

^1H NMR characterization of two crambin species

J.A.W.H. Vermeulen⁺, R.M.J.N. Lamerichs, L.J. Berliner[°], A. De Marco[†], M. Llinás*,
R. Boelens, J. Alleman and R. Kaptein

Department of Organic Chemistry, University of Utrecht, Padualaan 8, 3584 CH Utrecht,

⁺ Philips Medical Systems, PO Box 218, 5600 MD Eindhoven, The Netherlands,

[°] Department of Chemistry, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA,

[†] Istituto di Chimica delle Macromolecole del CNR, Via E. Bassini 15, I-20133, Milano, Italy and

** Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA*

Received 2 June 1987

Crambin displays amino acid heterogeneity at positions 22 (Pro or Ser) and 25 (Leu or Ile). Using reversed phase HPLC the crambin mixture can be resolved into two protein fractions. It is shown by amino acid analysis and NMR spectroscopy that these fractions represent single proteins (Ser-22/Ile-25 and Pro-22/Leu-25 species). A first characterization of the ^1H -NMR spectra of these species is presented.

Crambin; HPLC; Protein purification; Amino acid analysis; ^1H NMR; COSY; 2D NMR

1. INTRODUCTION

Crambin is a water-insoluble protein of M_r 4700 isolated from the seeds of *Crambe abyssinica* [1]. It has one of the best refined crystal structures [2,3] and has therefore served as a model in a number of protein structural studies [4,5]. Previous NMR studies of crambin have focussed mainly on the methyl region [6] and aromatic region [7,8] of the spectrum and led to some partial resonance assignments. Effects of solvent and temperature on the NMR spectrum indicated that crambin has a high thermal stability [8,9].

We have embarked on a full structural study of crambin using the methods of two-dimensional NMR [10]. The aim is to test recently developed methods for structure determination and refinement based on NMR data [10–13] and to compare

the solution structure of crambin with that in the crystal. One of the potential difficulties for an accurate structure determination is the compositional heterogeneity of the protein: Pro and Ser can occur at position 22 and Ile and Leu at position 25 [3]. Accordingly, crambin preparations are generally assumed to consist of 4 species with the Pro/Ile form as the most abundant one. In this letter we show that in fact only two crambin species occur, the Ser/Ile and the Pro/Leu forms in a 55:45 ratio. We show that these species can be isolated in pure form and present a first characterization of their ^1H -NMR spectra.

2. MATERIALS AND METHODS

2.1. Protein purification

Crambin was isolated according to the procedure of Van Etten [1]. All solvents were HPLC or sequence grade. Acetone- d_6 was from Janssen Chemicals and $^2\text{H}_2\text{O}$ (99.99%) was from Merck.

Crambin species were resolved by HPLC on a RV5 Chrompack column (250 \times 4.6 mm) packed

Correspondence address: R. Kaptein, Department of Organic Chemistry, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

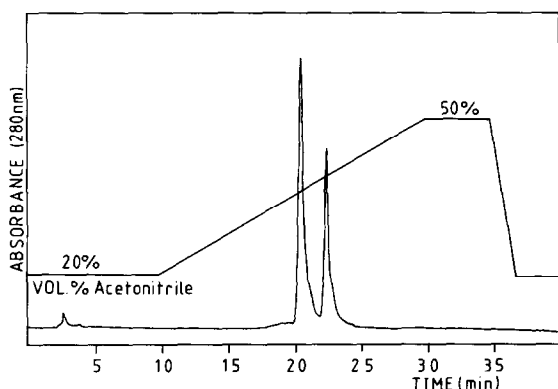


Fig.1. HPLC elution profile of the crambin mixture. A linear gradient from 20% to 50% acetonitrile-water is shown. Fractions are indicated I and II according to the order in which they are eluted from the column.

with Nucleosil 10 (C18). Elution was programmed over a linear gradient from 20% (v/v) CH₃CN/H₂O to 50% (v/v) CH₃CN/H₂O with UV-detection.

Table 1

Amino acid composition of crambin species^a

	Fraction I	Fraction II
Asp	4.3 (4)	4.2 (4)
Thr	6.1 (6)	6.0 (6)
Ser	2.9 (3)	2.1 (2)
Glu	1.0 (1)	0.8 (1)
Gly	4.4 (4)	4.3 (4)
Ala	5.3 (5)	5.1 (5)
Cys	5.3 (6)	5.4 (6)
Val	1.9 (2)	2.0 (2)
Ile	4.3 (5)	3.3 (4)
Leu	1.1 (1)	2.1 (2)
Tyr	1.8 (2)	1.9 (2)
Phe	1.0 (1)	1.0 (1)
Arg	2.2 (2)	2.1 (2)
Pro	4.4 (4)	5.4 (5)

^a Values measured for HPLC fractions I and II (see section 2). The numbers in parentheses are deduced assuming that these fractions contain single protein species (cf. [1])

For determination of the amino acid composition the HPLC fractions (diluted to 10 $\mu\text{mol} \cdot \text{l}^{-1}$) were hydrolyzed in 6 N HCl at 110°C for 48 h and then analyzed with a Kontron amino acid analyser.

2.2. NMR spectroscopy

Crambin was dissolved in a 3:1 mixture of ²H₆-acetone and ²H₂O for the NMR experiments. These experiments were performed at 303 K, 500 MHz on a Bruker AM 500, interfaced to an Aspect 3000 computer. The 1D and 2D spectra were obtained with presaturation of the residual water

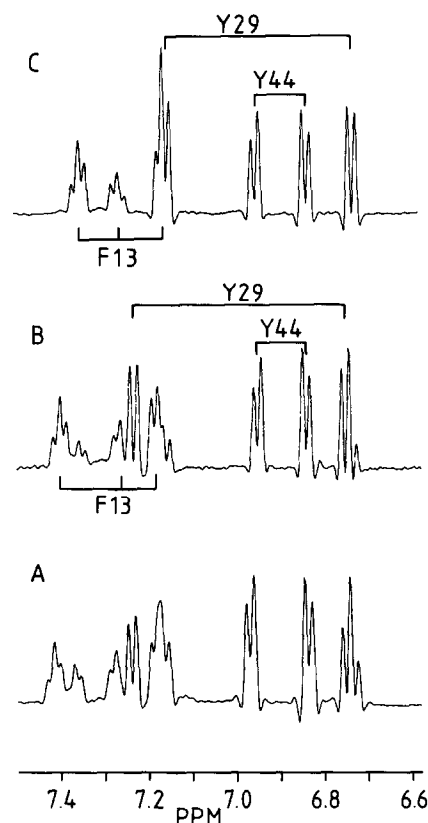
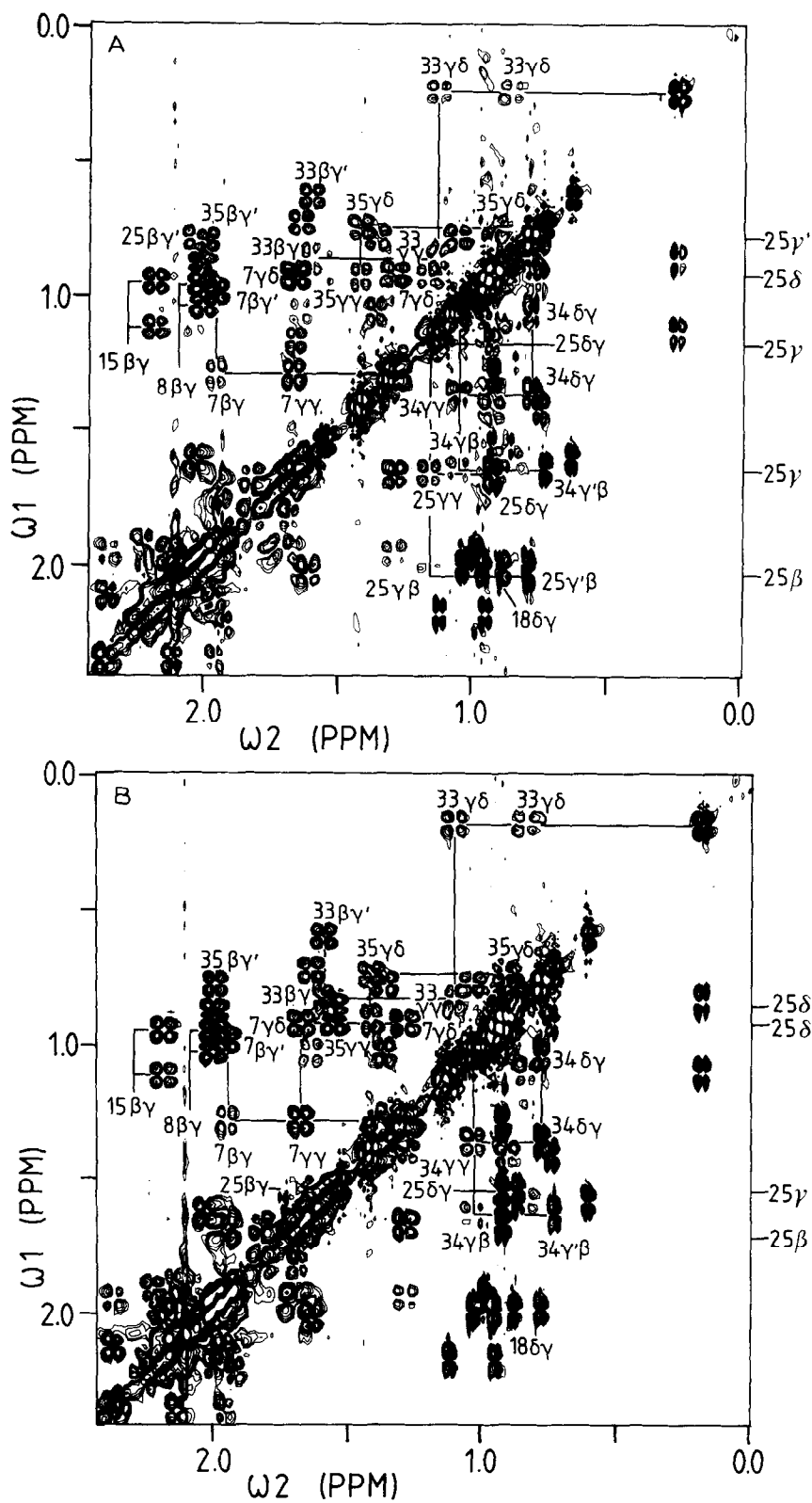


Fig.2. Aromatic region of the 500-MHz ¹H NMR spectra of (A) crambin mixture, (B) fraction I (Ser/Ile crambin), and (C) fraction II (Pro/Leu crambin). Assignments of Tyr 29, Tyr 44, and Phe 13 protons are indicated.

Fig.3. 500-MHz ¹H double quantum filtered COSY spectrum of (A) fraction I (Ser/Ile crambin) and (B) fraction II (Pro/Leu crambin). The region containing all cross-peaks involving methyl protons is shown. Complete assignment of these cross-peaks is indicated either on one side or on the other side of the diagonal. The resonance positions of the side chain protons of Ile 25 (A) and Leu 25 (B) are indicated on the right side of the spectra. The Ile γ -protons (A) and the Leu δ -protons (B) were only pairwise assigned.



resonance. The double-quantum filtered COSY spectra [14] were recorded with the time-proportional-phase-increment method [15]. For each spectrum the time domain data consisted of 512 FIDs of 2K data points with t_1 values ranging from 0 to 38.4 ms; 64 scans were accumulated for each t_1 value. The delay between the experiments was varied randomly between 1200 and 1400 ms for each t_1 value, for suppression of double quantum coherences evolving during the t_1 period. The rf pulses and receiver phase were cycled according to a 32-step scheme in order to obtain the double quantum coherence transfer and to suppress axial peaks. The time domain data were weighted with a sine-bell shifted by $\pi/3$ in both t_1 and t_2 domains. The baselines in the (ω_1, ω_2) data set were corrected in ω_1 with an automatic baseline correction using a third order polynomial. The resulting pure absorption phase spectra have a digital resolution of 6.67 Hz/point in both ω_1 and ω_2 .

3. RESULTS AND DISCUSSION

3.1. HPLC separation and amino acid analysis

HPLC separation of the crambin mixture results in two fractions, I and II, in an approximate ratio of 55:45; a typical elution pattern is shown in fig.1. To characterize these protein fractions amino acid analyses were carried out. The results are shown in table 1. First, it can be noted that the HPLC separation, which is likely to be based on the Pro/Ser substitution at position 22 and not at the Leu/Ile substitution at 25, has yielded a more hydrophobic Pro-22 species in fraction II as expected for the C18 column used. As to the Leu/Ile heterogeneity, the data of table 1 strongly indicate that fraction II consists of a pure Leu-25 species and that fraction I has Ile at position 25. For leucine this is particularly clear, while for isoleucine the yields are somewhat low. This is, however, presumably due to inefficient hydrolysis due to steric crowding [16] in the region 33–35, where crambin has three consecutive isoleucines. For an alternative possibility assuming four crambin species, two in each HPLC fraction and separation on the basis of Pro/Ser, one would expect the Leu/Ile ratios in the two fractions to be the same, which is clearly not observed. Thus, it appears that the crude crambin mixture separates into two single proteins, Ser-22/Ile-25 (fraction I) and

Pro-22/Leu-25 (fraction II). As is shown below this is confirmed by the NMR measurements.

3.2. NMR spectra

Previously, it has been observed [6–9] that the amino acid heterogeneity causes a splitting of many resonances in the ^1H -NMR spectrum of crambin. As shown in fig.2A in the aromatic region two sets of resonances are observed for Phe-13 and for one of two tyrosines. Previously, the aromatic region of the spectrum had been simulated in terms of two components [7]. On the basis of the spatial proximity to the Pro/Ser site at position 22, it had been suggested that the split tyrosine is Tyr-29 [9] and this is now confirmed by a full sequential assignment of the crambin ^1H resonances, which will be reported elsewhere. From the spectra of the fractions I and II shown in fig.2B and C, respectively, it is now clear which lines belong to the Pro/Leu and which to the Ser/Ile species. Thus, the aromatic region of the ^1H -NMR spectrum of the crambin mixture can be fully accounted for on the basis of a mixture of two crambin species (note that fraction I, fig.2B, contains the protein of fraction II as a contamination of about 15%).

Of course, the Leu/Ile substitution at position 25 may not affect the chemical shifts of the aromatic residues, so it is necessary to examine the aliphatic protons as well. Since this region of the spectrum is crowded, containing many overlapping resonances, it can best be studied in a two-dimensional NMR spectrum. Fig.3 shows parts of the double quantum filtered COSY spectra containing the methyl resonances of crambin for the fractions I and II. All cross-peaks involving methyl protons could be assigned assuming that fractions I and II contained single Ser-22/Ile-25 and Pro-22/Leu-25 proteins (apart from the 15% contamination of fraction I as discussed above). In particular, for Ile-25 in fraction I the β , γ' and δ protons resonate at 2.02, 0.78 and 0.92 ppm, respectively, while the γ_1 and γ_2 CH_2 protons could only be pairwise assigned at 1.16 and 1.64 ppm. Similarly, for Leu-25 in fraction II one β proton was found at 1.78 ppm (the other β proton has not yet been assigned), the γ -proton resonates at 1.61 ppm, and the δ_1 and δ_2 methyl protons were pairwise assigned at 0.98 and 0.93 ppm, respectively. Again, since all cross-peaks involving methyl

protons could be accounted for, there is no evidence for more than two crambin species.

4. CONCLUSIONS

Crambin as isolated from the seeds of *Crambe abyssinica* separates on HPLC in two protein fractions. The amino acid analysis and the NMR spectra of these fractions show that these are single proteins, a Ser-22/Ile-25 and a Pro-22/Leu-25 species, occurring in an approximate ratio of 55:45. This finding should be useful in the further refinement of the crystal structure [3], which so far has been carried out assuming a statistical mixture of four crambin species with Pro/Ile as the most abundant one. Also, since the crambin structure is used in model calculations to test potential energy functions [5] and in restrained molecular dynamics simulations [13], we suggest that these calculations should be carried out with the Ser/Ile or Pro/Leu proteins and not with the Pro/Ile species, for which we have found no evidence.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO). L.J.B. was supported by a grant from the National Science Foundation (grant no. INT-8603427) and by a visiting professor fellowship from ZWO.

REFERENCES

- [1] Van Etten, C.H., Nielsen, H.C. and Peters, J.E. (1965) *Phytochemistry* 4, 467-473.
- [2] Teeter, M.M. and Hendrickson, W.A. (1979) *J. Mol. Biol.* 127, 219-223.
- [3] Hendrickson, W.A. and Teeter, M.M. (1981) *Nature* 290, 107-113.
- [4] Teeter, M.M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 6014-6018.
- [5] Whitlow, M. and Teeter, M.M. (1986) *J. Am. Chem. Soc.* 108, 7163-7172.
- [6] Lecomte, J.T.J., De Marco, A. and Llinás, M. (1982) *Biochim. Biophys. Acta* 703, 223-230.
- [7] Lecomte, J.T.J. and Llinás, M. (1984) *Biochemistry* 23, 4799-4807.
- [8] Lecomte, J.T.J. and Llinás, M. (1984) *J. Am. Chem. Soc.* 106, 2741-2748.
- [9] De Marco, A., Lecomte, J.T.J. and Llinás, M. (1981) *Eur. J. Biochem.* 119, 483-490.
- [10] Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- [11] Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.* 182, 281-294.
- [12] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) *J. Mol. Biol.* 182, 179-182.
- [13] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3801-3805.
- [14] Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.* 104, 6800-6801.
- [15] Marion, D. and Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun.* 113, 967-974.
- [16] Needleman, S.B. (1970) *Protein Sequence Determination*, pp. 141-150, Springer, Berlin.