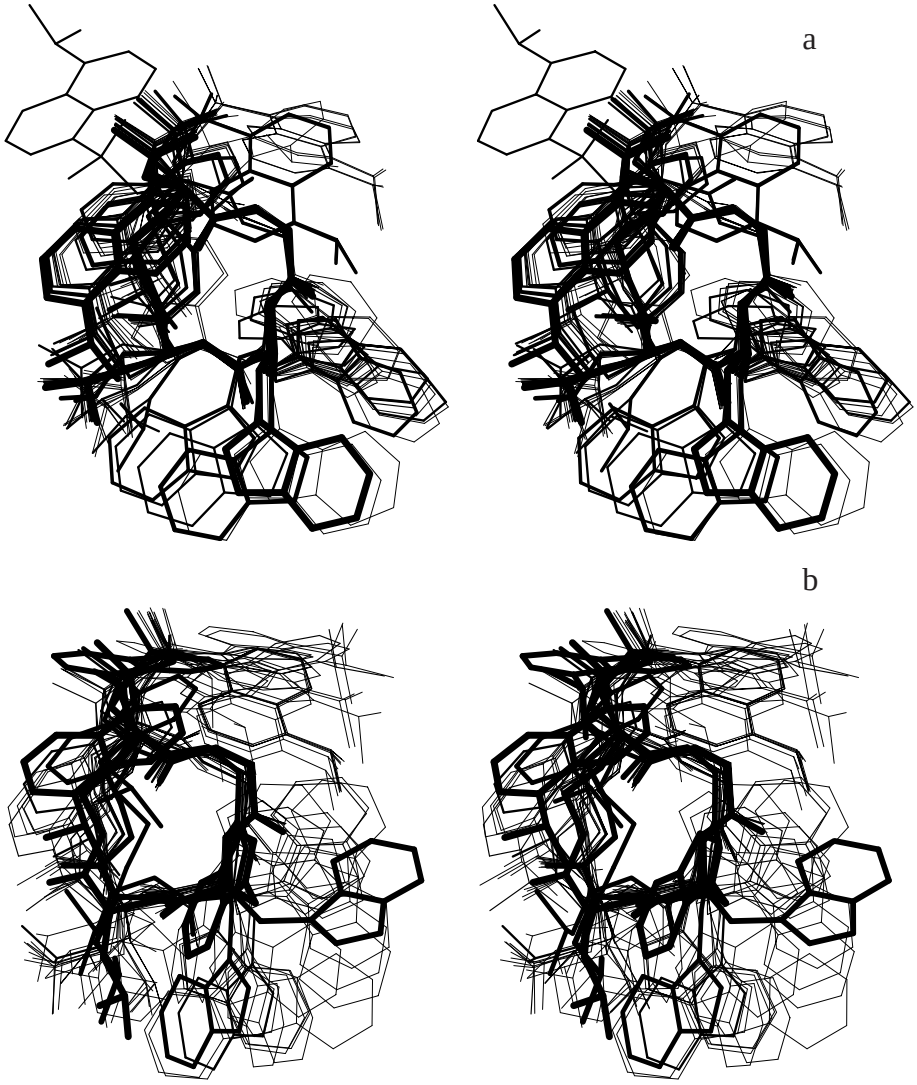

Figure 5. The conformational ensemble of ENKL (a) and ENKD (b) obtained by fitting the statistical weights of EDMC-generated conformations with $\alpha = 0.2$. The bond width is proportional to the statistical weights. The pictures were produced using MOLMOL (Koradi et al., 1996).

section). For ENKD the average standard deviation ranges from 0.051 to 0.062, which is greater than the experimental estimate. This is probably caused by the inaccuracy of the method used to compute the theoretical intensities.

It can be noted that the statistical weights of the conformations obtained with the ECEPP/3+SRFOPT force field are not correlated with their energies: conformations with energies by several kcal/mol above the lowest-energy conformations have the greatest statistical weights, while the lowest-energy conformations have low statistical weights (Figure 7). The standard deviations of the Boltzmann-averaged (using the ECEPP/3+SRFOPT energies) coupling constants and NOE intensities from the experimental values also are remarkably greater than those computed with best-fitting weights (Table 2). One obvious reason for the discrepancy between the Boltzmann weights and weights obtained by fitting to the NMR data is that the SRFOPT parameterization corresponds to solvation by water and not DMSO. However, the differences between the average values of the coupling constants calculated from MD ensembles, which have been obtained with explicit solvent, and the experimental coupling constants are also remarkably greater than the estimated experimental errors, although they are

Table 2. Measured and computed values of the vicinal coupling constants of ENKL and ENKD and other measures of the performance of the maximum-entropy fitting algorithm

Proton	J^{exp}	J^{calc} (Hz)									
		EDMC ensemble					MD ensemble				
	(Hz)	Fitted; α on the top[a]				ave.[b]	Fitted; α on the top[a]				ave.[b]
		0.0	0.1	0.2	0.3		0.0	0.5	1.0	2.0	
ENKL											
A_2bu^2(NH)–A_2bu^2($C^{\alpha}H$)	9.28	7.27	7.29	7.05	6.84	2.81	9.15	8.20	7.75	7.24	4.91
A_2bu^2($N^{\gamma}H$)–A_2bu^2($C^{\gamma}H_2$)	14.66	14.07	14.31	14.37	14.39	14.14	14.62	14.37	14.36	14.35	14.29
Gly^3(NH)–Gly^3($C^{\alpha}H_2$)	12.21	13.70	13.68	13.79	13.87	13.72	12.22	12.38	12.44	12.45	12.27
Trp^4(NH)–Trp^4($C^{\alpha}H$)	5.37	6.27	6.66	6.93	7.13	8.58	5.34	5.69	5.84	6.02	5.33
Leu^5(NH)–Leu^5($C^{\alpha}H$)	9.77	9.79	9.57	9.37	9.23	8.70	9.69	9.46	9.52	9.64	8.61
σ_J (Hz)[c]		1.10	1.16	1.29	1.42	3.34	0.07	0.54	0.75	1.18	2.03
σ_v[d]		0.015	0.017	0.017	0.017	0.019	0.015	0.018	0.018	0.018	0.018
Φ		0.94	1.15	1.36	1.57	6.91	0.01	0.41	0.57	0.81	2.40
Entropy		0.95	3.52	4.26	4.68	2.82	1.61	6.11	6.23	6.32	6.39
N[e]		2	15	28	45		3	224	275	324	
ENKD											
A_2bu^2(NH)-A_2bu^2($C^{\alpha}H$)	9.28	7.58	7.43	7.12	6.81	1.92	9.22	8.49	8.06	7.66	6.01
A_2bu^2($N^{\gamma}H$)–A_2bu^2($C^{\gamma}H_2$)	11.23	11.90	11.90	11.91	11.92	14.06	11.55	11.87	11.95	11.99	12.84
Gly^3(NH)–Gly^3($C^{\alpha}H_2$)	12.21	12.43	12.46	12.60	12.70	14.05	12.42	12.89	12.97	13.00	13.75
Trp^4(NH)–Trp^4($C^{\alpha}H$)	7.81	8.34	8.51	8.78	8.97	8.61	7.62	7.66	7.84	7.98	7.40
Leu^5(NH)–Leu^5($C^{\alpha}H$)	9.77	9.38	9.32	9.16	9.04	7.69	9.65	9.87	9.98	10.02	9.29
σ_J (Hz)[c]		0.88	0.96	1.15	1.32	3.76	0.20	0.55	0.72	0.89	1.79
σ_v[d]		0.060	0.060	0.061	0.061	0.127	0.051	0.058	0.059	0.060	0.062
Φ		2.52	2.65	2.97	3.28	15.36	0.11	2.06	2.28	2.50	5.27
Entropy		1.66	2.92	4.05	4.69	1.78	2.13	6.51	6.66	6.74	6.80
N[e]		3	7	25	49		5	347	438	510	

[a] α denotes the coefficient at the entropy term in Equation 11.
[b] Boltzmann average in the case of EDMC ensemble and unweighted average in the case of MD ensemble.
[c] Standard deviation in coupling constant.
[d] Weighted standard deviation in NOE integral intensity [weights calculated from Equation 9].
[e] Number of conformations comprising 70% or more of the ensemble.

smaller than the differences calculated from EDMC Boltzmann averages. This suggests that even considering the solvent at microscopic level is insufficient to reproduce experimental observables and work still has to be done on the parameterization of the force fields. With the current force fields, it is advisable not to apply a too tight energy cut-off on the conformations constituting the basis set used in fitting.

Dependence of the conformational ensemble on weights in the minimized sum and on the choice of entropy factor α

The functional Ψ of Equation 11 contains many constants (the scaling factors of error terms in Φ and of the entropy term) that can be estimated only within tighter or looser confidence limits. Such problems with scaling different terms also occur in other ensemble-fitting algorithms (Pearlman, 1996). It is therefore critical to determine how sensitive the derived conformational ensemble is on the choice of scaling factors. Figure 8 compares the weights calculated for the MD ensemble of ENKL with different a factors in NOE weights [Equation 9] and different w_J. As shown, the weights obtained with different scaling of error terms are in tight correlation, which indicates that the method is robust.

The sample dependence of the set of weights for the MD ensemble of ENKL on the entropy scaling factor α is shown in Figure 9. As shown, the weights gradually become more equal with increasing α. On

Figure 6. The conformational ensemble of ENKL (a) and ENKD (b) obtained by fitting the statistical weights of MD-generated conformations with $\alpha = 2.0$. The bond width and shade of grey are proportional to the statistical weights. The pictures were produced using MOLMOL (Koradi et al., 1996).

the other hand, the weights obtained with small α's are still tightly correlated with the weights obtained with $\alpha = 2$ that were used in further considerations. This indicates that the results do not change qualitatively, even within a broad range of α.

Analysis of the obtained conformational ensembles

As shown (Figures 5 and 6), the conformational ensembles derived from the EDMC and MD sets differ, although the same experimental data were used to compute the statistical weights. The most pronounced differences occur in the DNS and Trp aromatic groups, where little or no experimental data were collected. The conformations of the cyclic part of the analogs for which extensive NMR data were collected are more similar. The differences are likely to be caused by using different force fields and specifically different solvation models. Because an appropriate solvent was

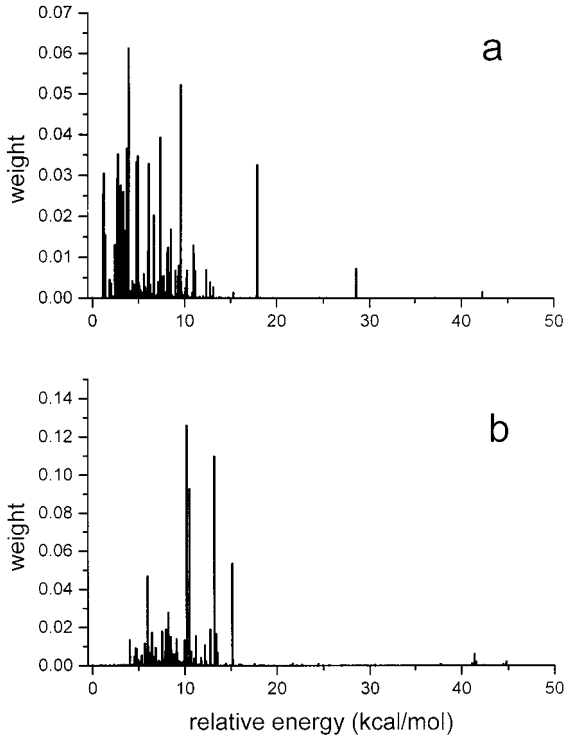

Figure 7. Impulse plot of statistical weights of the conformations of the EDMC ensembles of ENKD and ENKL obtained with $\alpha = 0.2$ versus relative conformational energy.

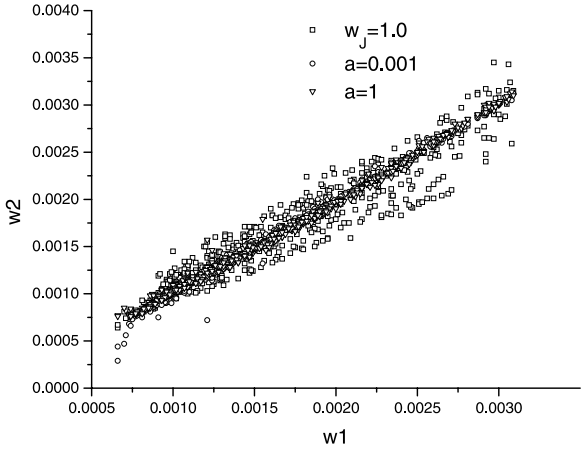

Figure 8. Correlation diagrams between the weights of the MD ensemble obtained with $w_J = 0.1$ and $a = 0.01$ of Equation 9 (w1, abscissae) with those obtained with different values of w_J and a (w2, ordinates). Squares: $w_J = 1.0$ (the entropy factor α was increased to 20, in order to maintain $\sigma_J = 1.0$), circles and triangles: $a = 0.001$ and 1.0, respectively.

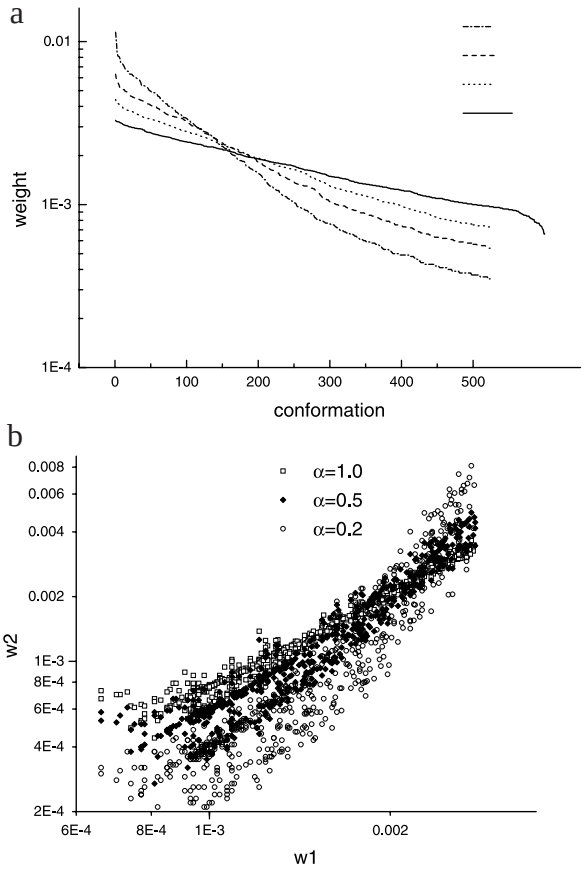

Figure 9. (a) Dependence of the statistical weights of the MD ensemble of ENKL on the scaling factor α in Equation 11. The weights are sorted in descending order. (b) Correlation between the weights determined with $\alpha = 2$ (w1, abscissae) and other values of α (w2, ordinates). Open squares: $\alpha = 1.0$, solid diamonds: $\alpha = 0.5$, open circles: $\alpha = 0.2$.

used in MD calculations, these results should be regarded more reliable, which is also supported by better fitting to the experimental data (Table 2). One can conclude that using an appropriate force field is still very important, even when experimental data are used; this is particularly remarkable in view of the fact that ensemble fitting or energy minimization subject to NMR restraints are often carried out using vacuum force fields.

A characteristic feature of the conformations of both ENKL and ENKD is a β-turn at the Gly^3–Trp^4 residues. This β-turn is forced by ring closure. The distribution of turn types (obtained using the weights calculated with $\alpha = 0.2$ for the EDMC ensembles and $\alpha = 2.0$ for the MD ensembles) is summarized in Table 3.

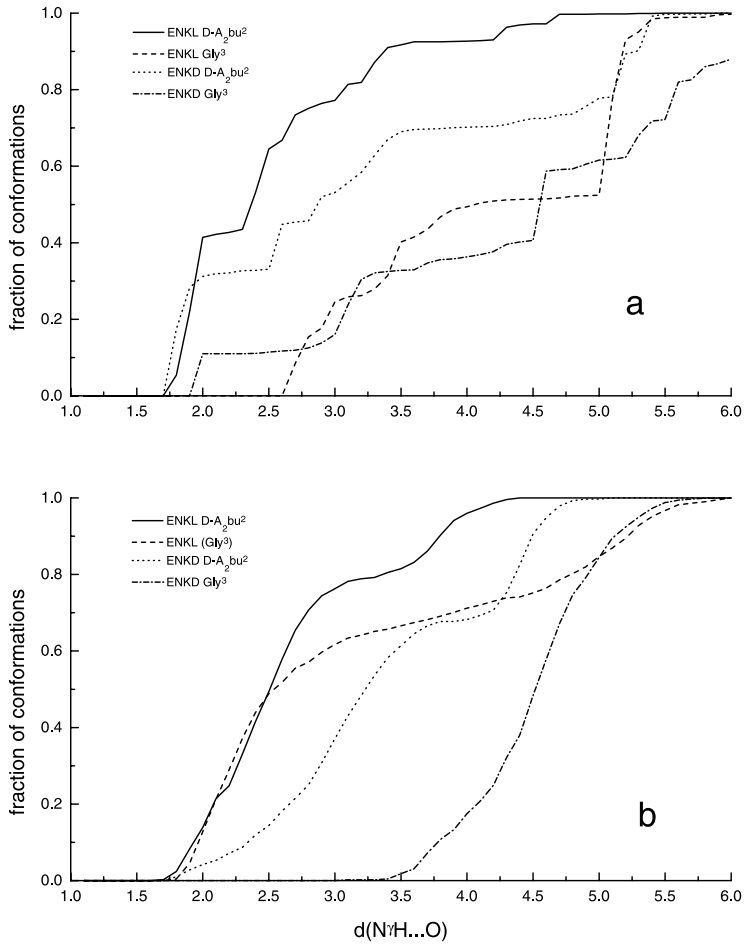

Figure 10. Distributions of the $A_2bu(N^{\gamma}H)\cdots A_2bu(CO)$ and $A_2bu(N^{\gamma}H)\cdots Gly(CO)$ distances (i.e. fractions of conformations for which the respective H···O distance is less than the abscissa value) for the EDMC (a) and MD (b) ensembles. The contributions from the individual conformations were weighted by the weights obtained with $\alpha = 0.2$ and 2.0 for the EDMC and MD ensembles, respectively.

Table 3. Fractions of β-turns of different types obtained by fitting EDMC and MD ensembles of ENKL and ENKD to NMR data

Type	EDMC		MD	
	ENKL	ENKD	ENKL	ENKD
I	0.070	0.075	0.029	0.087
II	0.000	0.003	0.000	0.000
III	0.133	0.029	0.004	0.000
IV	0.024	0.359	0.153	0.356
I′	0.000	0.000	0.000	0.000
II′	0.773	0.533	0.814	0.557
III′	0.000	0.001	0.000	0.000

The majority of defined turns are of type II′, the other turns are of type IV, with no defined hydrogen bonds, or of type III [according to the classification of Scheraga and co-workers (Lewis et al., 1973)]. The population of type II′ β-turns is greater in the case of ENKL. In the case of the EDMC ensemble of ENKL, the second group of turns contains mainly the type III turns and only a minor fraction of type IV turns. The preference of the type III β-turns in the ECEPP force field, compared to the AMBER force field, was also observed in our earlier conformational studies of vasopressin and oxytocin analogs (Shenderovich et al., 1991; Tarnowska et al., 1993). The type II′ β-turns are not tight and the Leu^5(NH) proton is not very much engaged in hydrogen bonding. In contrast to this, in the case of ENKL the D-$A_2bu^2(N^{\gamma}H)$ proton is hydrogen-bonded to D-A_2bu^2(CO) and to Gly^3(CO), while for ENKD only the first hydrogen bond exists and is weaker. This difference is illustrated in Figure 10,

Table 4. Values of temperature coefficients of the amide protons for ENKL and ENKD

Proton	$-\Delta\delta/\Delta T \times 10^3$ [ppm K^{-1}]	
	ENKL	ENKD
D-A_2bu($N^{\alpha}H$)	9.20 ± 0.07	6.10 ± 0.01
D-A_2bu($N^{\gamma}H$)	2.40 ± 0.07	5.12 ± 0.09
Gly(NH)	4.58 ± 0.07	3.4 ± 0.2
Trp($N^{\alpha}H$)	5.05 ± 0.01	7.4 ± 0.4
Trp($N^{indole}H$)	4.46 ± 0.09	3.53 ± 0.02
Leu(NH)	3.6 ± 0.2	6.7 ± 0.1

where the distributions of the D-A_2bu^2($N^{\gamma}H$)$\cdots$D-A_2bu^2(CO) and D-A_2bu^2($N^{\gamma}H$)$\cdots$$Gly^3$(CO) distances are shown. From the results of the MD simulations it follows that in the case of ENKL about 40% of the population has the D-A_2bu^2(N^{γ}-H)$\cdots$D-A_2bu^2(C=O) and D-A_2bu^2(N^{γ}-H)$\cdots$$Gly^3$(C=O) distances within 2.5 Å. For ENKD only about 15% of the population has the D-A_2bu^2(N^{γ}-H)$\cdots$D-A_2bu^2(C=O) distance within this value. Qualitatively similar, though less pronounced differences can be noticed in the case of the results of EDMC simulations. This result is in full agreement with the values of the temperature coefficients of the two analogs under study. The value of the $N^{\gamma}H$ amide proton temperature coefficient of ENKL is close to 2, which suggests that this proton is engaged in hydrogen bonds; the temperature coefficient is remarkably higher in the case of ENKD (Table 4).

The presence of more stable intra-annular hydrogen bonds in the case of ENKL probably contributes to its greater rigidity, which is, in turn, suggested to be the cause of the greater affinity of the L-Leu^5 cyclic enkephalin analogs for the μ opioid receptor, compared to the D-Leu^5 analogs (Mierke et al., 1987, 1990; Yamazaki et al., 1991).

Conclusions

In this work we have proposed a relatively fast method to determine the conformational equilibrium of small peptides by means of maximum-entropy fitting of the statistical weights of the conformations generated by theoretical calculations to NMR data. The method is not restricted to NMR data only and any combination of experimental data that carries conformation-related information can be implemented. Its advantage over the existing approaches is avoiding of 'overfitting' the experimental data, while keeping the computational cost at a reasonable level even for large input ensembles of conformations. We have shown that the resulting conformational ensembles are largely unaffected by the choice of the scaling factors of different error terms. The final conformational ensembles derived for the two cyclic enkephalin analogs under study turned out to be consistent with independent data on temperature coefficients, as well as with earlier conformational studies of related cyclic enkephalin analogs (Mierke et al., 1987, 1990; Yamazaki et al., 1991).

From the results of our study it follows that the best way to use the proposed method could consist of two stages, each comprising the conformational search and the fitting step. In the first stage, a global optimization method such as EDMC could be used, in order to generate a sparse, but widespread conformational ensemble. The conformations with top statistical weights resulting from this stage could then be used as starting conformations for the local, more detailed search by means of the MD or canonical Monte Carlo method and, finally, the statistical weights of the conformations of the ensemble obtained in the second stage could be determined. A similar approach to conformational search was recently proposed by Meirovich and co-workers (Meirovitch et al., 1995; Meirovitch and Meirovitch, 1996; Baysal and Meirovich, 1998).

Program availability

FORTRAN 77 source codes of the fitting programs, as well as test examples and documentation are available from the Cornell Theory Center software repository at http://www.tc.cornell.edu/reports/NIH/resource/CompBiologyTools/analyze/ or can be obtained from the authors upon request.

Acknowledgements

This work was supported by grants BW 8000-5-0221-6, 3 T09A 123 16, and 3T09A 102 16 from the Polish State Committee for Scientific Research (KBN). Calculations were carried out with the use of the resources and software of the Informatics Center of the Metropolitan Academic Network (IC MAN) at the Technical University of Gdańsk, Poland, and the Interdisciplinary Center for Mathematical Modeling (ICM) in Warsaw, Poland.

References

Allen, F.H., Kennard, O. and Taylor, R. (1983) *Acc. Chem. Res.*, **16**, 146–153.

Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) *J. Phys. Chem.*, **97**, 10269–10280.

Baysal, K. and Meirovich, H. (1998) *J. Am. Chem. Soc.*, **120**, 800–812.

Bhaskaran, R., Chuang, L.-C. and Yu, C. (1992) *Biopolymers*, **32**, 1599–1608.

Blackledge, M.J., Brüschweiler, R., Griesinger, C., Schmidt, J.M., Xu, P. and Ernst, R.R. (1993) *Biochemistry*, **32**, 10960–10974.

Bodanszky, M. and Bodanszky, A. (1984) *The Practice of Peptide Synthesis*, Springer, Berlin, pp. 20–25.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) *J. Biomol. NMR*, **1**, 305–309.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1994) *J. Biomol. NMR*, **4**, 143–149.

Brochon, J.C., Livesey, A.K., Pouget, J. and Valeur, B. (1990) *Chem. Phys. Lett.*, **174**, 517–522.

Brüschweiler, R., Blackledge, M. and Ernst, R.R. (1991) *J. Biomol. NMR*, **1**, 3–11.

Bystrov, V.F. (1976) *Progr. NMR Spectrosc.*, **10**, 41–81.

Carroll, R.J. and Ruppert, D. (1976) *Transformation and Weighting in Regression*, Chapman and Hall, New York, NY, pp. 13–18.

Daniell, G.J. (1991) In *Maximum Entropy in Action* (Buck, B. and Macaulay, V.A., Eds), Clarendon Press, Oxford, pp. 1–18.

Fields, G.B. and Noble, R.L. (1990) *Int. J. Pept. Protein Res.*, **35**, 161–214.

Gajewski, J.J., Gilbert, K.E. and McKevey, J. (1990) In *Advances in Molecular Modelling. A Research Annual*, Vol. 2 (Liotta, D., Ed.), Jai Press Inc., London, p. 65.

Gay, D.M. (1983) *Assoc. Comput. Math. Trans. Math. Software*, **9**, 503–524.

Gippert, G.P., Wright, P.E. and Case, D.A. (1998) *J. Biomol. NMR*, **11**, 241–263.

Gull, S.F. (1988) In *Maximum Entropy and Bayesian Methods* (Shilling, J., Ed.), Kluwer, Dordrecht, pp. 53–71.

Kemmink, J., Van Mierlo, C.P.M., Scheek, R.M. and Creighton, T.E. (1993) *J. Mol. Biol.*, **230**, 312–323.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Lewis, P.N., Momany, F.A. and Scheraga, H.A. (1973) *Biochim. Biophys. Acta*, **303**, 211–229.

Lieu, R. and Hicks, R.B. (1994) *Astrophys. J.*, **422**, 845–849.

Liu, H., Müller-Plathe, F. and van Gunsteren, W. (1995) *J. Am. Chem. Soc.*, **117**, 4363–4366.

Livesey, A.K. and Brochon, J.C. (1987) *Biophys. J.*, **52**, 693–706.

Liwo, A., Tempczyk, A., Ołdziej, St., Shenderovich, M.D., Hruby, V.J., Talluri, S., Ciarkowski, J., Kasprzykowski, F., Łankiewicz, L. and Grzonka, Z. (1996) *Biopolymers*, **38**, 157–175.

Lyon, R.G., Hollis, J.M. and Dorband, J.E. (1997) *Astrophys. J.*, **478**, 658–662.

Marquardt, D.W. (1963) *J. Soc. Industr. Appl. Math.*, **11**, 431-441.

Meadows, R.P., Post, C.B., Luxon, B.A. and Gorenstein, D.G. (1994) MORASS 2.1, Purdue University, W. Lafayette, IN.

Meirovitch, H., Meirovitch, E. and Lee, J. (1995) *J. Phys. Chem.*, **99**, 4847–4854.

Meirovitch, E. and Meirovitch, H. (1996) *Biopolymers*, **38**, 69–88.

Mierke, D.F., Lucietto, P. and Goodman, M. (1987) *Biopolymers*, **26**, 1573–1586.

Mierke, D.F., Said-Nejad, O.E., Schiller, P.W. and Goodman, M. (1990) *Biopolymers*, **29**, 179–196.

Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumseyn, S. and Scheraga, H.A. (1992) *J. Phys. Chem.*, **96**, 6472–6484.

Nikiforovich, G.V., Prakash, O., Gehrig, C.A. and Hruby, V.J. (1993) *J. Am. Chem. Soc.*, **115**, 3399–3406.

Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.

Pearlman, D.A. (1996) *J. Biomol. NMR*, **8**, 49–66.

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.A., Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1995) AMBER 4.1, University of California, San Francisco, CA.

Post, C.B., Meadows, R.P. and Gorenstein, D.G. (1990) *J. Am. Chem. Soc.*, **112**, 6796–6803.

Reichardt, C. (1988) *Solvent and Solvent Effects in Organic Chemistry*, VCH Publishers, Weinheim, pp. 365–378.

Riddick, J.A., Bunger, W.B. and Sakand, T.K. (1986) *Organic Solvents: Physical Properties and Methods of Purification*, Wiley, New York, NY.

Ripoll, D. and Scheraga, H.A. (1988) *Biopolymers*, **27**, 1283–1303.

Ripoll, D. and Scheraga, H.A. (1989) *J. Protein Chem.*, **8**, 263–287.

Ripoll, D. and Scheraga, H.A. (1990) *Biopolymers*, **30**, 165–176.

Ripoll, D.R., Liwo, A. and Scheraga, H.A. (1998) *Biopolymers*, **46**, 117–126.

Scheek, R.M., Torda, A.E., Kemmink, J. and Van Gunsteren, W.F. (1991) In *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J.C., Poulsen, F.M. and Redfield, C., Eds.), Plenum Press, New York, NY, pp. 209–217.

Schmidt, R. and Neubert, K. (1991) *Int. J. Pept. Protein Res.*, **37**, 502–507.

Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.J., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M. and Montgomery, J.A. (1993) *J. Comput. Chem.*, **14**, 1347–1363.

Shenderovich, M.D., Nikiforovich, G.V., Saulitis, J.B. and Chipens, G.I. (1988) *Biophys. Chem.*, **31**, 163–173.

Shenderovich, M.D., Kasprzykowski, F., Liwo, A., Sekacis, I., Saulitis, J. and Nikiforovich, G.V. (1991) *Int. J. Pept. Protein Res.*, **38**, 528–538.

Späth, H. (1980) *Cluster Analysis Algorithms*, Halsted Press, New York, NY, pp. 170–194.

Stewart, J.J.P. (1989) *J. Comput. Chem.*, **10**, 209–220.

Tarnowska, M., Liwo, A., Shenderovich, M.D., Liepiņa, I., Golbraikh, A.A., Grzonka, Z. and Tempczyk, A. (1993) *J. Comput.-Aided Mol. Design*, **7**, 699–720.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and Van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Vila, J., Williams, R.L., Vásquez, M. and Scheraga, H.A. (1991) *Proteins Struct. Funct. Genet.*, **10**, 199–218.

Wagner, G. (1990) *Progr. NMR Spectrosc.*, **23**, 101–139.

Waki, M., Kitajima, Y. and Izumiya, N. (1981) *Synthesis*, 266–268.

Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230–252.

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

Yamazaki, T., Said-Nejad, O.E., Schiller, P.W. and Goodman, M. (1991) *Biopolymers*, **31**, 877–898.

Yang, J.-X. and Havel, T.F. (1993) *J. Biomol. NMR*, **3**, 355–360.

Yu, C., Yang, T.-H., Yeh, C.-J. and Chuang, L.-C. (1992) *Can. J. Chem.*, **70**, 1950–1954.