Journal of Peptide Science, Vol. 3, 133–140 (1997)

Constrained Refinement Based on NOE and Chemical Shift Information: The Monomer Form of Arginine–Vasopressin-like Insect Factor

BERNARD BUSETTA and PHILIPPE PICARD

Unité de Biophysique Structurale, Université Bordeaux II, France

Recieved 1 April 1996
Accepted 17 June 1996

Abstract: Via the refinement process of the monomer form of an arginine–vasopressin-like insect factor, the paper analyses the most relevant NMR information to define the solution structure of a flexible peptide. The relative importance of the different NOE constraints is discussed. © 1997 European Peptide Society and John Wiley & Sons, Ltd.

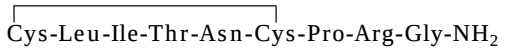
Keywords: NMR; chemical shift estimation; restrained refinement

INTRODUCTION

The simultaneous use of chemical shift and NOE information has already been suggested in order to obtain an accurate definition of the solution 3D structure of proteins [1,2] and peptides [3]. The final structure should provide an explanation for all the information recorded by the NMR spectroscopy.

Whereas NOE-derived constraints have been widely and successfully applied in protein structure determinations, flexible peptides are thought to provide too small a number of constraints to allow a good definition of their solution structure, and other sources of conformational informations are required.

In a previous paper [3], we emphasized that, in small peptides as in proteins [4], the information derived from NMR spectroscopy is more accurate when it corresponds to ordered parts of the molecule. Here we present the solution structure of the monomeric form of the diuretic arginine–vasopressin (AVP)-like insect (*Locusta migratoria*) factor. It consists of an hexapeptide ring closed by a disulphide bridge and extended by a C-terminal tripeptidic tail.



During the solution structure determination we discuss the relative importance which should be attributed to the different NOE and chemical shift-derived constraints.

MATERIALS AND METHODS

Synthesis

The chemical synthesis was performed on a Perkin Elmer peptide synthesizer model ABI 431A, using the solid-phase method with the tBoc strategy. The amino acid side-chain protections were as follows: benzyl for threonine, tosyl for arginine and 4-methylbenzyl for cysteines.

Elongation was achieved on N-α-protected glycine linked to a methylbenzhydrylamine (MBHA) resin (0.5 mmol/g) according to the operational cycle

Abbreviations: DSS: 2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt; COSY, correlated spectroscopy; HOHAHA, homonuclear Hartman–Hahn spectroscopy; NOESY, nuclear overhauser enhancement spectroscopy; ROESY, Rotating frame Overhauser enhancement spectroscopy.

Address for correspondence: Dr Bernard Busetta, Unité de Biophysique Structurale, CNRS ERS 133, 351 crs de la Libération, 33405 Talence Cédex, France.

CCC 1075-2617/97/020133-08 $17.50

protocol described earlier [5]. The peptide was cleaved from the resin by classical high HF treatment with anisol and ethanedithiol as scavengers. Good splitting of the MBHA resin and removal of the tosyl protection of arginine were achieved by allowing a reaction time of 90 min at 0° C. After HF evaporation, the crude peptide was precipitated with ether, extracted with 50% aqueous acetic acid and the extracts diluted with water and lyophilized.

The crude peptide cyclized by itself when left under stirring at room temperature and at pH 7.5 in water to yield the monomeric form of the AVP-like insect factor. The crude monomer was purified by semi-preparative HPLC performed with a C18-Delta Pak column (1.9 × 300 mm, 300 Å, 15 μm). The chromatographic conditions were the same as indicated for the analytical scale in Figure 1.

Amino acid analysis of the monomer was determined after hydrolysis in 6 M HCl (110° C, 12 h) using a Millipore-Waters Pico-Tag work station with amino acid standard H from Pierce. The result was as follows: Leu 1.13(1), Ile 0.99(1), Thr 1.03(1), Asx 0·98(1), Pro 1.08(1), Arg 0.95(1), Gly 1.08(1).

Purity of the monomer was checked by HPLC (Figure 1) and proved to be around 95% as judged from the 1D-NMR spectrum (Figure 2).

NMR Spectroscopy

1D and 2D NMR spectroscopy (COSY, HOHAHA and ROESY spectra) was achieved in water, on a Bruker DPX-400 spectrometer. Experiments were performed at 22° C, pH = 4.9 and DSS was used as the reference standard.

The observed proton chemical shifts are reported in Table 1 and the 1D spectrum of the amide protons in Figure 2. Every proton may be correctly assigned in the 1D spectrum except for the H_N proton of Leu_2 which gives a very weak and broad signal at 8.98 p.p.m. The Leu_2 spin system was unambiguously identified in the HOHAHA and ROESY 2D spectra. The line width and small intensity of the signal recorded for the H_N proton of the leucine suggest a relatively fast exchange between this proton and the water protons.

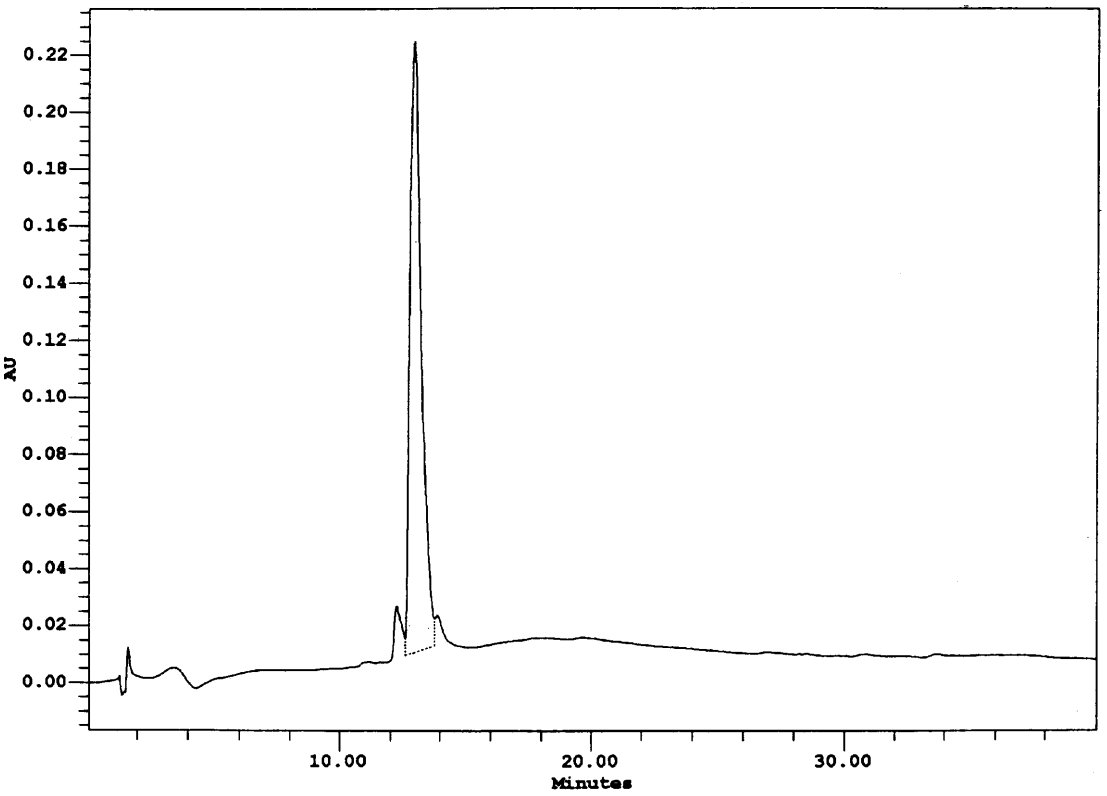


Figure 1 Reversed-phase HPLC chromatogram of AVP-like insect factor monomer. Conditions: linear gradient of 10–40% CH_3CN in 0.08% TFA for 30 min at a flow rate of 1 ml/min.

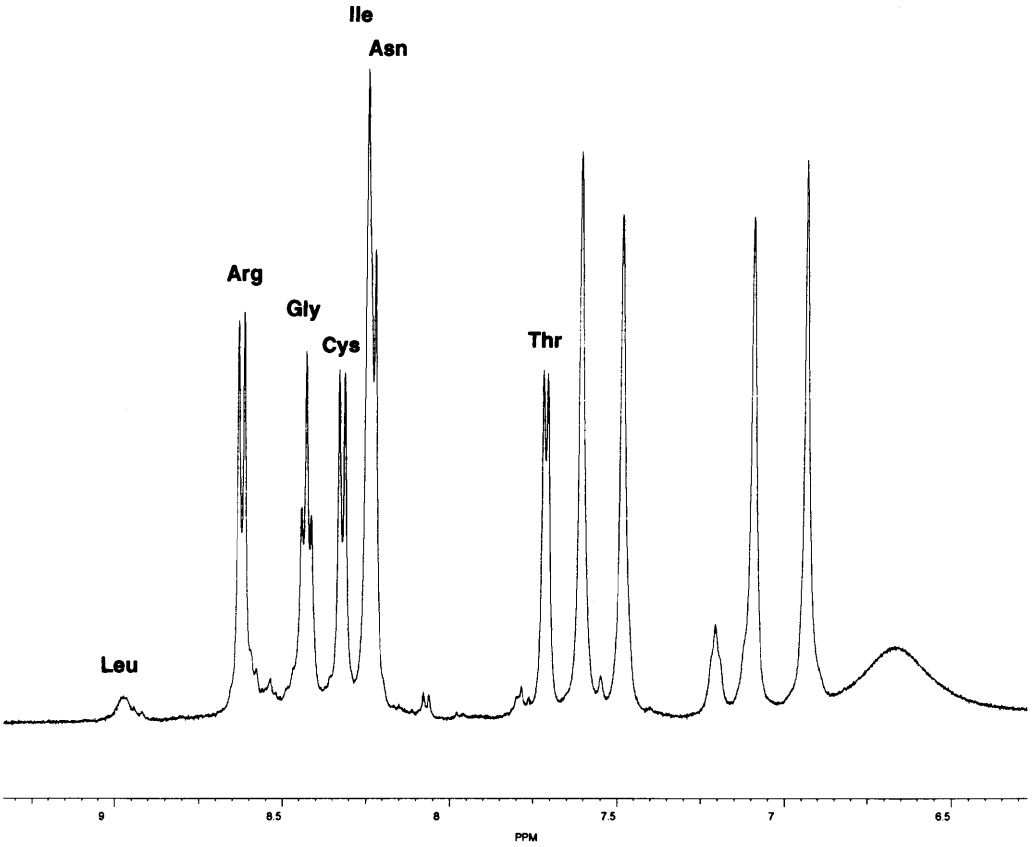


Figure 2 H_N proton region of the 1D ^{1}H spectrum of the AVP-like monomer. Each H_N proton is quoted by its residue name. The position of the Leu_2 H_N proton was unambiguously determined from the 2D HOHAHA spectrum.

Table 1. H_1 Chemical Shift (in ppm) for the AVP-like Monomer Recorded in Water at 22 °C, pH = 4.9

	H_N	H_α	H_β	H_γ	Others
Cys_1		4.29	3.30, 3.48		
Leu_2	8.98	4.54	1.68	1.68	H_δ 0.92, 0.94
Ile_3	8.25	4.20	2.03	1.22, 1.45	$H_{\gamma 2}$ 0.98 H_δ 0.92
Thr_4	7.71	4.08	4.22	1.24	
Asn_5	8.24	4.79	2.85		H_δ 6.93, 7.60
Cys_6	8.33	4.88	2.95, 3.25		
Pro_7		4.46	1.92, 2.31	2.04	H_δ 3.73, 3.76
Arg_8	8.63	4.31	1.78, 1.90	1.67	H_δ 3.22 H_ε 7.21
Gly_9	8.44	3.92			

From the ROESY spectrum, 26 intra-residue constraints and 16 constraints between adjacent residues were defined. Applying the constraints derived from the ROE cross peaks, the DIANA program [6] was used to generate 1000 random independent conformers. Then only the top 50 conformers with the lowest value of the DIANA target function were submitted to a classical constrained energy refinement involving 200 cycles of Powell's conjugate gradient process using the X-PLOR package [7].

Estimation of Proton Chemical Shifts from 3D Structures

The general protocol defined to estimate the secondary chemical shifts of the AVP-like parallel dimer [3] was used in this study.

For a peptide without any aromatic side chains, the secondary chemical shifts are expresses as the sum:

$$\Delta\delta = \delta_{\mathrm{m}} + \delta_{el} \qquad (1)$$

where δ_{m} and δ_{el} respectively represent the contributions due to the magnetic anisotropies of neighbouring amide groups and to the electrostatic effects. Both terms are estimated with the empirical parameters which were derived by Osapay and Case [8] from NMR protein structures.

When a H_N proton is involved in a hydrogen bond, an additive term is used in the estimation of its secondary chemical shift [9]:

$$\delta_{\mathrm{m}} = -0.92 + 7.6/r^3 \text{ p.p.m.} \qquad (2)$$

where r is the distance (in Å) between the H_N proton and the corresponding oxygen atom.

Owing to the random torsion around the valence bonds on each side of the carrying nitrogen or carbon atom, thermal vibrations of any proton are approximated by the use of a randomization factor p [3] in Eq. (1):

$$\Delta\delta = p(\delta_{\mathrm{m}} + \delta_{el}) \qquad (3)$$

where p is defined as the value giving the best fit between the observed and computed secondary chemical shifts. In comparison with the main-chain protons, a weaker randomization factor (i.e. a greater thermal vibration) is attributed to the side-chain protons. For the monomeric AVP-like insect factor, for which the disulphide-bride formation involves greater constraints on the polypeptide chain than in the flexible dimer peptide, $p = 0.65$ in spite of the small size of the molecule.

The average difference between the observed and computed secondary chemical shifts of all the H_N, H_α and cysteine H_β protons (further referred to as the test protons) was used to estimate the reliability of the different refined conformers.

Chemical Shift-dependent Constraints [3]

When a set of different conformers is generated, the one displaying the best estimate for the chemical shift of a given proton is used to define the optimum chemical environment of this proton as a set of distances between the proton and its neighbouring atoms.

The optimum chemical environment of each proton determines a set of local chemical shift-dependent distance constraints. The result is added to the NOE distance constraints to improve the definition of the solution structure.

Concerning the AVP-like monomer, for which no ring current effects occur, the required constraints are limited to distances between every proton and the surrounding carbonyl atoms.

Secondary Chemical Shifts of the N-terminal Residue

In a previous paper [3] we already noticed that the most stable part of a peptide gives the most reliable NMR information. In the case of the AVP-like peptide, the disulphide bridge represents the less flexible portion of the molecule. To determine the peptide solution structure, a correct estimation of the average chemical environment of each cysteine proton is required. When correctly estimated, the secondary chemical shifts of these protons may bring some information about their chemical environment. In the basic estimation process [8], the proton chemical shifts are estimated by reference to a 'random coil' conformation which is supposed to approximate 'diamagnetic' and 'paramagnetic' contributions of the local average environment. These local contributions are defined as the observed chemical shifts for each X residue of a linear tetrapeptide such as H-Gly-Gly-X-AlaOH measured in H_2O at 36 °C and pH = 6.0 [10]. In this tetrapeptide, both N and C terminal atoms of residue X are involved in a peptide bond. The reference chemical shifts are defined for a residue X submitted on each side to the magnetic anisotropies of a peptide group. In our study, the absence of an N terminal peptide group should be taken into account to estimate the chemical shifts of the N terminal cysteine protons.

From the linear tetrapeptide H-Gly-Gly-Cys-AlaOH [10], the reference chemical shifts are 4.68 p.p.m. for the H_α proton and 2.96/3.28 p.p.m. for each H_β protons of the cysteine residue.

The isolated reduced l-cysteine [11] displays signals at 3.95 and 3.02 p.p.m. for the H_α and H_β protons respectively. These chemical shifts are similar to those observed for the reduced human AVP peptide [12] which is supposed to adopt a random coil conformation (4.15 and 2.96/2.99 p.p.m.). In this case, the different protons are also submitted to ring current effects induced by

tyrosine and phenylalanine aromatic residues in positions 2 and 3. On the contrary, the chemical shifts of the Cys_6 protons (4.76 and 2.84/2.90 p.p.m.) reproduce the reference shifts described in [10].

In endothelin where a cysteine residue occurs in position 1, as in the AVP-like monomer, the observed chemical shifts (3.91 and 3.03/3.16 p.p.m.) [12] are also in agreement with the variations derived from the observed chemical shifts reported in Table 1 (4.29 and 3.30/3.48 p.p.m.).

Thus, in comparison with the reference chemical shifts [10], we estimated that the absence of an N-terminal peptide group involves a 0.50 p.p.m. upfield shift for the H_α proton and a 0.15 p.p.m. downfield shift for the H_β protons.

RESULTS AND DISCUSSION

Constrained Refinements

In order to determine the most probable solution structure of the monomeric form of the AVP-like insect factor, three different refinements were applied to the top 50 conformers provided by the DIANA process [6].

Use of Full NOE Constraints. All the 42 distance constraints derived from the ROESY spectrum are used to define the upper limits of the proton–proton distances. The lower limits were set as the sum of the van der Waals radii (1.90 Å). The constrained energy of the refined conformers stretched over 20 kcal and their mean average distance (r.m.s.d.) is 1.20 Å. There is no NOE violation greater than 0.1 Å. The average error ($\Delta\delta$) on the secondary chemical shifts of the test protons varies from 0.130 to 0.190 p.p.m. but is not energy-dependent as in the case of the parallel AVP-like dimer [3]. For five conformers among the top 50 ones the error on the estimation of the secondary chemical shifts of the test protons ($0.130 < \Delta\delta < 0.144$) differs significantly from the error associated with the other conformers. These five conformers were subsequently used to define the optimum chemical environment of the H_α and $H_{\beta cys}$ protons and to generate the chemical shift-dependent constraints.

Among the 50 conformers, 25% of them present a short hydrogen bond between the Asn_5 H_N proton and the Cys_1 carbonyl group. For these conformers the overall peptide backbone is somehow distorted and the average error, on the H_α protons, is higher (> 0.180 p.p.m.) than for the other conformers. For two conformers, an extra hydrogen bond links the Leu_2H_N proton with the Thr_4 carbonyl group.

At this point of the refinement, the solution structure of the AVP-like monomer might be defined as a quite flexible structure with the H_N proton of Asn_5 partially engaged in an intramolecular hydrogen bond or exposed to the surrounding solvent.

Use of NOE and Chemical Shift-dependent Constraints. The chemical shift-dependent constraints, defined during the previous refinement, were added to the NOE constraints and each conformer was then re-submitted to 200 cycles of Powell refinement. The five best solutions were initially, on average, 0.90 Å apart from one another. After refinement, the average distance between the refined conformers was reduced to 0.25 Å and the error on the estimation of the secondary chemical shifts of the test protons drops to $0.120 < \Delta\delta < 0.125$ p.p.m. This group of conformers will be further referred to as conformer A_1.

If we analyse the NOE violations at this stage of the refinement, we must notice that there is always no violation greater than 0.1 Å for the constraints which involve the main-chain protons but some violations occur for constraints between main-chain and side-chain protons.

Use of a Reduced Number of Tight NOE Constraints. As already observed in protein structure determination, even a single incorrectly estimated NOE constraint can be disastrous (see for instance [13]) and involves a spurious convergence of the constrained refinement process. NOE constraints between side-chain and main-chain protons are likely to be wrongly estimated owing to the side-chain flexibility. Flexibility means that all the distances obtained from NOE intensities appear to be somewhat shorter than the real ones [14]. However, we may assume that the 13 observed NOE constraints between main-chain protons are correctly evaluated. They define the upper limits of the proton–proton distances. The lower limits, greater than the sum of the van der Waals radii, were assumed to deviate from the corresponding upper limits by no more than 0.40 Å. The 50 best initial conformations which were generated by the DIANA process, were refined by 200 cycles of constrained energy using this reduced set of NOE constraints. The estimation of the test proton chemical shifts was used, as previously, to appraise the reliability of the refined structures. Two groups of five conformations emerge from the refinement:

Table 2. List of NOE Violations Greater than 0.1 Å

	NOE distances	Interproton distances for Conformers A_1 (full NOE constraints) (Å)	Interproton distances for Conformers A_2 or B[a] (reduced constraints) (Å)
Thr_4H_N-Asn_5H_N	2.80	1.90[b]	No violation
Leu_2H_α-Ile_3H_N	2.90	3.07	3.11
Ile_3H_α-Thr_4H_N	2.50	2.94	3.02
Arg_8H_α-Arg_8H_β	2.50	2.94	3.02
Cys_6H_N-$Cys_6H_{\beta 2}$	3.10	3.26	3.44
Ile_3H_α-Ile_3H_β	2.30	2.88	3.01

[a] The three last NOE constraints between side-chain and main-chain protons are not taken into account in the constrained energy refinement of conformer B.
[b] Would be an NOE violation when the lower limit is fixed to (upper limit -0.40 Å).

1. The first one, A2, does not significantly deviate from the solution structures previously defined as conformer A_1. The error on the estimation of the secondary chemical shifts within group A_2 drops to $0.112 < \Delta\delta < 0.118$ p.p.m.

2. In the other group (B) the hydrogen bond between the Asn_5 proton and the Cys_1 carbonyl group is retrieved. Within this group the average distance between the conformers is 0.30–0.40 Å. In contrast the average distance between the group conformer A_2 and B is 0.70 Å. For conformers B the error on the estimation of the secondary chemical shifts of test protons is $\Delta\delta = 0.110$ p.p.m.

In Table 2, we compare the violations of NOE constraints between the initial conformer 'A_1' refined with full set of NOE constraints and conformers A_2 and B refined with the reduced set of NOE constraints. The same kind of violation occurs to every conformer. In the last refinement, the structural improvement was obtained by reducing the number of NOE constraints: a greater accuracy was assigned to the definition of the constraint between the main-chain Thr_4 H_N and Asn_5 H_N proton (2.40–2.80 instead of 1.90–2.80 Å). On the contrary, NOE constraints between side-chain and main-chain protons which appear to be largely overestimated were avoided.

Analysis of the H_N Proton Chemical Shifts

At present, the different contributions which determine the secondary chemical shifts of H_N protons are not yet completely understood. The influence of the surrounding solvent and the intramolecular hydrogen bonds may be the main factors.

Table 3. Comparison of the Error (in p.p.m.) in the Estimation of the Secondary Chemical Shifts of the H_N Protons for Conformers A and B, and their Estimated Accessibility Expressed as the Number (No.) of Atoms in an 8 Å Radius Sphere (the Weaker this Number is, the More Exposed the H_N Proton is)

	Conformer A		Conformer B	
	δ_{error}	No.	δ_{error}	No.
H_NLeu_2	−0.490	38	−0.456	16
Ile_3	0.065	29	0.078	30
Thr_4	0.020	39	−0.040	45
Asn_5	0.193	44	(H bonded)	
Cys_6	−0.015	29	0.031	23
Arg_8	−0.228	14	−0.224	14

The error in the estimation of the secondary chemical shifts of conformers A and B is greater for H_N protons (0.150–0.160 p.p.m.) than for H_α protons (0.080–0.090 p.p.m.). The largest discrepancy is observed for the H_N proton of Leu_2, the downfield shift (0.530 p.p.m.) of which cannot be explained by eq. (1). In Table 3, we report, for both conformers, the error in the estimation of each H_N protons and we compare this error with its accessibility appraised by the number of atoms located in an 8 Å radius surrounding sphere.

Besides the existence of a hydrogen bond between the H_N proton of Asn_5 and the Cys_1 carbonyl group in conformers B, there is another obvious difference between the two conformers: the H_N proton of Leu_2 is solvent exposed in conformers B (only 16 surrounding atoms) and buried in conformers A (38 surrounding atoms).

When a H_N proton is fully exposed (as for instance Arg_8 H_N) the potential hydrogen bond is shorter than

the average hydrogen bond length in water, and the chemical shift is moved downfield as may be expected from eq. (2). On the contrary, the chemical shift of a buried H_N proton is expected to move upfield. The chemical shift of Leu_2 H_N proton moves downfield and therefore is expected to be fully exposed to solvent as in conformers B. This assumption is in agreement with the relatively fast exchange between this proton and the water protons suggested by 1D spectrum (Figure 2). Moreover in the related human AVP peptide, the H_N proton of residue 2 (Tyr) in the cyclic monomer is moved downfield in comparison with the reduced extended conformation [15].

CONCLUSIONS

The chemical shifts, calculated from high-resolution crystal structures, are usually observed to be significantly closer to the experimental values than those calculated from the corresponding NMR structures [2,16]. This observation suggests some failure in the usual NOE-constrained refinement.

In this paper, we tried to determine the most relevant NMR observations in order to define the solution structure of a flexible peptide. For this purpose, we adopt the molecular description currently used in X-ray analysis: one average structure with individual thermal motion for each atom (called here the individual randomization factor). The average structure is defined as the 3D structure which gives the best estimate of secondary chemical shifts. Whereas the use of the complete set of NOE distance constraints predicts an undefined flexible structure of the AVP-like monomer, the structural interpretation of the observed chemical shifts leads to a very restricted structure (conformers B) with a weak dispersion (0.30 Å) (see Figure 3).

If any piece of information provided by the NMR spectroscopy is of importance during the structure determination, it must be introduced at the right time during the refinement process. All the distance constraints, derived from NOE experiments, are required to define an approximate 3D solution structure, but distance constraints involving side chain protons are usually overestimated. The constrained energy refinement would converge towards a more reliable structure if only NOE and chemical shift-derived constraints are applied to the main-chain protons. Then the backbone polypeptide chain should remain in fix position when NOE constraints, involving the side-chain protons, are used to define the side-chain conformation.

The solvent accessibility of the H_N protons and their ability to form hydrogen bonds appear to be the main contribution which determines their secondary chemical shifts. To ensure a correct estimation of the chemical shift, a very accurate definition of the intramolecular hydrogen bond or the average hydrogen bond with the surrounding solvent molecule should be performed.

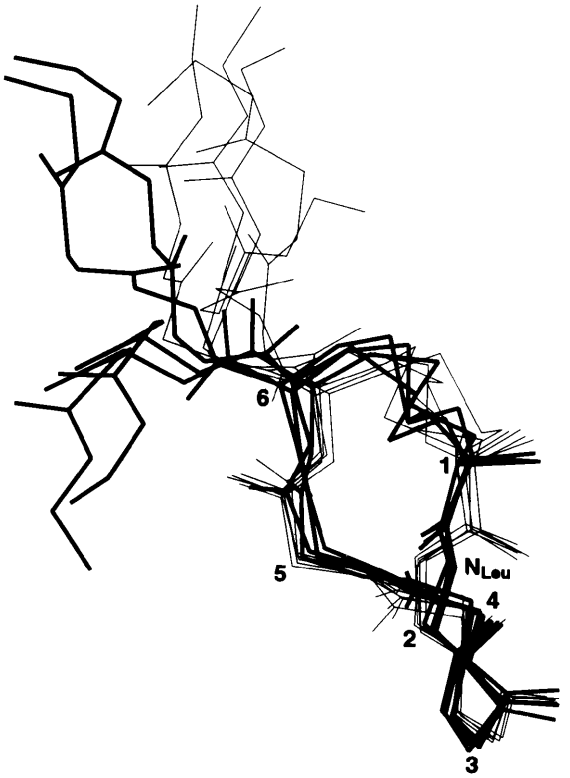


Figure 3 Superposition of the side chain of residues Cys_1 and Cys_6 and the backbone atoms of conformers A_2 (thin lines) and B (thick lines). The position of the H_N proton of residue Leu_2 (solvent exposed in B and buried in A_2) is indicated. For the N terminal ring the residue number is indicated in front of the position of each Cα.

Acknowledgements

We thank the CESAMO at the University of Bordeaux I, for NMR data collection facilities.

REFERENCES

1. H. J. G. Le Pearson, A. C. de Dios and E. Oldfield (1995). Protein structure refinement and prediction via NMR chemical shifts and quantum chemistry. *J. Am. Chem. Soc. 117*, 3800–3807.

2. J. Kuszewski, A. M. Gronenborn and G. M. Clore (1995). The impact of direct refinement against proton chemical shifts of protein structure determination by NMR. *J. Magn. Reson. 107*, 293–297.
3. B. Busetta and P. Picard (1996). The use of proton chemical shifts to define the solution of a dimeric peptide. *J. Peptide Sci. 2*, 233–239.
4. D. S. Wishart, B. D. Sykes and F. M. Richards (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol. 222*, 311–333.
5. J. Girardie, A. Girardie, A. Van Dorsselaer, O. Sorokine, S. Geoffre, M. Hospital and G. precigoux (1995). Chemical synthesis of a 7 kDa insect gonadotropic hormone. *J. Peptide Sci. 1*, 311–318.
6. P. Güntert, W. Braun and K. Wüthrich (1991). Efficient computation of three dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol. 217*, 517–530.
7. A. T. Brünger (1992). *X-PLOR, Version 2.1 Manual*, Yale University, New Haven, CT.
8. H. Osapay and D. A. Case (1991). A new analysis of proton chemical shifts in protein. *J. Am. Chem. Soc. 13*, 9436–9444.
9. A. Pardi, G. Wagner and K. Wüthrich (1983). Protein conformation and proton nuclear magnetic resonance chemical shifts. *Eur. J. Biochem. 137*, 445–454.
10. A. Bundi and K. Wüthrich (1979). ^{1}H-NMR parameters of the common amino-acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-Ala-OH. *Biopolymers 18*, 285–297.
11. C. C. McDonald and W. D. Phillips (1969). Proton magnetic resonance spectra of proteins in random coil configurations. *J. Am. Chem. Soc. 91*, 1513–1521.
12. C. K. Larive, L. Guerra and D. L. Rabenstein (1992). *Cis-trans* conformational equilibrium across the cysteine^{6}-proline peptide bond of oxytocin, arginine vasopressin and lysine vasopressin. *J. Am. Chem. Soc. 114*, 7331–7337.
13. N. H. Andersen, C. Chen, T. M. Marschner, S. R. Krystek and D. a. Bassolino (1992). Conformational isomerism of endothelin in acidic aqueous media: a quantitative NOESY analysis. *Biochemistry 31*, 1280–1295.
14. T. M. G. Koning, R. Boelens and R. Kaptein (1990). 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.
15. C. K. Larive and D. L. Rabenstein (1991). Two dimensional ^{1}H NMR spectroscopy of aqueous solutions with elimination of the water resonance by the transverse relaxation: application to assignment of the ^{1}H spectrum of reduced arginine vasopressin. *Magn. Reson. Chem. 29*, 409–417.
16. M. P. Williamson, T. Asakura, E. Nakamura and M. Demura (1992). A method for the calculation of protein α-CH chemical shifts. *J. Biomol. NMR 2*, 83–98.

