

# “Ensemble” Iterative Relaxation Matrix Approach: A New NMR Refinement Protocol Applied to the Solution Structure of Crambin

Alexandre M.J.J. Bonvin, J. Antoon C. Rullmann, Rolf M.J.N. Lamerichs, Rolf Boelens, and Robert Kaptein

*Bijvoet Center, University of Utrecht, 3584 CH Utrecht, The Netherlands*

**ABSTRACT** The structure in solution of crambin, a small protein of 46 residues, has been determined from 2D NMR data using an iterative relaxation matrix approach (IRMA) together with distance geometry, distance bound driven dynamics, molecular dynamics, and energy minimization. A new protocol based on an “ensemble” approach is proposed and compared to the more standard initial rate analysis approach and a “single structure” relaxation matrix approach. The effects of fast local motions are included and *R*-factor calculations are performed on NOE build-ups to describe the quality of agreement between theory and experiment. A new method for stereospecific assignment of prochiral groups, based on a comparison of theoretical and experimental NOE intensities, has been applied. The solution structure of crambin could be determined with a precision (rmsd from the average structure) of 0.7 Å on backbone atoms and 1.1 Å on all heavy atoms and is largely similar to the crystal structure with a small difference observed in the position of the side chain of Tyr-29 which is determined in solution by both *J*-coupling and NOE data. Regions of higher structural variability (suggesting higher mobility) are found in the solution structure, in particular for the loop between the two helices (Gly-20 to Pro-22).

© 1993 Wiley-Liss, Inc.

**Key words:** crambin, iterative relaxation matrix approach, ensemble averaging, solution structure, restrained molecular dynamics, stereospecific assignments, 2D NMR

## INTRODUCTION

Structure determination of proteins in solution by NMR spectroscopy is based primarily on proton–proton distances obtained from nuclear Overhauser effects (NOEs). The quality of these structures depends to a large extent on the number of NOEs per residue that is taken into account. However, for a given set of NOEs the accuracy of the structures can be improved by deriving more precise distance con-

straints from experimental data. Various approaches have been proposed to obtain such distances from experimental NOE intensities, in particular methods based on the relaxation matrix calculations.<sup>1,2</sup> The aim of this study is to test such methods for the determination of the solution structure of proteins and compare the results with the more standard initial rate analysis.<sup>3</sup> In addition, a new refinement protocol based on an “ensemble” iterative relaxation matrix approach is proposed which should better represent the experimental NOE data since NOEs are measured as an ensemble property. As a model we studied the solution structure of crambin, a small hydrophobic protein (MW = 4715) which is found in the seeds of *Crambe abyssinica*. It is homologous to several plant toxins, such as purothionin and phoratoxin, but its precise function is unknown. A good crystal structure with a resolution of 1.5 Å is available<sup>4</sup> of which coordinates have been deposited in the Brookhaven Protein Data Bank (entry 1CRN). It has been further refined to a resolution of 0.9 Å but coordinates have, to our knowledge, not been made public.<sup>5</sup> The good quality of the X-ray structure and the fact that the molecule despite its size of only 46 residues contains two  $\alpha$ -helices and a short  $\beta$ -sheet have made crambin into a favorite model system in a number of theoretical studies and simulations.<sup>6–8</sup>

The primary structure of crambin is given in Figure 1. NMR studies have shown that crambin displays amino acid heterogeneity at positions 22 (Pro or Ser) and 25 (Leu or Ile).<sup>9</sup> The mixture obtained in the isolation procedure consists of two species, the

Abbreviations: NMR, nuclear magnetic resonance; 2D, two-dimensional; NOE, nuclear Overhauser effect; IRMA, iterative relaxation matrix approach; DG, distance geometry; DDD, distance bounds driven dynamics; MD, molecular dynamics; EM, energy minimization; rmsd, root mean squared deviation.

Received May 12, 1992; revision accepted August 17, 1992. Address reprint requests to Dr. Robert Kaptein, Department of NMR Spectroscopy, Bijvoet Center, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Rolf M.J.N. Lamerichs' present address is the Philips Medical Systems, Best, The Netherlands.

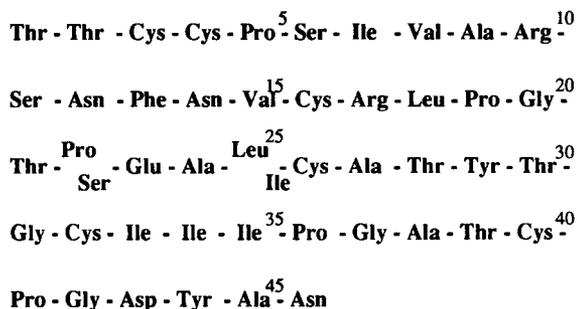


Fig. 1. Primary structure of crambin.

Ser/Ile and the Pro/Leu forms in a 55:45 ratio.<sup>9</sup> The secondary structure in solution has been found to be in agreement with the crystal structure.<sup>10,11</sup> Crambin contains two  $\alpha$ -helix regions from residue Ile-7 to Leu-18 and from Glu-23 to Thr-30, the last part of the first helix showing a transition to a  $3_{10}$ -helix for residues 17 and 18. Residues 1-4 and 32-34 form a short antiparallel  $\beta$ -sheet. For the complete resonance assignments of the  $^1\text{H}$  NMR spectrum of both crambin species see ref. 11.

Our analysis is based on NMR data for the Pro/Leu form. An extensive series of distance constraints has been derived from NOE data using an iterative relaxation matrix approach (IRMA).<sup>2,12,13</sup> A new "ensemble" approach was followed in which the calculation of theoretical NOE intensities is based on an averaged relaxation matrix. The effects of fast local motions together with methyl group rotation and aromatic ring flips were included in the calculations. Stereospecific assignments obtained from both experimental data and a model based approach and dihedral angle constraints were applied. An initial set of 20 structures was generated with distance geometry<sup>14,15</sup> and further refined in a parallel procedure combining the "ensemble" IRMA and restrained molecular dynamics calculations.<sup>16-19</sup> Two other sets of structures were generated for comparison using the more standard initial rate analysis ("NOE build-up") approach and a "single structure" IRMA. In the present paper we report the results of these calculations and show that the "ensemble" IRMA protocol leads to better structures than the initial rate analysis and allows for a better fit of the experimental NOE data. The solution structure of crambin will be described and compared to the crystal structure.

## MATERIALS AND METHODS

### Experimental

Crambin was a gift from Prof. Dr. M. Llinás (Carnegie-Mellon University, Pittsburgh) and the Pro/Leu species was purified as described before.<sup>9,20</sup> Crambin was dissolved in either acetone- $d_6$ / $^1\text{H}_2\text{O}$  (3:1 v/v) or acetone- $d_6$ / $^2\text{H}_2\text{O}$  (3:1 v/v). Since the acetone

protons exchange with the water protons the intensity of the residual protonated acetone signal in the NMR spectra increases with time and becomes important after  $\sim 48$  hr. Therefore all  $^1\text{H}_2\text{O}$  spectra were recorded on samples freshly prepared from lyophilized material.

The 2D NOE spectra for the build-up data were recorded at 500 MHz on a Bruker AM500 interfaced with an Aspect 3000 computer and processed on a VAX 11/750 or a micro VAX II using our "TRITON" software written in FORTRAN 77. Two build-up series in  $^1\text{H}_2\text{O}$  and  $^2\text{H}_2\text{O}$ , six spectra each, were recorded with a 32-step phase cycle<sup>21</sup> using TPPI.<sup>22</sup> The mixing times were 20, 40, 80, 120, 160, and 250 msec, respectively. Each set was recorded in one measuring session. Each spectrum consists of 424 FID of 32 scans each. In the  $^2\text{H}_2\text{O}$  spectra the spurious HDO signal was suppressed by irradiation during the relaxation delay (2s). All time domain data were weighted with a sine-bell shifted by  $\pi/3$ . Within one set the same phase correction parameters were used for all six spectra and a baseline correction was applied in both directions.<sup>23</sup> The final  $1\text{K} \times 1\text{K}$  spectra have a digital resolution of 5.43 Hz/point for the  $^2\text{H}_2\text{O}$  spectra and 6.10 Hz/point for the  $^1\text{H}_2\text{O}$  spectra. Peak intensities were determined from volume integrals. The volume integral consisted of the sum over all data points within a rectangular base plane. The correction factor for the difference in intensity between the  $^1\text{H}_2\text{O}$  and  $^2\text{H}_2\text{O}$  sets was determined from the build-up rates of the intrasidue aromatic cross-peaks of Phe-13 and Tyr-29 and checked on other cross-peaks involving the aromatic protons. Cross-peak intensities involving the exchangeable protons were integrated in the  $^1\text{H}_2\text{O}$  spectra, all others in the  $^2\text{H}_2\text{O}$  spectra.

A COSY spectrum in acetone- $d_6$ / $^1\text{H}_2\text{O}$  was recorded at 600 MHz with a resolution of 10.4 and 3.0 Hz/point in  $f_1$  and  $f_2$  dimensions, respectively. The coupling constants were measured directly from the antiphase splitting in the 2D spectrum after zero filling the spectrum to the final digital resolution of 6.1 and 0.76 Hz/point in  $f_1$  and  $f_2$  dimensions.

### Crystal Structure

As already mentioned crambin displays heterogeneity at positions 22 and 25. Two crambin species exists: Pro-22/Leu-25 and Ser-22/Ile-25. However, the crystal structure has been determined assuming that the (nonexisting) Pro/Ile form is the dominant component. Since 2D NMR data for the structure determination were obtained for the Pro/Leu crambin, the X-ray structure of 1.5 Å resolution as obtained from the Brookhaven Protein Data Bank had to be modified at position 25. After changing the Ile side chain in that of Leu, 30 steps of unrestrained energy minimization with the GROMOS force field<sup>24</sup> were carried out in order to remove unfavorable contacts. The total potential energy changed

from  $-1132$  to  $-1920$   $\text{kJ mol}^{-1}$ ; the rmsd between the backbone atoms of the minimized structure and the original one is  $0.06$  Å. In the following this structure will be denoted as “the X-ray structure.”

### Iterative Relaxation Matrix Approach (IRMA)

When a model for the structure and dynamics of a molecule is available, the NOEs can be calculated from the spectral densities by standard matrix techniques.<sup>25</sup> Mobility effects can be included in these calculations: methyl group rotation and aromatic ring flip can be described in a simple way by appropriate averaging of the corresponding elements of the relaxation matrix<sup>26–28</sup>, effects of fast local motions can be included via order parameters<sup>29</sup> according to the model-free approach of Lipari and Szabo<sup>30,31</sup> (see below). The opposite route from experimental NOEs to relaxation parameters is not possible directly, since the experimental NOE matrix is incomplete. We have shown, however, that the experimental data may be supplemented by NOEs calculated from a model.<sup>2,12,13</sup> The combined NOE matrix is transformed back to a corrected relaxation matrix from which new distances are calculated. Upper and lower bound margins are related to the precision with which the relaxation matrix elements can be calculated, i.e., their variation with the mixing time. Structure calculations are then performed using the new distance restraints, which now reflect the effect from direct and indirect magnetization transfers. The whole process is repeated until convergence is obtained. A complete description can be found elsewhere.<sup>2,12,13</sup> In the “ensemble” IRMA, the single structure providing the starting point to the calculations is replaced by an ensemble obtained from a set of different structures and an relaxation matrix is built from the distances of the corresponding contributions averaged as  $\langle 1/r^6 \rangle$ .

All IRMA calculations were performed with inclusion of local mobility, aromatic ring flips and methyl group rotation. The best value for the overall correlation time was found to be  $2$  nsec and an additional leakage rate of  $0.8 \text{ sec}^{-1}$  corresponding to  $1/T_1$  was added on the relaxation matrix diagonal. Six alanine  $\text{H}^\alpha$ -methyl cross-peaks on both sides of the diagonal were chosen as calibration peaks, the corresponding residues being part of the  $\alpha$ -helices.

### Model Based Stereospecific Assignments

Except for proline residues the procedures for obtaining stereospecific assignments of the prochiral centres are based on a combination of  $J$ -coupling and NOE data<sup>32,33</sup> and therefore depend on the availability of  $J$ -coupling information. Apart from this method another procedure based solely on experimental NOEs has also been used here. In this approach experimental NOEs are compared with theoretical values calculated from a model struc-

ture. All proton pairs for which at least one NOE contact has been found can be investigated. The method works for  $\text{C}^\alpha$  protons of Gly, for  $\text{C}^\beta$ ,  $\text{C}^\gamma$ , and  $\text{C}^\delta$  protons of methylene groups, for methyl groups in Val and Leu, and in principle also for the amide- $\text{NH}_2$  groups. The method can be particularly valuable for larger proteins, where the  $J$ -coupling information cannot be easily obtained.

The stereospecific assignment for a pair of prochiral protons or methyl groups ( $i^R$ ,  $i^S$ ) is based on a subset of all NOEs found for these protons. We denote here the initial experimental assignments with  $i'$ ,  $i''$ . Of all NOEs involving  $i'$  only those are used for which  $A_{i'j}^{\text{exp}} > a_{\text{lim}}$ ,  $\max(A_{i'R_j}^{\text{theo}}, A_{i'S_j}^{\text{theo}}) > a_{\text{lim}}$  and  $|A_{i'R_j}^{\text{theo}} - A_{i'S_j}^{\text{theo}}| > a_{\text{dif}}$ , where  $j$  runs over all uniquely assigned protons  $N_j$  for which an NOE involving  $i'$  has been found. The same selection criteria apply to  $i''$ . By choosing appropriate values for  $a_{\text{lim}}$  and  $a_{\text{dif}}$  only constraints that are sufficiently discriminating between the two possible assignments will be selected. The assignment is then based upon the ratio between experimental and theoretical values, after scaling, for a given mixing time. We calculate the quantities

$$f_1 = \frac{1}{(N_j + N_k)} \left( \sum_j \frac{A_{i'R_j}^{\text{theo}}}{A_{i'j}^{\text{exp}}} + \sum_k \frac{A_{i'S_k}^{\text{theo}}}{A_{i''k}^{\text{exp}}} \right) \quad (1a)$$

and for the alternative assignment:

$$f_2 = \frac{1}{(N_j + N_k)} \left( \sum_j \frac{A_{i'S_j}^{\text{theo}}}{A_{i'j}^{\text{exp}}} + \sum_k \frac{A_{i'R_k}^{\text{theo}}}{A_{i''k}^{\text{exp}}} \right) \quad (1b)$$

and their rmsd  $f_1^{\text{rms}}$  and  $f_2^{\text{rms}}$ , where  $j$  runs over the constraints set selected for  $i'$  ( $N_j$  NOEs) and  $k$  runs over the constraints set selected for  $i''$  ( $N_k$  NOEs). The following assignment is made:

$$\begin{aligned} i' &\rightarrow i^R & \text{if} & \quad |1-f_1| < |1-f_2| \\ i'' &\rightarrow i^S & & \quad f_1^{\text{rms}} < f_2^{\text{rms}} \\ & & & \quad f_1^{\text{rms}} < |f_1-f_2| \end{aligned} \quad (2a)$$

or

$$\begin{aligned} i' &\rightarrow i^S & \text{if} & \quad |1-f_2| < |1-f_1| \\ i'' &\rightarrow i^R & & \quad f_2^{\text{rms}} < f_1^{\text{rms}} \\ & & & \quad f_2^{\text{rms}} < |f_1-f_2|. \end{aligned} \quad (2b)$$

If not all three conditions can be met simultaneously, or if none of the constraints passes the selection criteria, the data are considered to be statistically unreliable. In order to avoid any bias IRMA then averages the corresponding experimental NOE intensities of the prochiral proton pairs before merging them with the theoretical data.

In the present work the selection parameters  $\alpha_{\text{im}}$  and  $\alpha_{\text{dif}}$  were both set to a value of 0.003. This guarantees the inclusion of strong NOEs corresponding to distances of about 3 Å, such as intrarésidue  $\alpha\beta$  and  $\beta\text{N}$  and many sequential contacts, plus some of the medium and long range contacts. At the same time the two assignments should correspond to reasonably different intensities: the given parameters require that for NOEs corresponding to distances of 3 Å the intensities differ by a factor of about 2, i.e., a difference in distance of about 0.3 Å.

### Generalized Order Parameters $S^2$

Effects of fast local motions can be included in the calculations of the NOE intensities by introduction of a generalized order parameter  $S_{ij}^2$  in the spectral density functions. Following the approach developed by Lipari and Szabo<sup>30,31</sup> for  $^{13}\text{C}$  relaxation, the generalized order parameter,  $S_{ij}^2$ , can be evaluated from the plateau values of the internal correlation function<sup>29</sup>  $C_{ij}^{\text{int}}(t)$ , normalized at the origin.  $C_{ij}^{\text{int}}(t)$  is defined as

$$C_{ij}^{\text{int}}(t) = \frac{4\pi}{5} \sum_{n=-2}^2 \left\langle \frac{Y_{2n}[\Phi_{ij}^{\text{mol}}(t)] Y_{2n}^*[\Phi_{ij}^{\text{mol}}(0)]}{r_{ij}^3(t) r_{ij}^3(0)} \right\rangle \quad (3)$$

Here the angular brackets indicate an ensemble average;  $Y_{2n}$  represents the second order spherical harmonics,  $r$  and  $\Phi^{\text{mol}}$  denote the length and polar angles of the interproton vector in the molecular frame of coordinates.  $S_{ij}^2$  has values between 0 and 1, and can be calculated from an MD trajectory by using Eq. (3) and estimating the plateau value for each interproton vector. These order parameters can be introduced in the spectral density functions used to calculate the elements of the relaxation matrix.

Generalized order parameters were calculated from a free MD simulation in water starting from the X-ray structure using the GROMOS force field.<sup>24</sup> Crambin was put in a box containing 1178 water molecules with periodic boundary conditions. The system was first subjected to an energy minimization with harmonic position restraining on the crambin coordinates ( $K_{\text{pos}} = 80 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ ) to remove the unfavorable interactions introduced by adding water. Initial velocities for the molecular dynamics simulation were taken from a Maxwellian distribution at 300 K. The time step was set to 1.25 fsec and temperature and pressure in the system were kept constant with relaxation times of 0.5 and 2.5 psec, respectively.<sup>34</sup> These relaxation times were sufficiently long to avoid disturbing the dynamic processes under consideration. The other parameters were the same as in the restrained MD simulations (see below). After equilibration a 50 psec MD trajectory was generated and structures were saved every 0.05 psec. This trajectory was used for analysis. The internal correlation functions  $C_{ij}^{\text{int}}(t)$  were calculated for all interproton vectors shorter than 5

Å using a fast Fourier transform algorithm. After visual inspection, generalised order parameters  $S_{ij}^2$  were computed as the averaged value of the internal correlation functions, normalized at the origin, between  $t = 8$  and  $t = 10$  psec.

### R-Factor Calculations

The quality of an NMR structure has often been expressed in terms of the residual restraint energy or sum of violations. A more direct comparison between experimental data and model structure is possible in terms of the agreement between measured and calculated NOE intensities. This can be expressed as an NMR  $R$ -factor,<sup>35–37</sup> analogous to what is done in X-ray structure refinement. We used the simple but effective definition as proposed by Gonzales et al.<sup>36</sup>

$$R = \frac{\sum_{\tau_m} \sum_i w_i(\tau_m) |A_i^{\text{theo}}(\tau_m) - A_i^{\text{exp}}(\tau_m)|}{\sum_{\tau_m} \sum_i w_i(\tau_m) |A_i^{\text{exp}}(\tau_m)|} \quad (4)$$

where the weights were chosen equal to the mixing times. A distinction was made between  $R$ -factors calculated on all NOEs or only interresidue NOEs, the latter giving a more sensitive measure of the quality of the NMR structures.

### Refinement Protocols

New structures were generated with the distance geometry algorithm (DG) derived from the original EMBED program.<sup>14</sup> This was followed by distance bounds driven dynamics (DDD),<sup>38</sup> a Newtonian dynamics simulation using a chiral constraint function and the distance constraint function of Eq. (5)

$$V_{\text{dis}} = K_{\text{dis}} \left( \sum_{d_{ij} > u_{ij}} (d_{ij}^2 - u_{ij}^2)^2 + \sum_{d_{ij} < l_{ij}} (d_{ij}^2 - l_{ij}^2)^2 \right) \quad (5)$$

where  $u$  and  $l$  represent upper and lower bounds, respectively, and  $d$  the actual distance. The force constant  $K_{\text{dis}}$  was set to  $1 \text{ kJ mol}^{-1} \text{ \AA}^{-4}$ . Two runs, the first at 300 K, the second at 1 K, were performed. By “shaking up” the molecule the DDD procedure improves the sampling of allowed conformational space, while keeping the violations at the same level or reducing them further. All information about the covalent structure is represented by upper and lower bounds between atoms separated by less than 4 bonds.

Structure refinement was performed with molecular dynamics and energy minimization using the GROMOS force field and programs.<sup>24</sup> In this force field only polar protons are treated explicitly; the

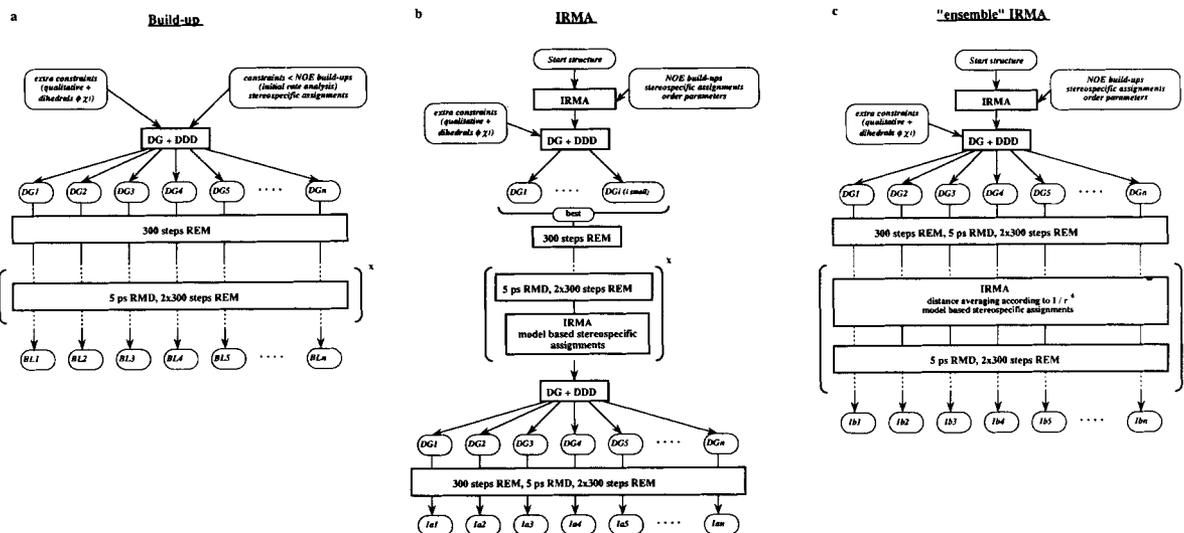


Fig. 2. Flow diagrams of the three refinement protocols: (a) build-up, (b) “single structure” IRMA, and (c) “ensemble” IRMA. The abbreviations used are DG (distance geometry), DDD (distance bound driven dynamics), REM (restrained energy minimization), and RMD (restrained molecular dynamics).

others are incorporated using the united atom technique. Distance constraints were entered via an extra harmonic pseudo potential  $V_{dis}$  defined in the following way<sup>18</sup>:

$$V_{dis} = \begin{cases} \frac{1}{2} K_{dis} (d_{ij} - l_{ij})^2 & 0 \leq d_{ij} \leq l_{ij} \\ 0 & l_{ij} \leq d_{ij} \leq d_{ij} \\ \frac{1}{2} K_{dis} (d_{ij} - u_{ij})^2 & u_{ij} \leq d_{ij} \end{cases} \quad (6)$$

where  $u$  and  $l$  represent the upper and lower bounds. Dihedral constraints were also applied with a potential  $V_{dihed}$  defined as

$$V_{dihed} = K_{dihed} [1 - \cos(\phi - \phi_0)] \quad (7)$$

where  $\phi_0$  represents the dihedral angle value obtained from  $J$ -coupling data. Force constants were set in all calculations to  $40 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$  for  $K_{dis}$  and  $110 \text{ kJ mol}^{-1}$  for  $K_{dihed}$ , this latter value allowing fluctuations for the dihedral angles of  $\pm 20^\circ$  at 300 K. Restrained MD simulations were done in vacuo with a time step of 2 fsec. The temperature in the system was kept constant with a relaxation time of 0.1 psec. A cut-off of  $8 \text{ \AA}$  was used and the non-bonded pair list was updated every 10 steps. Initial velocities were taken from a Maxwellian distribution at 300 K.

Three different refinement protocols presented in Figure 2 were tested.

### Build-up

In this protocol distance constraints are obtained from an initial rate analysis of the build-up curves

without corrections for spin diffusion. These were used in DG calculations followed by two times 400 steps DDD simulations at 300 and 1 K, respectively. The resulting structures are first energy minimized using 300 steps of steepest descent and then refined using a combination of 5 psec restrained MD followed by two times 300 steps restrained EM with steepest descent and conjugated gradient, respectively. This last refinement cycle can be repeated a number of times.

### “Single structure” IRMA

In this protocol the iterative relaxation matrix approach is applied to the refinement of one structure until convergence, monitored in terms of  $R$ -factors, is achieved. We can start from a linear chain, in which case an initial set of structures has to be generated using DG, the best structure being selected for further refinement. Alternatively, another structure, corresponding to a low resolution structure, for example, can be used if available. This structure is refined with the same combination of restrained MD and EM steps as in the build-up protocol. After each cycle a new IRMA calculation is performed and the distance constraints are updated. The process is repeated until convergence is achieved. A new set of structures is then generated from the final IRMA constraints with DG followed by DDD and the MD/EM refinement procedure of the build-up protocol.

### “Ensemble” IRMA

The starting point for this protocol is identical to the “single structure” IRMA protocol. However, an ensemble of structures is directly generated in the

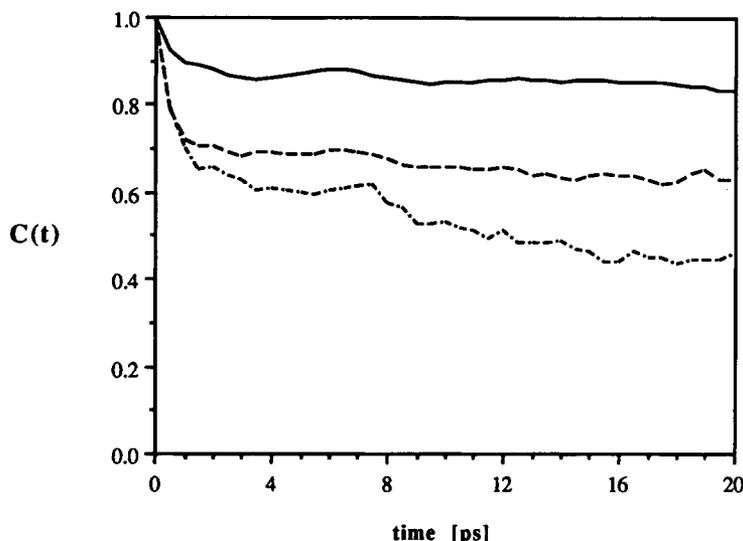


Fig. 3. First 20 psec of the internal correlation functions of the Val-8 HN-Ala-9 HN (—), Ile-33 HN-Ile-33 H $\gamma$  (- - -), and Thr-2 H $\beta$ -Phe-13 H $\alpha$  (- · - · -) interproton vectors normalized at  $t = 0$ .

first cycle with DG followed by DDD. All structures are then refined in parallel with the standard EM, MD, EM combination. At this stage an averaged relaxation matrix is built from the ensemble and used for a new IRMA calculation. Parallel refinement of the ensemble is then pursued with the new constraints reflecting now both spin diffusion effects and ensemble averaging. The "ensemble" IRMA refinement is repeated until convergence is achieved.

Additional constraints corresponding to NOE peaks for which no build-up series are available or from other sources and dihedral angle constraints obtained from  $J$ -coupling data are included in all three protocols. In addition, both IRMA protocols make use of the model based stereospecific assignments.

## RESULTS

### Order Parameters

The internal correlation functions of all interproton vectors within 5 Å in the X-ray structure were computed from the 50 psec free MD trajectory in water. This resulted in a set of 2,037 correlation functions from which order parameters were evaluated. Figure 3 presents the first 20 psec of three of these correlation functions corresponding to the amide HN $_i$ -HN $_{i+1}$  interproton vector between Val-8 and Ala-9, the amide-side chain HN $_i$ -H $_{i\gamma}$  interproton vector in Ile-33 and a long range interproton vector between a methylene proton of Thr-2 and an aromatic proton of Phe-13. A fast initial decay within a few picoseconds can be observed for all three functions. As expected, more mobility is found for side chain and long range vectors. A correlation can be found between order parameter values and secondary structure elements. The plot of the amide

HN $_i$ -H $_{i\alpha}$  order parameters against the residue sequence in Figure 4 allows one to recognize the two  $\alpha$ -helices and  $\beta$ -sheet regions in crambin. Regions of higher mobility correspond to turns and loops in the structure. The computed order parameters were used in all IRMA and  $R$ -factor calculations.

### Stereospecific Assignments

Crambin contains 38 diastereotopic methylene protons pairs and two valine residues. From an analysis of the  $J$ -coupling data 16 proton pairs and the methyl groups of one valine residue could be stereospecifically assigned, which represents 43% of the possible assignments.<sup>11</sup> Using the model based method 17 proton pairs and the methyl groups of the two valine residues could be assigned, which represents 48% of the possible assignments. The latter were based on theoretical NOE intensities calculated from the last "ensemble" IRMA cycle. The procedure was also tested with the X-ray structure as a model, but only the assignments obtained from the solution structures were used in the calculations. Combination of the experimental and model based results leads to stereospecific assignment of 20 proton pairs and of the methyl groups of the two valine residues which corresponds to 55% of the possible assignments. Experimental and model based stereospecific assignments are presented in Table I. It can be noted that there is very good correspondence between assignments based on  $J$ -coupling data and NOE intensities.

### Distance Constraints

For both build-up series the intensities of 735 cross-peaks on both side of the diagonal could be determined.<sup>11</sup> After a visual inspection of the build-

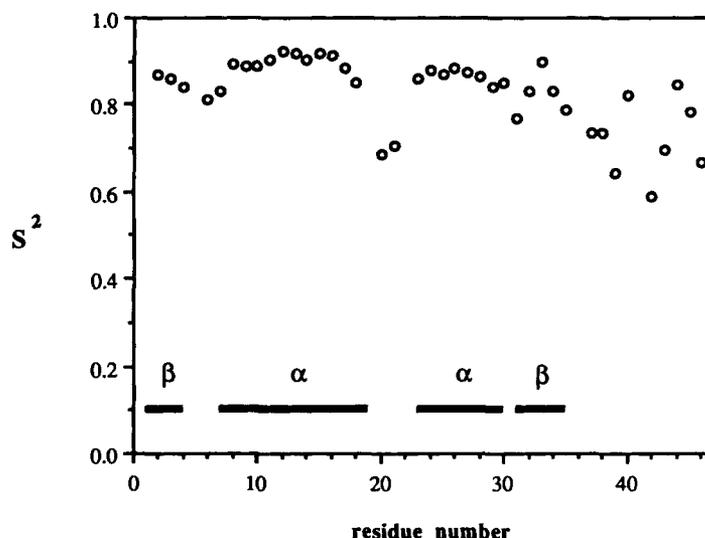


Fig. 4. Plot of the intraresidue  $HN-H^{\alpha}$  order parameters against the residue sequence and comparison with the secondary structure elements.

TABLE I. Stereospecific Assignments From  $J$ -Coupling and NOE Data\*

Residue	Atom	J	NOE <sup>†</sup>		Residue	Atom	J	NOE <sup>†</sup>	
			NMR	X-ray				NMR	X-ray
Cys-3	H <sup>β</sup>	s	s	s	Pro-19	H <sup>δ</sup>			
Pro-5	H <sup>β</sup>	s	s	s	Gly-20	H <sup>α</sup>			n
Pro-5	H <sup>δ</sup>				Pro-22	H <sup>β</sup>	s		
Ser-6	H <sup>β</sup>				Cys-26	H <sup>β</sup>	s	s	
Ile-7	H <sup>β</sup>			s	Tyr-29	H <sup>β</sup>	n		
Val-8	H <sup>γ1</sup>		n	n	Gly-31	H <sup>α</sup>		n	n
Arg-10	H <sup>γ</sup>		n		Cys-32	H <sup>β</sup>	n	n	n
Arg-10	H <sup>δ</sup>			s	Ile-33	H <sup>γ1</sup>			s
Arg-10	H <sup>γ1</sup>				Ile-34	H <sup>γ1</sup>			s
Asn-12	H <sup>β</sup>	n	n		Ile-35	H <sup>γ1</sup>		n	s
Asn-12	H <sup>δ2</sup>				Pro-36	H <sup>β</sup>	s	s	
Phe-13	H <sup>β</sup>				Pro-36	H <sup>δ</sup>		s	s
Asn-14	H <sup>β</sup>	n	n	n	Cys-40	H <sup>β</sup>	s	s	s
Asn-14	H <sup>δ2</sup>				Pro-41	H <sup>β</sup>	s		
Val-15	H <sup>γ1</sup>	n	n	n	Pro-41	H <sup>δ</sup>			
Cys-16	H <sup>β</sup>	s			Asp-43	H <sup>β</sup>	s	s	
Arg-17	H <sup>β</sup>		s		Tyr-44	H <sup>β</sup>	s	s	s
Arg-17	H <sup>γ</sup>				Asn-46	H <sup>β</sup>	s	s	s
Arg-17	H <sup>δ</sup>			n	Asn-46	H <sup>δ2</sup>			
Pro-19	H <sup>β</sup>	s	s						

\*The label n (no swap) means that the initial assignment based upon resonance position is correct, the label s (swap) means that it must be reversed.

†Model based stereospecific assignments were made for mixing times = 80, 120, 160, and 250 msec. Only if the results obtained for all four mixing times are not conflicting are they included in the table. X-ray denotes the assignments obtained for the X-ray structure and NMR the assignments obtained in the "ensemble" IRMA protocol. NOEs were calculated with inclusion of local mobility, methyl group rotation, and aromatic ring flip.

up curves 91 cross-peaks were removed from the list, since the quality of the build-up curves was too poor. These latter were, however, used as qualitative constraints in the calculations with upper bounds values of 2.8, 3.5 and 4.5 Å—before pseudoatom correction—for strong, medium, and weak NOEs, respectively. Fifty additional NOEs which are close

to the water frequency were identified and introduced as upper bounds of 5 Å—before pseudo atom correction. This resulted in a set of 132 qualitative constraints—after removing redundant information—for which lower bounds were chosen equal to the sum of the van der Waals radii. The three disulfide bridges were introduced in the distance geome-

**TABLE II. Final NMR Constraints Used in the "Ensemble" IRMA Protocol**

Distances*	Number
Intraresidue	248
Sequential	129
Medium	75
Long range	91
Total	543
Dihedral angles	Number
${}^3J_{\text{HN}\alpha} < 6 \text{ Hz}$	18
$8 \text{ Hz} < {}^3J_{\text{HN}\alpha} < 9 \text{ Hz}$	2
${}^3J_{\text{HN}\alpha} < 9 \text{ Hz}$	11
$\chi_1$ constraints	13

\*411 IRMA constraints + 132 qualitative constraints corresponding to observed NOEs without build-up series.

try calculations as having an upper bound of 2.1 Å and a lower bound of 1.9 Å.

Constraints were obtained from the 644 remaining build-up series in two different ways:

1. Initial rate analysis was applied for the build-up protocol. All distances were calibrated using a single proportionality constant corresponding to a distance of 3.5 Å. Upper and lower bounds were obtained by taking an additional margin of  $\pm 10\%$  of the actual distance. Pseudoatoms were introduced on the prochiral centers when no experimental stereospecific assignment was available and the bounds were relaxed accordingly. As proposed by Holak et al.,<sup>39</sup> explicit lower bounds were used instead of the sum of the van der Waals radii to obtain a good fit between calculated and observed NOE intensities.

2. In both IRMA protocols, distance limits issued from the relaxation matrix calculations were used directly as upper and lower distance constraints. Pseudoatoms were introduced on the prochiral centers when no stereospecific assignment was available and the bounds were relaxed accordingly. In addition to the experimental stereospecific assignments model based assignments were applied except in the first cycle. For methyl groups pseudo atoms were used with a correction of 0.3 Å as suggested by Koning et al.<sup>28</sup> when fast methyl rotation is taken into account.

Some statistics concerning the final distance constraints from the "ensemble" IRMA protocol are presented in Table II.

### Dihedral Angle Constraints

Backbone  $\phi$  dihedral angle constraints were obtained from  ${}^3J_{\text{HN}\alpha}$  coupling constants measured in the COSY spectrum; 35  ${}^3J_{\text{HN}\alpha}$  coupling constants were measured, but only 31 could be transformed into dihedral angle constraints. For  ${}^3J_{\text{HN}\alpha} > 9 \text{ Hz}$

the dihedral angle constraint was  $-120^\circ \pm 30^\circ$ , for  $8 < {}^3J_{\text{HN}\alpha} < 9 \text{ Hz}$  the constraint was  $-120^\circ \pm 40^\circ$  and for  ${}^3J_{\text{HN}\alpha} < 6 \text{ Hz}$  and the residue being part of an  $\alpha$ -helix the constraint was  $-60^\circ \pm 30^\circ$ . These dihedral angle constraints were transformed into distance constraints for both the DG and MD calculations. From the experimental stereospecific assignment procedure 13  $\chi_1$  dihedral angle constraints could be imposed. These were also transformed into distance constraints for the DG calculations but directly used in the potential function of Eq. (7) for the MD refinement. These constraints are summarized in Table II.

### Structure Calculations

Sets of 20 structures were generated with each protocol. The *build-up* structures were refined with one cycle of 5 psec restrained MD followed by energy minimization. In the "single structure" IRMA protocol, a total of three cycles ( $x = 2$  in Fig. 2b) were necessary to obtain convergence. A linear chain of crambin provided the initial model for the IRMA calculations. Four structures were then generated with DG followed by DDD and the best one was further refined with two IRMA cycles. From the constraints obtained at this point 20 structures were generated and refined as indicated in Figure 2b. Two of the final structures with high energies were discarded for the analysis. In the "ensemble" IRMA protocol, two cycles ( $x = 1$  in Fig. 2c) were sufficient to achieve convergence. The initial model was also a linear chain and 20 structures were generated in the first DG calculations. After a first refinement with restrained MD, a relaxation matrix was built from the ensemble by averaging the different contributions according to  $\langle 1/r^6 \rangle$ . The three sets of structures are presented in Figure 5. A visual inspection directly reveals that the "ensemble" approach leads to better results. This is confirmed by both the rmsd and *R*-factors within each set. Table III lists the average rmsd from the average structure; results are presented for all heavy, for all backbone atoms (C,C $^\alpha$ ,N) and for the well-defined region excluding the first and last residues and the three residues constituting the loop between the two helices (Gly-20, Thr-21 and Pro-22). Both IRMA protocols give better results than the build-up approach. The lowest rmsd are obtained for the "ensemble" IRMA structures with values of 1.1 and 0.7 Å, average rmsd from the average structure for all heavy and backbone atoms for the well-defined region, respectively. No statistical difference with the X-ray structure is found since all rmsd from the X-ray structure are within the average pairwise rmsd for each of the three sets. Table IV presents energies and *R*-factors for all three sets and for the X-ray structure. The total potential energies, excluding restraint energies, are comprised between -1,647 and -1,998 kJ mol<sup>-1</sup> with the lowest value found for an "ensemble"

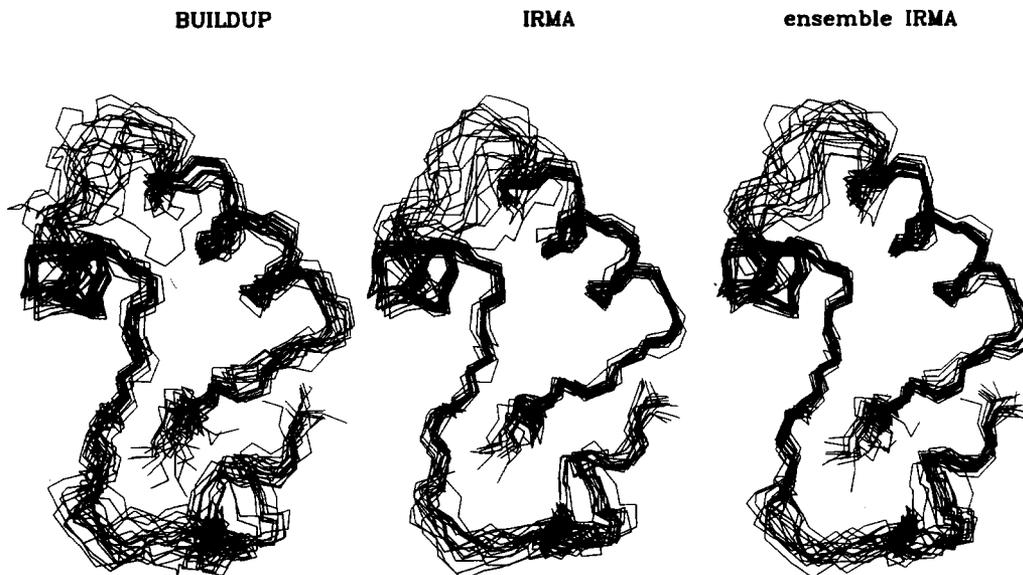


Fig. 5. View of the backbone of the three sets of structures generated according to the protocols of Figure 2. The structures were fitted on backbone C, N, C $\alpha$  atoms.

**TABLE III. Averaged rms Deviations in Å for the Three Sets of Structures Obtained According to the Protocols of Figure 2.**

Residues (atoms)	Build-up	IRMA	“Ensemble” IRMA
rmsd from average structure			
1–46 (all heavy)	1.5	1.3	1.2
2–19, 23–45* (all heavy)	1.4	1.2	1.1
1–46 (backbone)	1.1	1.0	0.9
2–19, 23–45* (backbone)	1.0	0.8	0.7
rmsd from X-ray structure			
1–46 (all heavy)	2.1	1.9	1.8
2–19, 23–45* (all heavy)	2.0	1.6	1.6
1–46 (backbone)	1.3	1.2	1.1
2–19, 23–45* (backbone)	1.2	0.9	0.8

\*First and last residues and the three residues constituting the loop between the two helices were omitted.

IRMA structure. Distance restraint energies between build-up and IRMA structures are not directly comparable since the IRMA constraints are tighter than the corresponding ones obtained from the initial rates analysis. All NMR structures satisfy quite well the experimental constraints: only a few constraints are violated by more than 0.5 Å and none by more than 1 Å. Larger violations are found in the X-ray structure for both dihedral angle and distance constraints as can be seen from the corresponding restraint energies. A direct comparison be-

tween the three sets and the X-ray structure is provided by the NMR *R*-factors: the X-ray structure scores quite well with an interresidue *R*-factor of 0.41 which is slightly lower than those calculated for the *build-up* and “*single structure*” IRMA sets. The performance of the “*ensemble*” IRMA is now well highlighted by the lowest *R*-factors of 0.35 and 0.38 calculated on all NOE peaks and interresidue peaks, respectively. Note that the *R*-factor calculations for all three sets were performed by computing the relaxation matrices from distances averaged as  $\langle 1/r^6 \rangle$  over the ensemble.

#### Solution Structure Analysis and Comparison With the X-Ray Structure

Our analysis is based on the structures obtained with the “*ensemble*” IRMA protocol. The solution structure of crambin is well determined for the two helices and the short antiparallel  $\beta$ -sheet as can be seen from Figure 5 and from the plot of the average rmsd on C $\alpha$  atoms from the average structure against the residue sequence in Figure 6. Regions of higher variability are found in the loop between the helices and in turns. These observations are in line with the order parameter values shown in Figure 4 computed from the free MD simulation in water. This suggests that the high structural variability such as in the loop region (Gly-20 to Pro-22) is in fact due to a high mobility of this peptide region with motions on the picosecond time scale. There is also a reasonable correspondence with the crystallographic *B*-factors given in Figure 6, the variability in the loop, however, being larger in solution.

The number of violations calculated from aver-

TABLE IV. Total Potential ( $E_{\text{pot}}$ ),\* Distance Restraint ( $E_{\text{dis}}$ ),<sup>†</sup> and Dihedral Restraint ( $E_{\text{dihed}}$ )<sup>‡</sup> Energies and R-Factors ( $R$ )<sup>§</sup> of the Crambin Structures

	Build-up**	IRMA <sup>††</sup>	"Ensemble" IRMA**	X-ray
$\langle E_{\text{pot}} \rangle$	-1848	-1846	-1891	-1920
$\langle E_{\text{dis}} \rangle$	183 <sup>‡‡</sup>	360 <sup>§§</sup>	288 <sup>§§</sup>	882 <sup>§§</sup>
$\langle E_{\text{dihed}} \rangle$	9	7	6	162
$E_{\text{pot}}$ range	-1991 to -1742	-1991 to -1647	-1998 to -1771	
$E_{\text{dis}}$ range	105 to 242	305 to 459	221 to 361	
$E_{\text{dihed}}$ range	6 to 16	3 to 11	3 to 10	
$R_{\text{all}}$	0.37	0.36	0.35	0.38
$R_{\text{interresidue}}$	0.42	0.42	0.38	0.41

\*GROMOS force field,<sup>24</sup> excluding distance and dihedral restraint energies.

<sup>†</sup>Harmonic potential of Eq. (6), force constant = 40 kJ mol<sup>-1</sup> Å<sup>-2</sup>. Energy calculated from averaged distances.

<sup>‡</sup>Potential of Eq. (7), force constant = 110 kJ mol<sup>-1</sup>.

<sup>§</sup>R-factors from Eq. (4). Theoretical NOEs calculated from an averaged relaxation matrix.

\*\*Set of 20 structures

<sup>††</sup>Set of 18 structures (two high energy conformation were discarded).

<sup>‡‡</sup>525 stereospecific constraints.

<sup>§§</sup>543 stereospecific constraints.

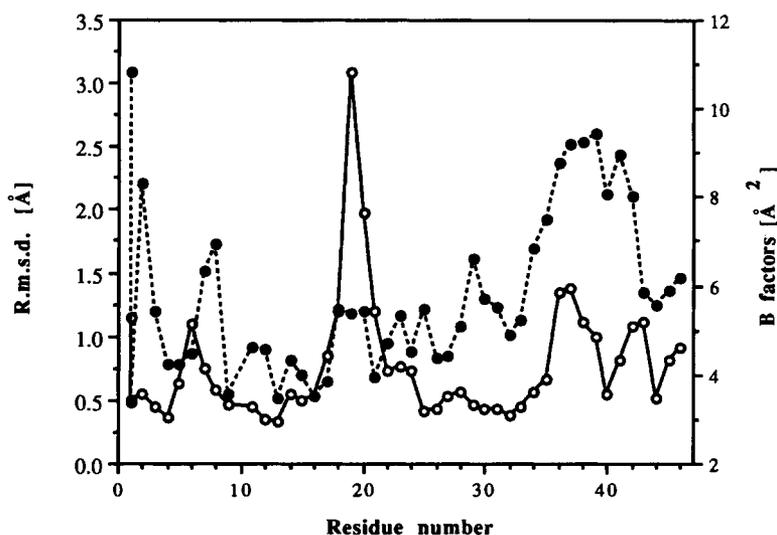


Fig. 6. Rms deviations from the average structure on C $\alpha$  atoms per residue (○) for the set of structures obtained with the "ensemble" IRMA protocol of Figure 2c and C $\alpha$  crystallographic B-factors (●).

aged distances is 185, 102 of which are upper bound violations. Figure 7 presents the size distribution of these violations. None exceeds 1.0 Å, only 8 are larger than 0.5 Å, and 72% is less than or equal to 0.2 Å. We checked our structures for bond distances, bond angles, and improper dihedral angles. Due to the procedure used in the GROMOS MD simulations to maintain constant bond distances (SHAKE),<sup>40</sup> no deviation of the bond distances exceeding 0.004 Å is observed! For the bond angles the average of the absolute value of the deviation from the GROMOS standard values is found to be 2.1° with a maximum deviation of 14.5° for the N-C $\alpha$ -C angle in Pro-41. The average deviation on the improper dihedral angles is 3.0° with a maximal deviation of 16.0° for the N-C $\alpha$ -C-C $\beta$  improper dihedral of Pro-19.

The side chain  $\chi_1$  dihedral angles satisfy the experimental values obtained from  $J$ -coupling data. Table V lists their averaged values. One significant difference between the solution and crystal structures is found in the  $\chi_1$  angle of Tyr-29 which is around -60° in solution, while a value of -172° was reported for the crystal. However, disorder was observed at this position in the X-ray structure which is possibly related to the fact that the structure was determined for a mixture of both crambin species (Pro/Leu and Ser/Ile).

An analysis of the conformation per residue is presented in Figure 8 in the form of a Ramachandran map of the  $\phi$  and  $\psi$  dihedral angles for the set of 20 NMR structures. Typical values are found for the helices and the antiparallel  $\beta$ -sheet. The majority of

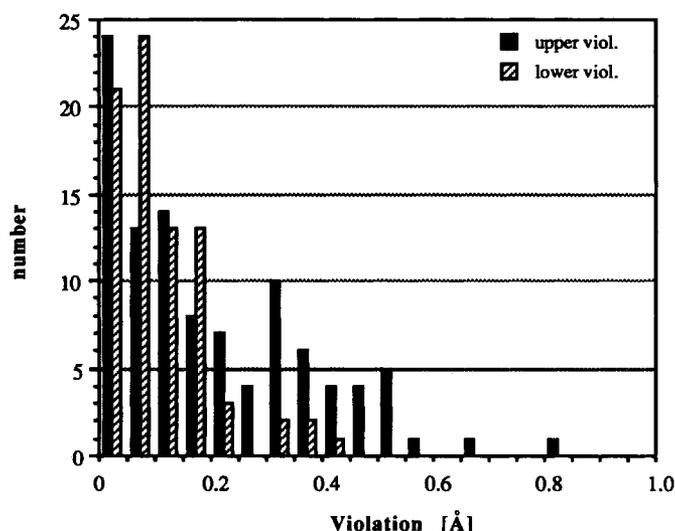


Fig. 7. Size distribution of the constraint violations in Å calculated from the averaged distances on the 20 "ensemble" IRMA structures.

the  $\phi \psi$  combinations found in the forbidden area corresponds to glycine residues with a few exceptions indicated in the figure. In the average structure all the  $\phi \psi$  dihedral angles lie in the allowed regions.

The hydrogen bond pattern of the structures explains to a large extent the exchange behavior of the backbone amide protons.<sup>10</sup> Experimental and theoretical results are summarized in Table VI. Two sets of criteria were used in the search for hydrogen bonds. The first one used a maximum distance between proton and acceptor of 2.8 Å and a minimum angle of 135°; the second one used a maximum distance of 3.0 Å and an minimum angle of 90°. For the solution data, hydrogen bonds were indicated only if

**TABLE V. Side Chain Dihedral Angles of Those Residues for Which Experimental Data Are Available for the Crambin Structures\***

Residue	$\chi_1$		
	X	Exp.	NMR
Cys-3	-53	-60	-58
Asn-12	-69	-60	-64
Phe-13	-174	-180	-171
Asn-14	-70	-60	-62
Val-15	-65	-60	-65
Cys-26	-69	-60	-60
Tyr-29	-172	-60	-62
Cys-32	-59	-60	-56
Cys-40	-69	-60	-68
Asp-43	+62	+60	+63
Tyr-44	-71	-60	-64
Asn-46	-59	-60	-62

\*X denotes the X-ray structure and exp. the experimental values based on *J*-coupling data. NMR denote the averaged angles from the final 20 structures of the "ensemble" IRMA protocol.

found in a relevant number of structures. Hydrogen bonds were found in both the X-ray and the NMR structures for the slow exchanging amide protons of the two helices (Ala-9 to Leu-18 and Leu-25 to Thr-30) and of the antiparallel  $\beta$ -sheet (Cys-3, Ile-33, and Ile-35). The slow or intermediate exchange of some amide protons can be explained by hydrogen bonds formed in turns. The intermediate exchange of the amides of Gly-20 and Thr-21 is due to hydrogen bonds to the carbonyls of Cys-16 and Pro-19, respectively. A different pattern is found for these two residues in the X-ray structure, the amide of Thr-21 forming hydrogen bond to its own  $\gamma$ -oxygen and a strong hydrogen bond being found between the amide of Gly-20 and the carbonyl of Arg-17. These observations point to somewhat different conformations of this loop region between the solution and crystal structures. Hydrogen bonds to the carbonyls of Thr-28 and Ala-27 explain the slow exchange of Gly-31 and Cys-32, respectively. A type-I turn is found in this region confirmed by the NOEs found between the amide HN of Cys-32 and both the H $^\alpha$  and amide HN of Thr-30. Another Type-I turn is found at Pro-41 which explains the slow exchange of the amide of Tyr-44 by an hydrogen bond to the carbonyl of Pro-41. Weak hydrogen bonds are found in the solution structures for the amide protons of Val-8 and Glu-23 explaining their intermediate exchange. There are further some weak hydrogen bonds observed for the amide protons of Cys-4, Ala-45, and Asn-46 which are in contradiction with the fast exchange observed for these protons. The slow exchange of Ser-6 is not well accounted either in the solution or in the crystal structures. The same is true for the amide proton of Leu-25. A possible explanation in the latter case is offered by the obser-

TABLE VI. Comparison Between the Hydrogen Bonds Found in the Final NMR (N)\* and X-Ray (X) Structures and the Amide Exchange† NMR Data<sup>10</sup>

Amide	Acceptor	N	X	E	Amide	Acceptor	N	X	E		
Thr-1	—			f	Leu-25	Pro-22	C=O	±	±	s	
Thr-2	Thr-2	O <sub>γ</sub>		±	f	Cys-26	Pro-22	C=O	+	+	s
Cys-3	Ile-33	C=O	+	+	s	Ala-27	Glu-23	C=O	+	+	s
Cys-4	Tyr-44	C=O	±		f	Thr-28	Ala-24	C=O	+	+	s
Cys-4	Asn-46	O <sub>δ</sub>		±	f	Tyr-29	Leu-25	C=O	+	+	s
Pro-5	—				Thr-30	Cys-26	C=O	+	+	s	
Ser-6	Ser-6	C=O		±	s	Thr-30	Ala-27	C=O	±	+	s
Ile-7	Asn-46	O <sub>δ</sub>			i	Gly-31	Ala-27	C=O	±	+	s
Val-8	Ser-6	O <sub>γ</sub>	±		i	Gly-31	Thr-28	C=O	+		s
Ala-9	Ser-6	O <sub>γ</sub>	±	+	s	Cys-32	Ala-27	C=O	+	±	s
Arg-10	Ser-6	C=O	+	+	s	Cys-32	Thr-30	O <sub>γ</sub>		+	s
Ser-11	Ile-7	C=O	+	+	s	Ile-33	Cys-3	C=O	+	+	s
Asn-12	Val-8	C=O	+	+	s	Ile-34	—				s
Phe-13	Ala-9	C=O	+	+	s	Ile-35	Thr-1	C=O	+	+	s
Asn-14	Arg-10	C=O	+	+	s	Pro-36	—				
Val-15	Ser-11	C=O	+	+	s	Gly-37	—				f
Cys-16	Asn-12	C=O	+	+	s	Ala-38	—				f
Arg-17	Phe-13	C=O	+	+	s	Thr-39	—				f
Leu-18	Val-15	C=O	+	+	s	Cys-40	—				f
Pro-19	—				Pro-41	—					
Gly-20	Cys-16	C=O	±		i	Gly-42	—				f
Gly-20	Arg-17	C=O		+	i	Asp-43	Asp-43	O <sub>δ</sub>	±	±	f
Thr-21	Pro-19	C=O	±		i	Tyr-44	Pro-41	C=O	±	+	s
Thr-21	Thr-21	O <sub>γ</sub>		±	i	Tyr-44	Gly-42	C=O	±		s
Pro-22	—					Ala-45	Gly-42	C=O	±		f
GLu-23	Thr-21	C=O	±		i	Asn-46	Cys-4	C=O	±	+	f
Ala-24	—				f	Asn-46	Tyr-44	C=O	±		f

\*The hydrogen bonds were calculated from the final set of 20 structures obtained with the “ensemble” IRMA protocol and indicated if found in a relevant number of structures.

† + indicates a hydrogen bond found with a maximum distance of 2.8 Å and a minimum angle of 135°. ± indicates a hydrogen bond found with a maximum distance of 3.0 Å and a minimum angle of 90°. f, s, i indicate that the exchange is fast, slow, and at an intermediate rate, respectively.

vation that crambin can form aggregates; since only hydrophobic side chains are found at this side of the molecule, we believe that this part of the surface may be favourable for dimerisation.

A stereoview of the superimposed solution and crystal structures is given in Figure 9. Backbone and C<sup>β</sup> atoms are presented with in addition the complete side chain of Tyr-29. No significant difference between solution and crystal structures is found except for the side chain of Tyr-29 which is almost parallel to the axis of the second helix in the solution structure but perpendicular to this latter in the crystal structure. Its position is defined in the solution structure by both *J*-coupling data and NOEs between the aromatic protons and the helix.

## DISCUSSION

Relaxation matrix approaches have become a common tool in structure refinement with NMR data (for review see Borgias et al.<sup>37</sup>). They have been essentially applied to DNA structure determination in a quest for more accurate structures.<sup>12,29,41–44</sup> Contradictory views on their usefulness for this purpose have been published.<sup>45,46</sup> We

have presented here an application to protein structure determination, introducing a new “ensemble” IRMA, and compared the performance this procedure with standard methods. NOEs are an ensemble property: several conformers can be present in solution and therefore the ensemble relaxation matrix approach should better reproduce the experimental situation, distance constraints obtained in this way reflecting both spin diffusion effects and the conformational averaging in solution. It should be possible to combine the “ensemble” IRMA with the time averaged distance restrained Molecular Dynamics proposed by Torda et al.<sup>47</sup> The applicability of this latter in our protocol will depend upon computational resources since longer simulation times are required, at least one order of magnitude longer than the period on which averaging is performed. We demonstrated with crambin that IRMA leads to better results than standard initial rates analysis: the more accurate distance constraints allow for a better structure determination and a better fit of the experimental NOEs.

Relaxation matrix calculations require dynamical and structural models for the biomolecule under

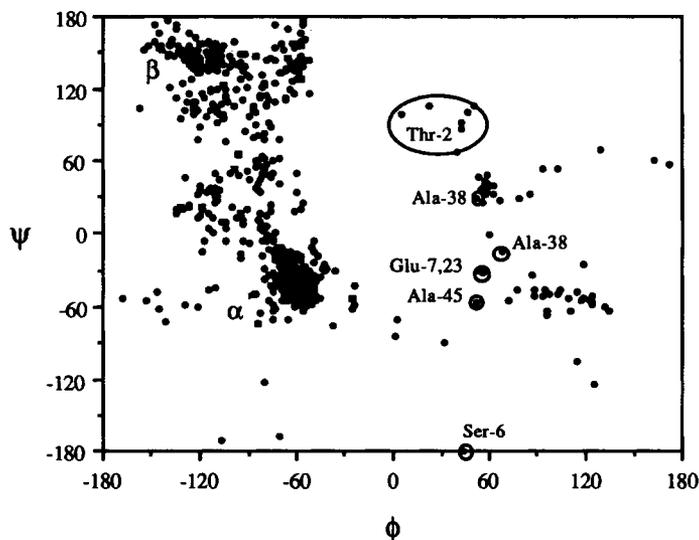


Fig. 8. Ramachandran plot of the  $\phi$ - $\psi$  dihedral angles of the 20 "ensemble" IRMA structures. Nonglycine residues found in the forbidden area are indicated in the figure.

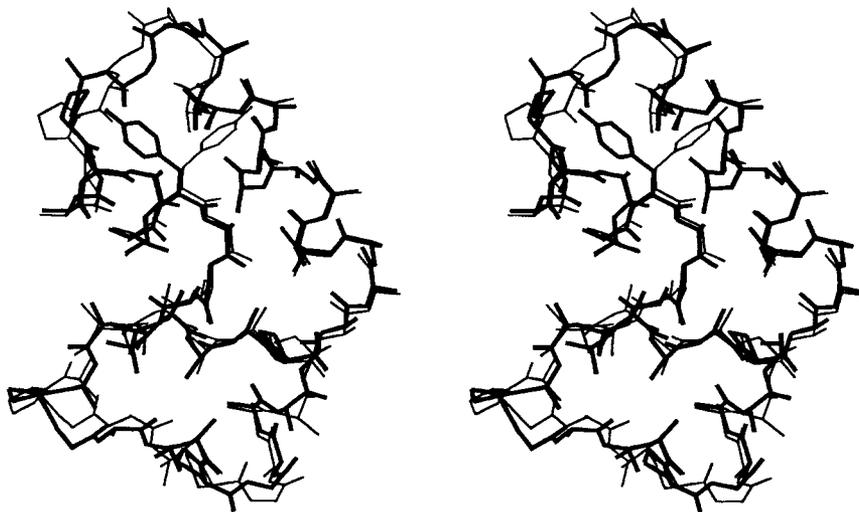


Fig. 9. Stereoview of the backbones and  $C^\beta$  atoms of the average "ensemble" IRMA structure (broad lines) and the X-ray structure of crambin. The complete side chain of Tyr-29 is also shown. The structures were fitted on backbone C, N,  $C'$  atoms.

consideration. Some dynamic processes like methyl group rotation, aromatic ring flips, and fast local motions can be described in a simple way and their effects have already been discussed.<sup>28,29</sup> The description of fast local motions in the NOE simulations via the order parameters has some limitations. The model of Lipari and Szabo<sup>30,31</sup> is valid only if internal motions occur at a much faster time scale than the overall rotation of the molecule. If no experimental relaxation data are available, order parameters can be obtained from simulations, but the parameters of such simulations should be carefully chosen, especially the relaxation time of the coupling with the temperature bath used in MD calcu-

lations to maintain a constant temperature. A too strong coupling could perturb the dynamics and the correlation functions computed from such trajectories will no more reflect the real dynamic behavior.<sup>34</sup> The starting structure and the length of the MD trajectory can also affect the results. In our case, the simulation was started from the crystal structure and in principle this might introduce a bias in the solution structures toward the X-ray structure. For crambin, however, the solution structure was already known to be very close to the crystal one and, moreover, the fluctuations in the free MD simulation were in the order of the difference observed between the solution and crystal struc-

tures (0.7 Å rmsd on backbone atoms from the average structure). Therefore, we do not think that any bias has been introduced in our results. In a more general refinement protocol, however, a structure obtained after the first IRMA cycle(s) could be used as starting point for the free MD simulation. The length of the simulation is another critical point in this approach. First we have to make sure that the system has equilibrated before using the trajectory for computing correlation functions. The 50 psec trajectory we used for this purpose may seem quite short, but, for crambin, the computed correlation functions reached a plateau within 5 to 10 psec and it has been shown from other work<sup>29</sup> that longer trajectories may generate somewhat lower  $S^2$  values without affecting too much the ratios between order parameter values for different parts of the molecule. Our experience showed that order parameters extracted from such short simulations already allow a better fit between the theoretical and experimental NOE intensities. Longer simulations may give better results, but, for larger systems, computational time can become a limiting factor.

Regarding the second requirement for relaxation matrix calculations, we investigated the influence of the structural model in IRMA by comparing distances obtained from a linear chain of crambin with those obtained from the X-ray structure. The inter-residue  $R$ -factors for these two structures were 0.93 and 0.41, respectively. The averaged deviation between the two sets of constraints was smaller than 0.15 Å, the largest differences being observed for long range contacts with a maximum of 0.65 Å for a  $\beta$ -sheet contact (Cys-4 H $^{\alpha}$  – Ile-33 HN), which is expected since short range interactions can already reasonably be satisfied in the extended structure as reflected by an intraresidue  $R$ -factor of 0.5. For crambin, all calculations were started from an extended chain for testing purpose and converged in a few IRMA cycles. This shows that the method is not sensitive to the starting model used. In practice, refinement could start from a better model, if available. Of course the performance will depend on the quality of the experimental data set.

Stereospecific assignment of prochiral groups is also an important problem in structure determination of proteins and several methods have been described to obtain additional assignments when experimental information fails.<sup>48–50</sup> Our approach makes use of the theoretical NOE intensities computed in IRMA. No contradictions with the experimental results based on  $J$ -coupling and NOE data were observed and, in general, the concordance was good between assignments predicted for the X-ray and the NMR structures, except for Ile-35 as can be seen from Table 1. Predicted assignments should depend more strongly on the model used in the calculations than the IRMA distances do. Therefore, the reliability of the method will increase with the qual-

ity of the model structure(s) and stereospecific assignments obtained from theoretical NOEs corresponding to the lowest  $R$ -factors should be preferred.

## CONCLUSION

The iterative relaxation matrix approach has proven a valuable method for structure determination of proteins. The method is quite robust and has good and fast convergence properties. By taking into account all indirect magnetisation transfer more accurate distance constraints can be derived from the NOE build-ups. The inclusion of internal motions and local mobility effects allows an interpretation of the NOE data in terms of dynamic systems rather than static structures, which, together with the “ensemble” approach, leads to a better fit of the experimental data and thus better distance constraints. The present set of constraints for crambin in combination with the GROMOS force field corresponds to one well-defined minimum, which is close to the crystal structure. Small differences are observed in the position of the side chain of Tyr-29 and in the loop between the two helices, the solution structure satisfying better the experimental exchange data for this region. Regions of higher mobility corresponding to turns in the structure are found in solution.

For a given set of NOEs, relaxation matrix approaches are a valuable tool toward more precise structures. The method presented in this paper allows us to obtain more accurate distances by back-transformation of NOE matrices in which the missing experimental information has been supplemented with intensities calculated from a model. We also have to mention a new type of refinement where NOE intensities are directly used as constraints without back-transformation to distances. This direct NOE refinement approach, first proposed by Yip and Case,<sup>51</sup> has already seen several implementations<sup>52–56</sup> and a first application to the refinement of a small protein<sup>56</sup> has resulted in an increased precision of the NMR solution structures. These direct approaches are, however, extremely time-consuming and should be used in a final refinement stage. Direct NOE refinement of our crambin structures has already shown that  $R$ -factors can be further minimized and an article on this work is presently in preparation.

## ACKNOWLEDGMENTS

The authors thank Miguel Llinás, Thea Koning, and Carlos Gonzales for many contributions to the project. A.B. acknowledges the Netherlands Organisation for Scientific Research (NWO) and the University of Lausanne (Switzerland) for financial support. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from NWO.

## REFERENCES

- Olejniczak, E.T., Gampe, R.T., Jr, Fesik, S.W. Accounting for spin diffusion in the analysis of 2D NOE data. *J. Magn. Reson.* 67:28-41, 1986.
- Boelens, R., Koning, T.M.G., Kaptein, R. Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach. *J. Mol. Struct.* 173:299-311, 1988.
- Kumar, A., Wagner G., Ernst, R.R., Wüthrich K. Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: Implications for studies of protein conformation. *J. Am. Chem. Soc.* 103:3654-3658, 1981.
- Hendrickson, W.A., Teeter, M.M. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature (London)* 220: 107-113, 1981.
- Teeter, M.M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. USA* 81:6014-6018 1984.
- Brünger, A.T., Clore, M.G., Gronenborn, A.M., Karplus, M. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to crambin. *Proc. Natl. Acad. Sci. USA* 83: 3801-3805, 1986.
- Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* 229:317-324, 1988.
- Teeter, M.M., Case, D.A. Harmonic and quasi-harmonic description of crambin. *J. Phys. Chem.* 94:8091-8097, 1990.
- Vermeulen, J.A.W.H., Lamerichs, R.M.J.N., Berliner, L.J., DeMarco, A., Llinás, M., Boelens, R., Alleman, J., Kaptein, R. <sup>1</sup>H NMR characterization of two crambin species. *FEBS Lett.* 219:426-430, 1987.
- Lamerichs, R.M.J.N., Berliner, L.J., Boelens, R., DeMarco, A., Llinás, M., Kaptein, R. Secondary structure and hydrogen bonding of crambin in solution. A two-dimensional NMR study. *Eur. J. Biochem.* 171:307-312, 1988.
- Lamerichs, R.M.J.N. 2D NMR studies of biomolecules: Protein structures and protein-DNA interactions. Ph.D. thesis, University of Utrecht, The Netherlands, 1989.
- Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H., Kaptein, R. Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson.* 82:290-308, 1989.
- Koning, T.M.G. IRMA: Iterative relaxation matrix approach for NMR structure determination. Application to DNA fragments. Ph.D. thesis, University of Utrecht, The Netherlands, 1990.
- Havel, T.F., Kuntz, I.D., Crippen, G.M. The theory and practice of distance geometry. *Bull. Math. Biol.* 45:665-720, 1983.
- Havel, T.F., Wüthrich, K. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurement of intramolecular <sup>1</sup>H-<sup>1</sup>H proximities in solution. *Bull. Math. Biol.* 46:673-698, 1984.
- van Gunsteren, W.F., Kaptein, R., Zuiderweg, E.R.P. Use of molecular dynamics computer simulations when determining protein structure by 2D-NMR. In: "Nucleic Acid Conformation and Dynamics." Olson, W.K., ed. Report of the NATO/CECAM workshop, Orsay, 1983: 79-92.
- Clore, M.G., Gronenborn, A.M., Brünger, A.T., Karplus, M. Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*. Combined use of <sup>1</sup>H nuclear magnetic resonance and restrained molecular dynamics. *J. Mol. Biol.* 186:435-455, 1985.
- Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data. Lac repressor headpiece. *J. Mol. Biol.* 182:179-182, 1985.
- Scheek, R.M., van Gunsteren, W.F., Kaptein, R. (1989) Molecular dynamics techniques for determination of molecular structures from nuclear magnetic resonance data. *Methods Enzymol.* 177:204-218.
- van Etten, C.H., Nielsen, H.C., Peter, J.E. A crystalline polypeptide from the seeds of *Crambe abyssinica*. *Phytochemistry* 4:467-473, 1965.
- States, D.J., Haberkorn, R.A., Ruben, D.J. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* 57:404-426, 1982.
- Marion, D., Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of <sup>1</sup>H-<sup>1</sup>H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* 113:967-974, 1984.
- Boelens, R., Scheek, R.M., Dijkstra, K., Kaptein, R. Sequential assignments of imino- and amino-proton resonances in <sup>1</sup>H NMR spectra of oligonucleotides by two-dimensional NMR spectroscopy. Application to a lac operator fragment. *J. Magn. Reson.* 62:378-386, 1985.
- van Gunsteren, W.F., Berendsen, H.J.C. Groningen Molecular Simulation (GROMOS) Library Manual. Biomos BV, Nijenborgh 16, 9747 AG Groningen, the Netherlands, 1987.
- Macura, S., Ernst, R.R. Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* 41:95-117, 1980.
- Olejniczak, E.T. Including methyl rotation in simulation of spin-lattice relaxation experiments. *J. Magn. Reson.* 81: 392-394, 1989.
- Landy, S.B., Rao, B.D.N. Dynamical NOE in multiple spin systems undergoing chemical exchange. *J. Magn. Reson.* 81:371-377, 1989.
- Koning, T.M.G., Boelens, R., Kaptein, R. Calculation of the nuclear Overhauser effect and the determination of proton-proton distances in the presence of internal motions. *J. Magn. Reson.* 90:111-123, 1990.
- Koning, T.M.G., Boelens, R., van der Marel, G.A., van Boom, J.H., Kaptein, R. Structure determination of a DNA octamer in solution by NMR spectroscopy. Effect of fast local motions. *Biochemistry* 30:3787-3797, 1991.
- Lipari, G., Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546-4559, 1982.
- Lipari, G., Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559-4570, 1982.
- Hyberts, S.G., Märki, W., Wagner, G. Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin C. *Eur. J. Biochem.* 164:625-635, 1987.
- Zuiderweg, E.R.P., Boelens, R., Kaptein, R. Stereospecific assignments of <sup>1</sup>H NMR methyl lines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* 24:601-611, 1985.
- Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R., Kaptein, R. Structure determination from NMR—application to crambin. In: "Computation of Biomolecular Structures: Achievements, Problems and Perspectives." Springer: Proceedings of a workshop held at Schloss Ringberg, march 1990, in press.
- Rullmann, J.A.C., Lamerichs, R.M.J.N., Gonzalez, C., Koning, T.M.G., Boelens, R., Kaptein, R. Structure determination from NMR using a relaxation matrix approach; application to the solution structure of crambin. In: "Modelling of Molecular Structures and Properties, Studies in Physical and Theoretical Chemistry," Vol. 71. Rivail J.-L., ed. Amsterdam: Elsevier, 1990:703-710.
- Gonzalez, C., Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R., Kaptein, R. Toward an NMR R factor. *J. Magn. Reson.* 91:659-664, 1991.
- Borgias, B.A., Cochin, M., Kerwood, D.J., James, T.L. Relaxation matrix analysis of 2D NMR data. *Prog. NMR Spectrosc.* 22:83-100, 1990.
- Kaptein, R., Boelens, R., Scheek, R.M., van Gunsteren, W.F. Protein structures from NMR. *Biochemistry* 27: 5389-5395, 1988.
- Holak, T.A., Gondol, D., Otlewski, J., Wilusz, T. Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance

- geometry and dynamical simulated annealing. *J. Mol. Biol.* 210:635–648, 1989.
40. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motions of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* 23:327–341, 1977.
  41. Metzler, W.J., Wang, C., Kitchen, D.B., Levy, R.M., Pardi, A. Determining local conformational variation in DNA. Nuclear magnetic resonance structures of the DNA duplexes d(CGCTAATCG) and d(CGTCACGCGC) generated using back-calculation of the nuclear Overhauser effect spectra, a distance geometry algorithm and constrained molecular dynamics. *J. Mol. Biol.* 214:711–736, 1990.
  42. Powers, R., Gorenstein, D.G. Two-dimensional  $^1\text{H}$  and  $^{31}\text{P}$  NMR spectra and restrained molecular dynamics structure of a covalent CPI-CDPI2-oligodeoxyribonucleotide decamer complex. *Biochemistry* 29:9994–10008, 1990.
  43. Powers, R., Jones, C.R., Gorenstein, D.G. Two-dimensional  $^1\text{H}$  and  $^{31}\text{P}$  NMR spectra and restrained molecular dynamics structure of an oligonucleotide duplex refined via a hybrid relaxation matrix procedure. *J. Biomol. Struct. Dynam.* 8:253–294, 1990.
  44. Nikonowicz, E., Meadows, R.P., Gorenstein, D.G. NMR structural refinement of an extrahelical adenosine tridecamer d(CGCGAATTCGCG) $^2$  via a hybrid relaxation matrix procedure. *Biochemistry* 29:4193–4204, 1990.
  45. Clore, G.M., Gronenberg A.M. How accurate can interproton distances in macromolecules really be determined by full relaxation matrix analysis of nuclear Overhauser enhancement data? *J. Magn. Reson.* 84:398–409, 1989.
  46. Kaluarachchi, K., Meadows, R.P., Gorenstein, D.G. How accurately can oligonucleotide structures be determined from the hybrid relaxation rate matrix/NOESY distance restrained molecular dynamic approach? *Biochemistry* 30: 8785–8797, 1991.
  47. Torda, A.E., Scheek, R.M., van Gunsteren, W.F. Time-averaged Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* 214:223–235, 1990.
  48. Güntert, P., Braun, W., Billeter, M., Wüthrich, K. Automated stereospecific  $^1\text{H}$  NMR assignments and their impact on the precision of protein structure determination in solution. *J. Am. Chem. Soc.* 111:3997–4004, 1989.
  49. Nilges, M., Clore, G.M., Gronenberg, A.M.  $^1\text{H}$ -NMR stereospecific assignments by conformational database searches. *Biopolymers* 29:813–822, 1990.
  50. Habazettl, J., Cieslar, C., Oschkinat, H., Holak, T.A.  $^1\text{H}$  NMR assignments of sidechain conformations in proteins using a high-dimensional potential in the simulated annealing calculations. *FEBS Lett.* 268:141–145, 1990.
  51. Yip, P., Case D.A. A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra. *J. Magn. Reson.* 83:643–648, 1989.
  52. Baleja, J.D., Moul, J., Sykes, B.D.J. Distance measurement and structure refinement with NOE data. *Magn. Reson.* 87:375–384, 1990.
  53. Mertz, J.E., Güntert, P., Wüthrich, K., Braun, W. Complete relaxation matrix refinement of NMR structures of proteins using analytically calculated dihedral angle derivatives of NOE intensities. *J. Biomol. NMR* 1:257–269, 1991.
  54. Bonvin, A.M.J.J., Boelens, R., Kaptein R. Direct NOE refinement of biomolecular structures using 2D NMR data. *J. Biomol. NMR* 1:305–309, 1991.
  55. Bonvin, A.M.J.J., Boelens, R., Kaptein, R. Direct structure refinement using 3D NOE-NOE spectra of biomolecules. *J. Magn. Reson.* 95:626–631, 1991.
  56. Nilges, M., Habazettl, J., Brünger, A.T., Holak, T.A. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* 219:499–510, 1991.