Two-Dimensional ^{1}H Nuclear Magnetic Resonance Study of AaH IT, an Anti-Insect Toxin from the Scorpion *Androctonus australis* Hector. Sequential Resonance Assignments and Folding of the Polypeptide Chain†

H. Darbon,*,‡ C. Weber,§ and W. Braun§

CNRS URA 1179, Laboratoire de Biochimie, Faculté de Médecine Nord, Boulevard Pierre-Dramard, F 13326 Marseille Cédex 15, France, and Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule—Hönggerberg, CH-8093 Zürich, Switzerland

Received June 25, 1990; Revised Manuscript Received October 9, 1990

ABSTRACT: Sequence-specific nuclear magnetic resonance assignments for the polypeptide backbone and for most of the amino acid side-chain protons, as well as the general folding of AaH IT, are described. AaH IT is a neurotoxin purified from the venom of the scorpion *Androctonus australis* Hector and is specifically active on the insect nervous system. The secondary structure and the hydrogen-bonding patterns in the regular secondary structure elements are deduced from nuclear Overhauser effects and the sequence locations of the slowly exchanging amide protons. The backbone folding is determined by distance geometry calculations with the DISMAN program. The regular secondary structure includes two and a half turns of α-helix running from residues 21 to 30 and a three-stranded antiparallel β-sheet including peptides 3–5, 34–38, and 41–46. Two tight turns are present, one connecting the end of the α-helix to an external strand of the β-sheet, i.e., turn 31–34, and another connecting this same strand to the central one, i.e., turn 38–41. These structure elements are very similar to the secondary structure reported in single crystals for either variant 3 from the scorpion *Centruroides sculpturatus* Ewing (CsE V_3) or toxin II from the scorpion *A. australis* Hector (AaH II). The differences in the specificity of these related proteins, which are able to discriminate between mammalian and insect voltage-dependent sodium channels of excitable tissues, are most probably brought about by the position of the C-terminal peptide with regard to a hydrophobic surface common to all scorpion toxins examined thus far. This surface is made of an aromatic cluster that is surrounded by long hydrophobic side-chain residues, as well as the loops protruding out of it. Thus, the interaction of a given scorpion toxin with its receptor might well be governed by the presence of this solvent-exposed hydrophobic surface, whereas adjacent areas modulate the specificity of the interaction.

Scorpion neurotoxins form a family of homologous proteins (Rochat et al., 1970, 1979). They bind with high affinity to the voltage-dependent sodium channel, thereby affecting the gating properties of the channel (Catteral, 1977; Beneski & Catteral, 1980). Depending on their mode of action, cooperative binding with other toxins, and differences in binding site, these toxins have been divided into two groups, namely α and β (Jover et al., 1980; Couraud et al., 1982).

They display various degrees of toxicity toward different animal classes. In that respect they have been divided into toxins specifically acting on the mammalian nervous system (mammal toxins) and toxins specifically binding to the sodium channel of the insect nervous system (insect toxins) (Zlotkin et al., 1971a,b; Gordon et al., 1985; De Lima et al., 1988).

Within the insect toxin group, a further discrimination may be made between insect toxins inducing, in the tested insect larvae, a flaccid paralysis (Lester et al., 1982; Zlotkin et al., 1985) or a paralysis in a contractive form (Zlotkin et al., 1971a,b; De Dianous et al., 1987). *Androctonus australis* Hector insect toxin (AaH IT), the protein studied here, belongs to this last group. The two toxins of which the three-dimensional X-ray crystal structures are known, *A. australis* Hector toxin II (AaH II) and *Centruroides sculpturatus* Ewing variant 3 (CsE V_3), are representative of respectively α and β scorpion toxins active on mammals.

Knowledge of the covalent structure (Figure 1) (Rochat et al., 1979; Possani, 1984) of some of the toxins allows us to draw a general picture of these miniproteins. Most generally basic, they are made of a single chain, 60–70 amino acids long and reticulated by four disulfide bridges. Alignment of the eight half-cysteine residues shows the existence of strong sequential homologies and allows the classification of scorpion toxins into several structural groups in which the homologies are maximum.

Although AaH II and CsE V_3 belong to two different pharmacological groups, they share common three-dimensional features: two and a half turns of α-helix, a three-stranded antiparallel β-sheet, and a solvent-exposed hydrophobic surface. Their main differences correspond to a five-residue extra loop found only in α-toxins and to the orientation of the C-terminal stretch. These two regions, which are adjacent to the conserved hydrophobic surface, are thought to be involved in the modulation of the activity of the toxins (Fontecilla-Camps et al., 1980, 1981, 1988). Two scorpion toxins, toxin M9 from the scorpion *Buthus eupeus* (BE M9) (Pashkov et al., 1988) and insect toxin I_5A from the same scorpion (Arseniev et al., 1984), have been studied by 2D ^{1}H NMR. The first one is a mammal toxin that belongs to the α-type pharmacological classification. The solution spatial structure of BE M9 obtained by ^{1}H NMR spectroscopy may be compared to the X-ray structure of toxin II from the scorpion *A. australis*

† Part of this work was done during a postdoctoral stay of H.D. at the ETH Zürich, where he was recipient of a grant from the Fondation pour la Recherche Médicale, France.

‡ Laboratoire de Biochimie, Faculté de Médecine Nord.

§ Institut für Molekularbiologie und Biophysik.

0006-2960/91/0430-1836$02.50/0

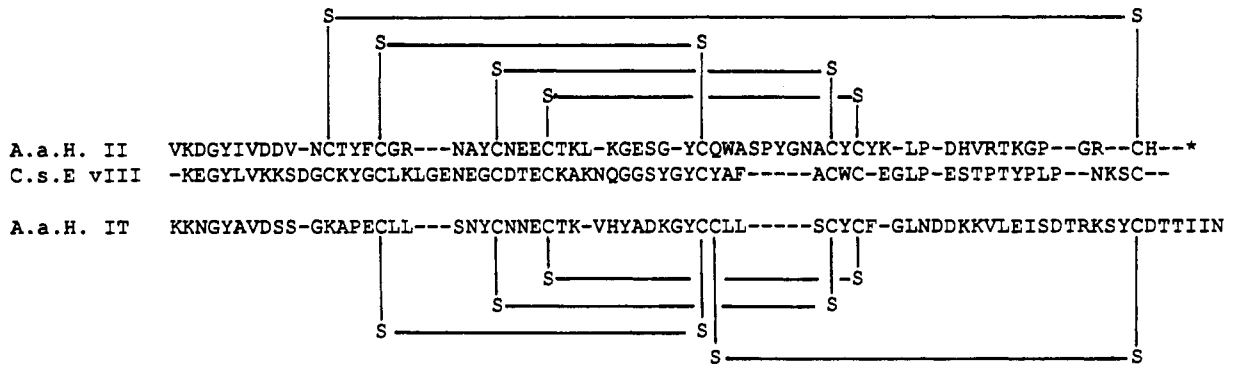

FIGURE 1: Amino acid sequence of AaH IT, AaH II, and CsE V3. The gaps are placed to optimize the sequence homologies. The star indicates that this terminal carboxyl group is amidated.

Hector (AaH II) (Fontecilla-Camps et al., 1988), which also belongs to the α pharmacological class. They possess very similar three-dimensional structures. They both contain a three-stranded antiparallel β structure and two and a half turns of α-helix. The other solution spatial structure of a scorpion toxin obtained by ^{1}H NMR spectroscopy is that of BE I_5A. This toxin belongs to a different scorpion toxin family and presumably acts on the glutamate receptor of the postsynaptic membrane (Grishin et al., 1982). This toxin is also much shorter, 35 amino acids long instead of 60–70 for "long" neurotoxins. However, despite these strong dissimilarities, BE I_5A also contains a three-stranded antiparallel β-sheet structure and two and a half turns of α-helical structure. Arseniev et al. (1984) have attempted to compare this structure with the X-ray structure of variant 3 from the scorpion *C. sculpturatus* Ewing (CsE V_3). CsE V_3 again contains a three-stranded antiparallel β structure and an α-helix (Fontecilla-Camps et al., 1980). The secondary stucture organization of BE I_5A seems to be very similar to that of CsE V_3, despite their very different mode of action.

Overall, up to now it appears that scorpion toxins share the same common secondary structure organization. The remaining question is to determine the molecular basis of the specificity demonstrated by these proteins. In that respect, determination of the three-dimensional structure of AaH IT is of fundamental importance because of its strong anti-insect specificity and because of a unique structural feature among scorpion toxins, i.e., an atypic disulfide bridge and an unusually long C-terminal stretch [see Figure 1 and Darbon et al. (1982)].

We herein report the sequence-specific ^{1}H NMR assignment for AaH IT as well as the secondary structure derived from NOE's between different backbone protons and from observation of slowed exchange of hydrogen-bonded amide protons with the solvent. The backbone folding of the protein is described as given by distance geometry calculations and is compared with those of related mammal toxins.

Experimental Procedures

A. australis Hector insect toxin (AaH IT) was purified from the venom of AaH scorpions collected in the area of Tozeur (Tunisia) according to the procedure previously described by Zlotkin et al. (1971b) and recently modified by Loret et al. (1990).

The 3.5 mM concentration of AaH IT was obtained without any sign of precipitation at pH 5.2; the pH was adjusted by adding drops of dilute HCl or NaOH. The solvent was either a mixture of 90% H_2O and 10% ^2H_2O or 99.99% ^2H_2O.

Protein samples in ^2H_2O with the labile protons completely exchanged against deuterium were prepared by heating the solution to 45 °C for 2 h, lyophilizing, twice redissolving in ^2H_2O and lyophilizing, and then solubilizing in 99.99% ^2H_2O to get the final solution.

All the 2D ^{1}H NMR spectra were recorded on a Bruker AM600 spectrometer, except for NH-exchange experiments where a Bruker WM500 spectrometer was used. 2QF-COSY (Piantini et al., 1982; Braun et al., 1981), 3QF-COSY (Müller et al., 1985), P-COSY (Marion & Bax, 1988), Clean TOCSY (Braunschweiler & Ernst, 1983; Bax & Davis, 1986; Griesinger et al., 1988), and NOESY (Kumar et al., 1980; Jeener et al., 1979) experiments were recorded by standard procedures. 2D spectra were Fourier transformed with standard Bruker software.

NOESY data sets acquired in H_2O solution were routinely strip-transformed such that only the left half of the spectrum was retained. The spectra were base-line corrected in both dimensions by fitting individual rows and columns to a third-order polynomial, which was then subtracted.

Amide proton exchange was studied by NOESY with the following procedure: The fully protonated sample was lyophilized from H_2O. The toxin was redissolved in 99.99% ^2H_2O and immediately inserted into the preshimmed spectrometer, equilibrated at 22 °C. Data acquisition was started 1 h later to obtain a NOESY spectrum in 6 h. Amide protons still giving rise to off-diagonal signals were then considered to be slowly exchangeable.

For the collection of NOE's used in distance geometry calculations, NOESY spectra with a mixing time of 80 ms (spectra recorded in H_2O) or 150 ms (spectra recorded in ^2H_2O) were recorded. Distance geometry calculations were performed on a VAX Station 2000 with the program DISMAN (Braun & Go, 1985) Stereospecific assignments for the interpretation of NOE distance constraints were not done in this first round of structure calculation. The structures were displayed for analysis on a Silicon Graphics System (Iris 3020) by using the program FRODO (Jones, 1978).

Results

As judged by circular dichroism experiments (data not shown), no major structural change of the toxin occurs at 22 °C between pH 3.6 and pH 8.6. In the same manner, structural integrity was ensured at pH 5 below 40 °C.

Identification of the Amino Acid Spin Systems Comprised of Nonlabile Protons. The general strategy described by Wüthrich (1986) has been closely followed for the identification of signals belonging to the various amino acid spin systems.

Figure 2A shows the αH–βH cross-peak region of a 2QF-COSY recorded in H_2O. In that spectrum, four AX spin systems pertaining to four glycines stand out clearly by virtue

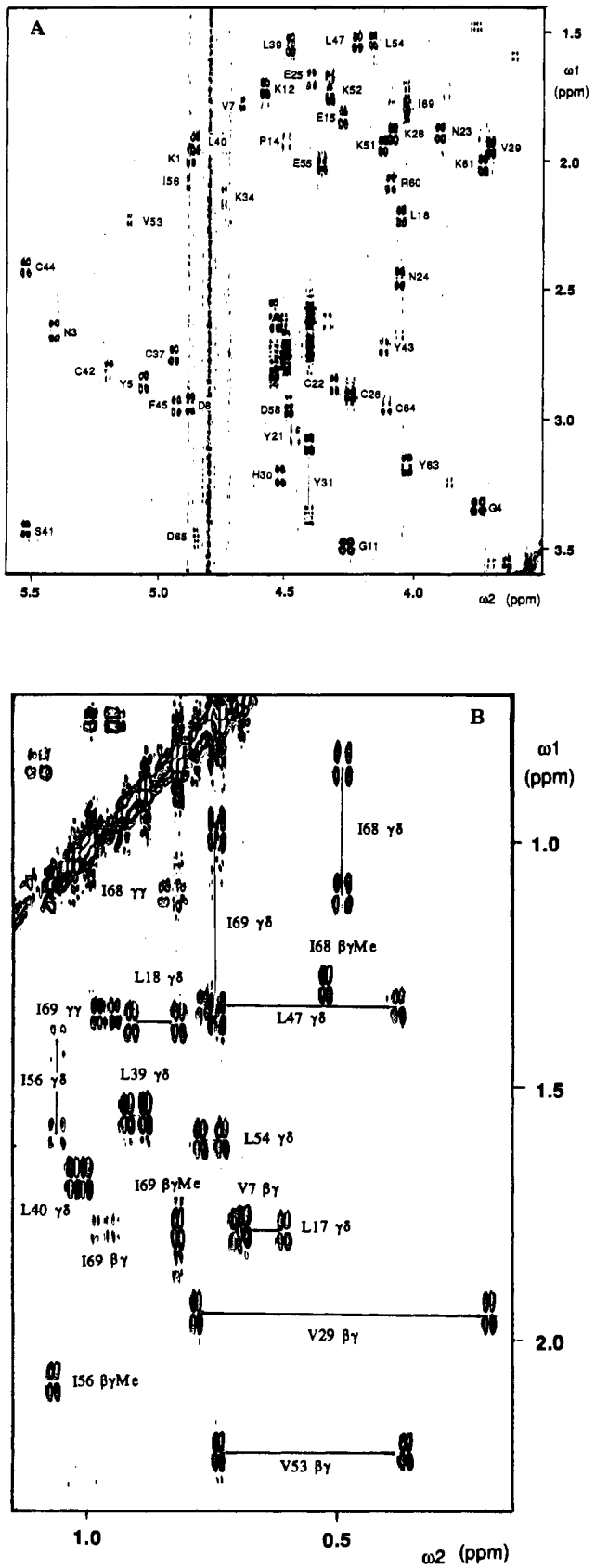

FIGURE 2: (A) 2QF-COSY spectrum of a 3.5 mM AaH IT solution in ^2H_2O at 22 °C, p^{2}H 5.2, after complete exchange of the labile protons. The region containing most αH–βH cross-peaks is shown. Positive and negative levels are reproduced without distinction. The cross-peaks are identified with the amino acid type and the sequence number obtained as the final result of the assignment procedure. In most of the cases only one β-proton is seen. A total of 920 experiments of 32 scans each were recorded. $t_1$max and $t_2$max were set to 71 and 320 ms, respectively. (B) Methyl region of the same spectrum.

of the large active coupling constant. The lack of these cross-peaks in 3QF-COSY confirms this assignment.

Identification of the three valines started from COSY cross-peaks connecting the methyl doublets to the β-proton (Figure 2B). For Val29 and Val7, the α-proton was located by 2QF-COSY; for Val53, it was located by TOCSY on the basis of the two CαH–CγH$_3$ cross-peaks. The three isoleucines, i.e., Ile56, Ile68, and Ile69, were completely identified by 2QF-COSY by starting from CγH–CδH$_3$ cross-peaks (Figure 2B).

The CαH–CβH$_3$ cross-peaks of the three alanines were identified by their chemical shifts and distinguished from CβH–CγH$_3$ of the four threonines by the lack of CαH–CγH$_3$ relayed cross-peaks in the TOCSY spectrum. These relayed cross-peaks also identified the α-proton of the four threonines. Five of the six seryl residues were recognized as AMX spin systems, all with chemical shifts greater than 3.9 ppm (Table I). Ser41 was identified as an unknown AMX spin system at this stage due to the high-field-shifted β-proton chemical shift. The αH–βH COSY cross-peaks of nine AMX spin systems are highly overlapping (Figure 2A) and were impossible to unambiguously assign at that stage. They were later identified on the basis of TOCSY cross-peaks with amide protons and were found to belong to Cys16, Asn23, Asp33, Cys38, Asn48, Asp49, Asp50, Asp58, and Asn70. All other AMX spin systems were easily analyzed on the basis of ^2H_2O-COSY spectra.

The long-side-chain spin systems include Glu, Leu, Pro, Arg, and Lys. Identification of the six leucines started from the remaining cross-peaks in the methyl region (Figure 2B) and was completed by combination of 2QF-COSY and TOCSY. The remaining COSY cross-peaks in Figure 2B ($\omega_1 < 2.2$ ppm; see Table I) were at that stage identified as αH–βH cross-peaks belonging to long-side-chain spin systems.

Identification of the aromatic spin system was obtained by 2QF-COSY and NOESY spectra in ^2H_2O in which the usually strong NOE between the β-proton and the 2,6-position protons was taken into account.

Completion of the Spin Systems with the Labile Protons. The P-COSY fingerprint (Figure 3) contains 65 assigned cross-peaks attributed to one of the spin systems identified, in ^2H_2O solution, either by COSY on the basis of unique Hα chemical shifts or by TOCSY through relayed cross-peaks. Some cross-peaks (labeled with a star in Figure 3) could not be attributed to any of the side-chain spin systems identified in ^2H_2O solution. These cross-peaks were absent in an early H_2O 2QF-COSY spectrum recorded at pH 3.5; they appear with aging of the sample and show a small coupling constant. Taken together, these observations led us to assume that they come from impurities or degradation products.

The amide proton resonances not seen in the fingerprint were further identified on H_2O NOESY spectra by using the sequential resonance assignment procedure (see below), through $d_{\alpha N}$ connectivities for the residues Gly4, Glu15, and Cys26, $d_{\beta N}$ connectivities for Cys38, and d_{NN} connectivities for Asn20.

Sequence Resonance Assignments. Following the usual procedure (Wüthrich, 1986, and references therein; Wüthrich, 1989; Wagner, 1990), sequential relations were established by using the connectivities $d_{\alpha N}$, $d_{\beta N}$, and d_{NN}. Figure 4 shows that sequential connectivities were observed throughout, except at Asn48 and Cys64. The previously unassigned amide proton resonances were assigned within the sequential assignment.

Alanines are good starting points, due to their strictly unambiguous assignment. Starting from Ala13, the assignment proceeded in the reverse direction via $d_{\alpha N}$ connectivities until

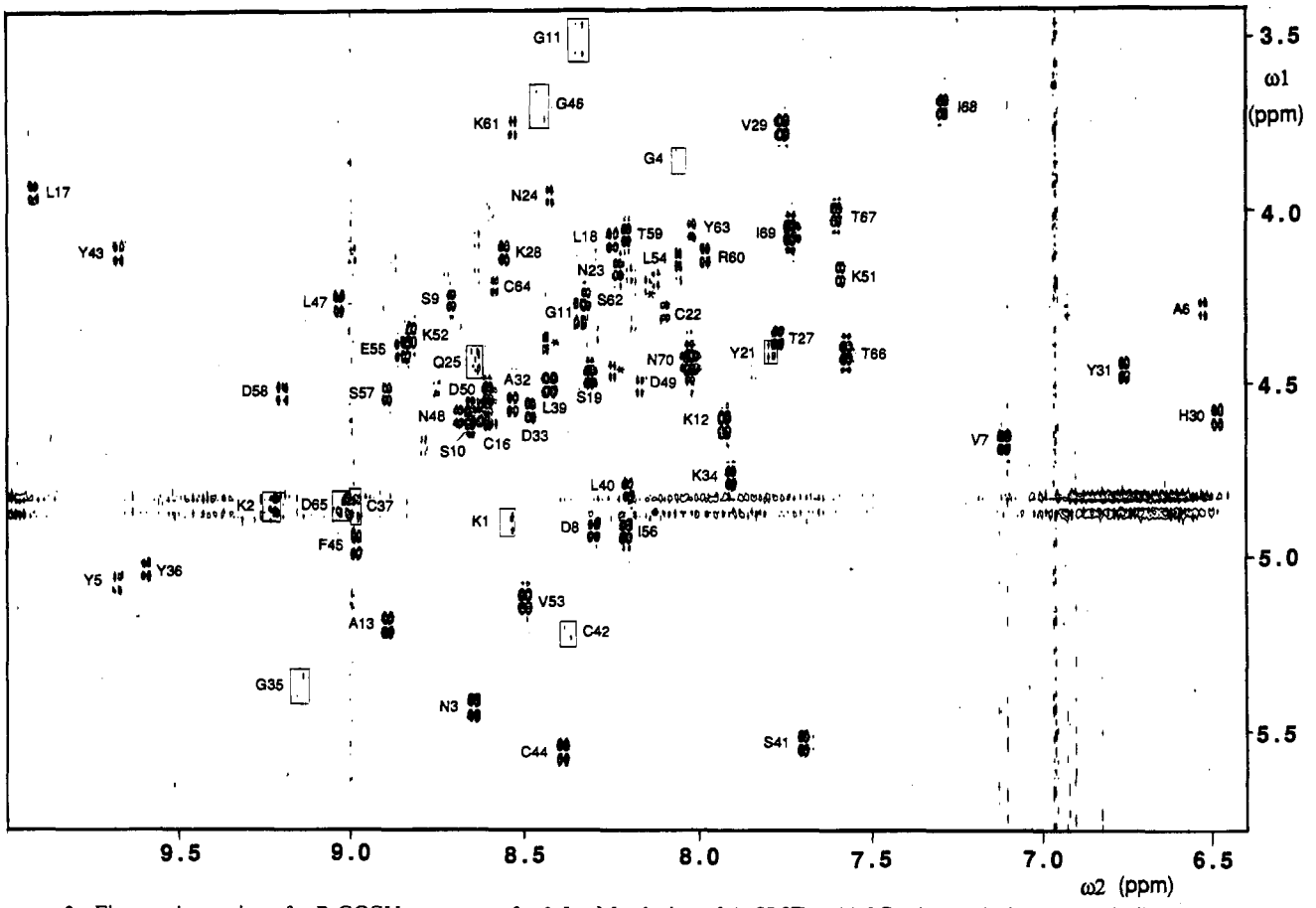

FIGURE 3: Fingerprint region of a P-COSY spectrum of a 3.5 mM solution of AaH IT at 22 °C, pH 5.2, in H_2O. The individual cross-peaks are labeled with the amino acid type and the sequence number obtained as the result of the sequential assignments. The stars label unassigned cross-peaks.

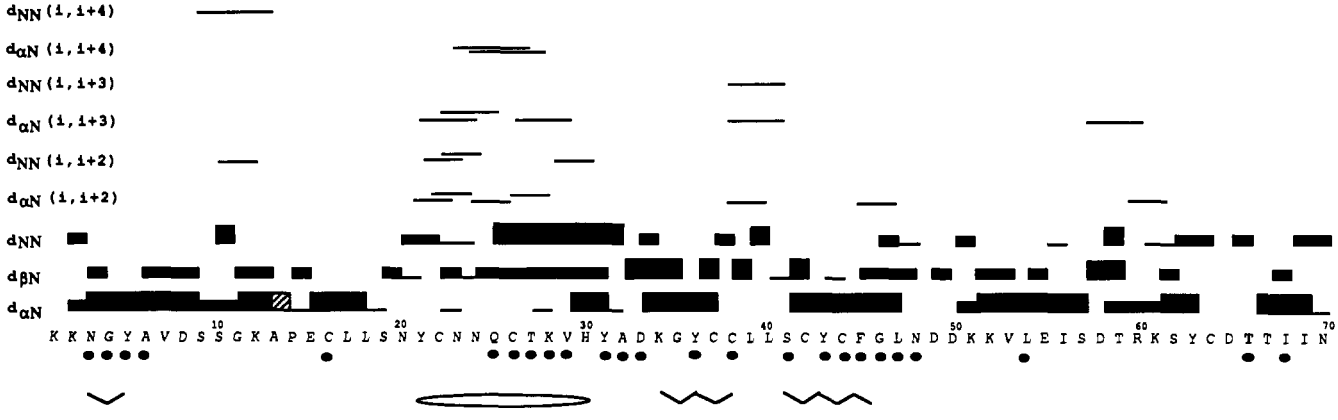

FIGURE 4: Survey of the sequential connectivities that were used to obtain the individual resonance assignments listed in Table I. Sequential $d_{\alpha N}$ (or $d_{\alpha\delta}$ for the Ala13–Pro14 dipeptide segment, indicated with a dashed box), d_{NN}, and $d_{\beta N}$ connectivities manifested by strong NOESY cross-peaks, are indicated by large boxes, those observed as medium NOESY peaks are shown by small boxes, and those observed as weak NOESY peaks are shown by lines. Filled circles identify residues for which the amide proton is slowly exchangeable. Medium-range NOE's are shown by lines, and secondary structures are schematized at the bottom.

Asn3 was reached. A strong $d_{\alpha N}$ connectivity between NH of Asn3 and an unassigned α-proton allowed us to localize the amide line of Lys2, the α-proton of which resonates at the chemical shift of water. Starting from Ala13 in the normal direction, the NOESY spectrum recorded in H_2O revealed strong cross-peaks with the possible δ-protons from the unique proline. Then, using ^2H_2O-NOESY, we assigned a possible spin system to this proline from which a strong $d_{\beta N}$ connectivity allowed us to assign the amide proton line of Glu15, and from there we proceeded, via $d_{\alpha N}$, to Leu18. The $d_{\alpha N}$ between Leu18 and Ser19 could not be observed due to the degeneracy of the two amide protons; however, a strong $d_{\beta N}$ allowed us to unambiguously assign the amide proton of Ser19. In that partial sequence, both glycines gave rise to two $d_{\alpha N}$ connectivities with the residue that followed.

Starting from the unique phenylalanine, Phe45, it was possible, by working in both directions, to sequentially assign the residues from Ser43 to Leu47.

The C-terminal peptide was easily assigned due to the two contiguous threonines (Thr66 and Thr67) and the two contiguous isoleucines (Ile68 and Ile69).

The remaining isoleucine, Ile56, was used as a starting point to assign the sequence 50–57. The connectivity between Ser57 and Asp58 was not observable due to the degeneracy of the

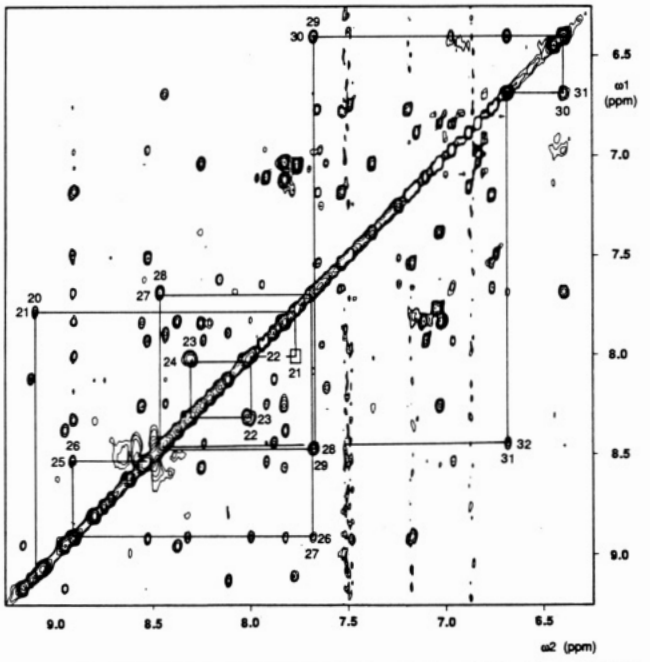

FIGURE 5: Illustration of sequential d_{NN} connectivities in AaH IT. A phase-sensitive NOESY spectrum of a 3.5 mM solution in 90% H_2O/10% ^2H_2O at pH 5.2 and 22 °C is shown ($t_1$max = 80 ms; $t_2$max = 320 ms; SW = 6400 Hz). The spectrum is the result of the addition of four time domain data sets acquired in immediate succession (13 h each). This approach significantly reduces t_1 noise. Solid lines connect the NH–NH COSY cross-peak positions with the sequential NOESY cross-peaks for the Tyr21–Ala32 peptide segment.

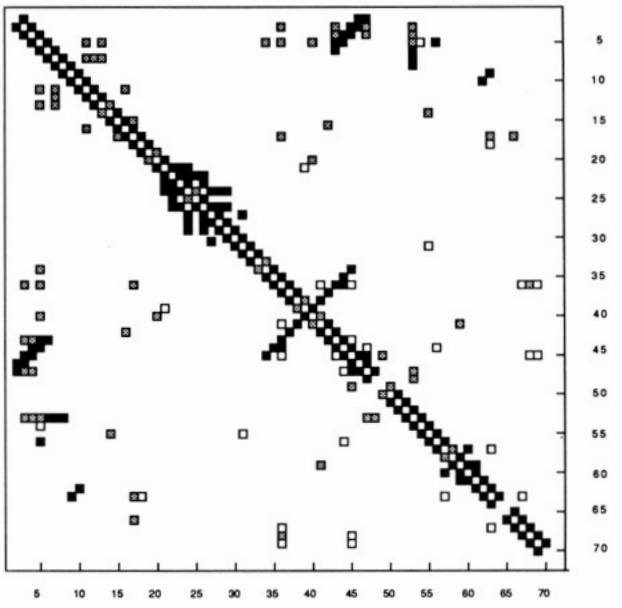

FIGURE 6: Diagonal plot for AaH IT of the NOE's observed in NOESY spectra. A filled square at the position x, y indicates that one or more NOE's were observed between backbone protons of the two residues at sequences positions x and y. A shaded square indicates an NOE between a backbone proton of one residue and a side-chain proton of the other residue. An open square indicates an NOE between side-chain protons of the two residues. Intraresidue NOE's are not shown. When two residues are connected by more than one NOE, only the one involving the largest number of backbone protons is shown.

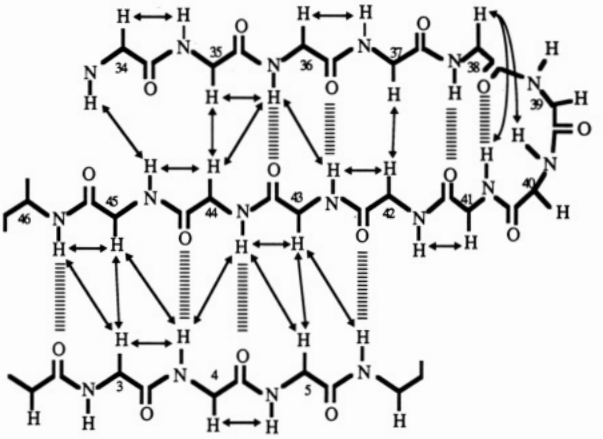

FIGURE 7: Schematic representation of the β-sheet secondary structure in AaH IT. Hydrogen bonds postulated on the basis of the NOE's and the amide proton exchange data are drawn as broken lines. Arrows indicate sequential medium, and long-range NOE's.

α-protons, but it was possible to progress through $d_{\beta N}$ and then through $d_{\alpha N}$ from Asp53 up to Tyr63.

Using the remaining glycine residue (Gly35) as the starting point, it was possible, by working in both directions, to assign sequence 33–37.

We were allowed to connect the remaining Val29 to His30 and continue up to Ala32 by $d_{\alpha N}$ connectivities. $d_{\beta N}$ connectivities were used to connect Ala32/Asp33, Cys38/Leu39, Leu40/Ser41, Leu47/Asn48, Asp49/Asp50, and Ser57/Asp58 and also to check the validity of the $d_{\alpha N}$ connectivities.

Finally, it was possible to step from Asn20 up to Ala32 through d_{NN} connectivities (Figure 5), with the exception of Asn24/Glu25, a gap which was closed by the $d_{\beta N}$ connectivity. A number of medium-range connectivities were also determined in this sequence [$d_{NN}(i, i + 2)$; $d_{\alpha N}(i, i + 2)$; $d_{\alpha N}(i, i + 3)$; $d_{\alpha N}(i, i + 4)$]. Gaps between Cys37/Cys38, Leu39/Leu40, Tyr63/Cys64, and Asp65/Thr66 were closed by d_{NN} connectivities.

NOE Constraints Defining Regular Secondary Structure Elements. A continuous set of d_{NN} NOE's (Figures 4 and 5) was observed from Asn20 to Ala32, which is indicative of α-helical secondary structure (Wüthrich, 1986). This is confirmed by a number of medium-range d_{NN} and $d_{\alpha N}$ NOE's (Figure 4). Taking into account the amide proton exchange data (see below), we assign an α-helical structure to residues Tyr21 to His30.

Strong sequential $d_{\alpha N}$ NOE's all along the polypeptide chain, with the exception of the α-helical structure, indicate that the scorpion toxin contains predominantly extended structures. A number of long-range backbone NOE's between peptides 3–5, 34–38, and 41–46 (Figures 6 and 7) indicate a three-stranded antiparallel β-sheet structure. The central strand runs from Ser41 to Gly46.

NH Exchange Data. Analysis of the amide proton exchange gives additional information for the pattern recognition procedure because the observation of slowed exchange is evidence for hydrogen-bonded amide protons (Wagner & Wüthrich, 1982). In that respect, the pattern of five successive, slowly exchanged amide protons from residues 25 to 29 confirms our assignment of the above structure as α-helical, because the amide protons of the first four residues in an α-helix are usually solvent exposed.

The location of the β-sheet structure is also confirmed by the proton exchange data; all the amide protons of the central stretch (peptide 41–46) show low exchange rates, whereas an alternating pattern of slowly and rapidly exchanged amide protons is typical of an external strand. In that respect, our exchange data confirm the external position of peptide 34–38 but are not strictly consistent with the external position of peptide 3–5 due to the low exchange rate of the amide proton of Asn3 and Tyr5. DISMAN calculations (see below) place these protons close in space to the slowly exchanged amide proton of Leu54, thus possibly forming a short parallel β structure

Table I: ^{1}H Chemical Shifts for AaH IT in H_2O Solution at 22 °C, pH 5.2

amino acid residue	NH	αH	βH	others
Lys1	8.45	4.89	1.98, 2.09	
Lys2	9.50	4.82	1.76, 1.88	
Asn3	8.54	5.42	2.65, 2.55	δNH_2 7.83, 7.12
Gly4	7.98	3.76, 3.33		
Tyr5	9.58	5.06	2.85, 2.65	δH 7.38; εH 7.04
Ala6	6.47	4.27	0.62	
Val7	7.03	4.64	1.80	γCH_3 0.70, 0.70
Asp8	8.21	4.86	3.45, 2.94	
Ser9	8.60	4.25	4.05, 4.00	
Ser10	8.58	4.57	4.06, 3.96	
Gly11	8.25	4.26, 3.49		
Lys12	7.83	4.59	1.81, 1.71	γCH_2 1.32; δCH_2 1.66; ϵCH_2 3.00
Ala13	8.78	5.18	1.25	
Pro14		4.51	2.65, 1.92	γCH_2 2.29, 2.14; δCH_2 4.15, 3.90
Glu15	9.21	4.30	1.94, 1.86	
Cys16	8.59	4.57	2.77, 2.85	
Leu17	9.83	3.92	1.87, 1.45	γH 1.78; δCH_3 0.71, 0.61
Leu18	8.14	4.07	2.22, 1.92	γH 1.38; δCH_3 0.94, 0.85
Ser19	8.18	4.36	4.29, 4.06	
Asn20	9.10	4.19	2.97, 2.87	δNH_2 7.76, 7.11
Tyr21	7.74	4.44	3.45, 3.05	δH 7.15; εH 6.88
Cys22	7.99	4.27	2.90, 2.86	
Asn23	8.36	3.92	1.95, 1.75	
Asn24	8.11	4.08	2.66, 2.45	
Glu25	8.53	4.42	1.84, 1.68	
Cys26	8.91	4.26	4.11, 2.83	
Thr27	7.69	4.35	4.11	γCH_3 1.19
Lys28	8.46	4.11	1.90, 1.75	
Val29	7.66	3.74	1.96	γCH_3 0.81, 0.23
His30	6.40	4.62	3.22, 3.16	4H 6.81; 2H 7.45; 1H 7.82
Tyr31	6.68	4.44	3.37, 3.10	δH 7.03; εH 6.83
Ala32	8.45	4.55	1.28	
Asp33	8.39	4.55	2.64, 2.58	
Lys34	7.83	4.75	2.13, 1.93	γCH_2 1.32, 1.19; δCH_2 1.47; ϵCH_2 2.56, 2.34
Gly35	9.05	5.34, 4.27		
Tyr36	9.50	5.02	3.28, 2.68	δH 7.19; εH 6.76
Cys37	8.90	4.95	2.92, 2.75	
Cys38	8.09	4.56	3.24, 2.75	
Leu39	8.31	4.48	1.50, 1.55	γH 1.55; δCH_3 0.92, 0.89
Leu40	8.02	4.85	1.95, 1.84	γH 1.69; δCH_3 1.03, 1.01
Ser41	7.61	5.51	3.47, 3.42	
Cys42	8.29	5.20	2.89, 2.79	
Tyr43	9.59	4.11	2.62, 2.72	δH 5.55; εH 6.00
Cys44	8.31	5.54	3.00, 2.41	
Phe45	8.90	4.95	3.08, 2.95	δH 7.18; εH 7.65; ζH 7.53
Gly46	8.36	4.25, 3.67		
Leu47	8.95	4.26	1.65, 1.54	γH 1.34; δCH_3 0.76, 0.39
Asn48	8.60	4.58	2.86, 2.80	δNH_2 7.92, 7.10
Asp49	8.00	4.42	2.65, 2.5	
Asp50	8.52	4.51	2.72, 2.63	
Lys51	7.50	4.17	2.57, 1.94	γCH_2 1.45; δCH_2 1.62, 1.51; ϵCH_2 2.56
Lys52	8.72	4.34	1.73, 1.68	γCH_2 1.36
Val53	8.39	5.10	2.24	γCH_3 0.76, 0.36
Leu54	8.03	4.18	1.61, 1.54	γH 1.61; δCH_3 0.76, 0.70
Glu55	8.74	4.37	2.02, 1.61	γCH_2 2.35, 2.25
Ile56	8.11	4.90	2.08	γCH_2 1.59, 1.41; γCH_3 1.08; δCH_3 1.06
Ser57	8.79	4.41	4.51, 4.08	
Asp58	9.15	4.51	2.98, 2.92	
Thr59	8.11	4.06	4.12	γCH_3 1.37
Arg60	7.68	4.10	2.08, 1.95	γCH_2 1.82, 1.70; δCH_2 3.42, 3.34
Lys61	8.44	3.75	2.01, 1.88	γCH_2 1.53; δCh_2 1.85, 1.75; ϵCH_2 3.09
Ser62	8.24	4.24	4.03, 3.98	
Tyr63	7.92	4.05	3.15, 3.10	δH 6.96; εH 6.85
Cys64	8.53	4.08	2.86, 2.95	
Asp65	8.91	4.86	3.48, 3.43	
Thr66	7.49	4.40	4.26	γCH_3 1.25
Thr67	7.51	4.00	3.88	γCH_3 1.00
Ile68	7.21	3.68	1.29	γCH_2 1.11, 084; γCH_3 0.54; δCH_3 0.50
Ile69	7.62	4.06	1.78	γCH_2 1.34, 098; γCH_3 0.83; δCH_3 0.74
Asn70	7.93	4.43	2.73, 2.60	δNH_2 7.83, 7.04

including Val53 and Leu54. This last structure is probably irregular because the NOE data (Figure 6) do not show a clear pattern for a β-sheet.

NOE Measurements and Calibration of H–H Distance Constraints. NOE constraints are the primary source of experimental information in the determination of a protein structure from NMR measurements (Wüthrich, 1986, 1989). Most of these constraints were obtained from 80-ms mixing time spectra in order to decrease spin diffusion contributions as much as possible.

Table II: Analysis of Residual Constraint Violations in the Four AaH IT Structures Calculated with DISMAN

structure	av violation of NMR constraints (nm)[a] upper (382)	lower (28)	no. of violations for distances 0.05–0.1 nm SS	VW	NOE	>0.1 nm SS	VW	NOE
I	0.056	0.025	0	0	4	0	0	0
II	0.083	0.062	0	0	10	0	0	0
III	0.087	0.031	0	0	14	0	0	2
IV	0.112	0.042	0	0	13	0	0	0

[a] The average violation is given by the sum of distance violations divided by the total number (given in parentheses) of these distance constraints.

To relate the NOE data to intramolecular interproton distances, an estimate of the relative peak intensities was obtained by counting the number of contour levels at the peak maximum. The calibration curve was based on the $d_{\alpha N}$ cross-peaks, which correspond to a known distance [$d_{\alpha N}$(β-sheet) = 0.22 nm] (Wüthrich, 1986), and on the uniform averaging model (Braun et al., 1981). For intraresidual and sequential constraints, the NOE's were divided into three classes with upper limit distances of 0.24, 0.30, and 0.40 nm. When an NOE corresponding to a medium-range or long-range backbone constraint was observed, the upper limit was set to 0.40 nm irrespective to the peak intensity, as well as 0.50 nm for interresidue NOE's with side-chain protons. Pseudoatom corrections were added where required because of the lack of stereospecific assignments (Wüthrich et al., 1981). In Figure 6, the NOE data are presented in the form of a diagonal plot, in which the previously mentioned antiparallel β-sheet and α-helix secondary structures are clearly revealed.

Determination of Other Constraints. Additional constraints were obtained from the known location of the disulfide bridges (Darbon et al., 1982) (Figure 1). For each of the four disulfide bridges, i.e., Cys16/Cys37, Cys22/Cys42, Cys26/Cys44, and Cys38/Cys64, the following set of upper and lower limits was used: 0.20 nm < $d_{\gamma Si,\gamma Sj}$ < 0.21 nm and 0.30 nm < $d_{\beta Ci,\gamma Sj}$ < 0.31 nm. In later calculations (see below), hydrogen-bond constraints were added, where the presence of a hydrogen bond was revealed both by the amide proton exchange data and by the geometry of the backbone determined in the early stage of the calculation. These hydrogen bonds are those involved in the β-sheet and α-helical structures as determined above.

Completion of the Input List of Experimental Constraints and Structural Interpretation of the NMR Data with the Program DISMAN. Several rounds of DISMAN calculations and reexamination of the experimental NMR data were used to complete the assignments of the NOESY cross-peaks observed in our spectra. One hundred calculations with randomly generated starting conformations were initiated. The input contained about 280 NOE's, all disulfide constraints, and all van der Waals lower bound constraints with an equal weighting. The variable target function in DISMAN was increased from an initial level of two sequential residues, in steps of between five and ten residues, until all constraints were included. At each step of the calculation, the violations of the constraints were analyzed and in most cases the errors were determined to come from misassignments of NOESY cross-peaks and were corrected. Then, as judged by the residual error function, the ten best structures were selected for further calculations. Analysis of these preliminary structures allowed us to improve the input list of constraints; these structures were used as a guide for assignments of additional NOESY cross-peaks. These structures also allowed us to check for hydrogen bonding. The resulting input file thus contained 32 hydrogen-bond constraints (16 upper bonds and 16 lower bounds), 24 disulfide-bond constraints (12 upper bounds and 12 lower bounds), and 356 NOE constraints, including 120 intraresidual constraints, 106 sequential constraints, 36 medium-range backbone and long-range backbone constraints, and 94 interresidual constraints with side-chain protons. With this final constraint list and the ten best preliminary structures previously obtained (see above), a determination of the folding of AaH IT was made with the following procedure: for the five best among the ten preliminary structures mentioned above, the DISMAN calculations were continued and the number of calculation cycles was increased up to 1000 until the error function showed no further improvement and structures I–IV (Table II) were obtained. The fifth calculated structure was discarded because of residual violations that were too high Structure I is clearly a good solution, almost completely compatible with the set of constraints used, since there are only a few residual distance constraint violations larger than 0.05 nm, no violation larger than 0.1 nm, and small average residual violations.

Besides the analysis of the residual constraint violations, a check for the quality of the structures calculated by distance geometry can be obtained from the pairwise minimal RMSD values of these structures (Table III). For the complete structures, we obtained rather large RMSD values. There is a correlation between the size of the RMSD values and the number of residual constraint violations. Much better fits can be obtained in pairwise comparisons of the secondary structure elements (Table III). The large RMSD values obtained for the complete structures are thus due to the poorly determined conformation of the loops connecting these secondary structure elements, as revealed by the RMSD values obtained when the obtained structures of the large loop 47–65 are compared.

Figure 8 shows the α-carbon backbone fold in this structure, revealing the secondary structure determined by pattern recognition, i.e., a three-stranded antiparallel β-sheet made of residues 3–5, 41–46, and 34–38; and two and a half turns of α-helix, from residues 21 to 30, running roughly parallel to the β-strand. The relative position of the α-helix and the β-sheet is known from several long-range NOE's. The C-terminal peptide from residues 66 to 70 is in an extended fold and is also parallel to the β-sheet. The α-helix is connected by a tight turn to the upper strand of the β-sheet; the latter is connected by another tight turn to the central strand. The 46–65 region seems highly contorted but corresponds to a region for which very few NOE's were recorded, so the folding of this peptide is still questionable.

Discussion

The present work deals with the elucidation of the folding of an anti-insect scorpion toxin, with the aim of determining the molecular basis of its high specificity toward the insect nervous system. Consequently, the discussion will be centered

Table III: Root-Mean-Square Distance (RMSD) (in Å) between All Pairs of Structures Formed by Combining the Four DISMAN Structures[a]

residues	all	2–5	2–46	21–30	34–46	47–65	65–69	2–5, 21–30, 34–46
av	4.23	0.95	3.74	1.05	1.76	3.96	1.44	2.93
SD	1.95	0.62	1.73	0.54	0.83	1.57	0.80	1.36

[a] Only the backbone atoms N, Cα, and C′ were used in these calculations.

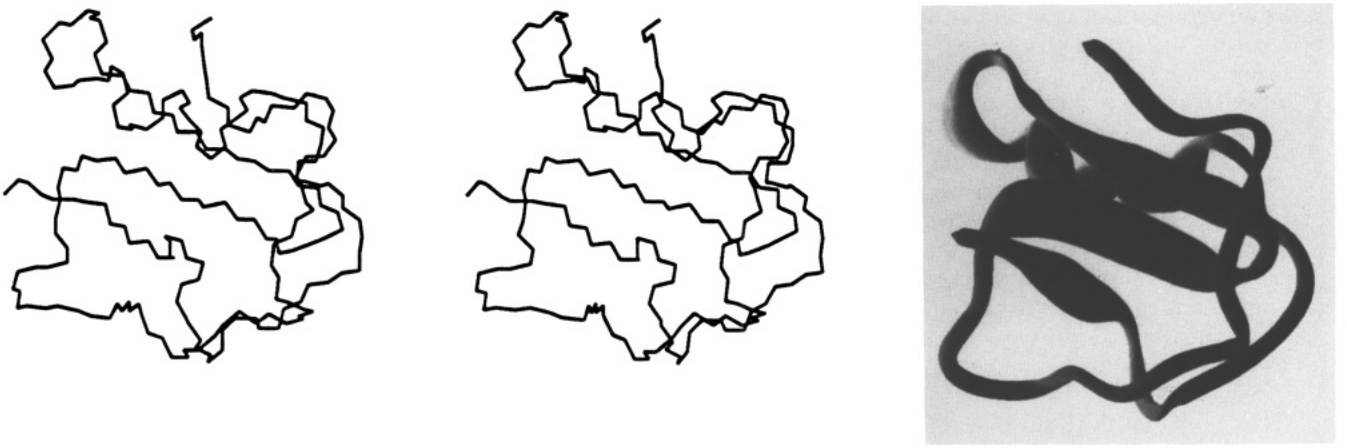

FIGURE 8: (Left) Stereoview of the solution structure I (Table II) of AaH IT calculated by DISMAN. Only the backbone atoms (C, Cα, N) are used. (Right) Schematic drawing of the polypeptide chain folding in the same structure, obtained with the program TURBO FRODO (Roussel & Cambillau, 1989).

FIGURE 9: Comparison of the Cα backbone folding of AaH IT (heavy lines) and AaH II (thin lines).

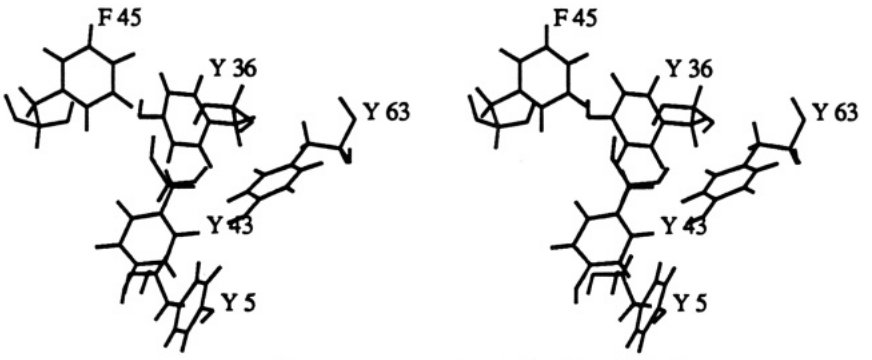

FIGURE 10: Stereoview of the aromatic cluster including Tyr5, Tyr36, Tyr43, Phe45, and Tyr63.

on the comparison of this structure with the crystal structure of two mammal-directed scorpion toxins, i.e., AaH II and CsE V_3, representatives of the α- and β-toxin group, respectively.

Figure 9 depicts the comparison of the α-carbon backbones of AaH IT and AaH II. The two molecules have a similar, highly structured region that contains the α-helix and the three-stranded antiparallel β-sheet moieties. The main differences occur in the loops connecting these secondary structures and in the C-terminal stretch of the molecules. The limited extent of our knowledge of the 46–63 loop (Table III) is due to the scarcity of medium- and long-range NOE's between this loop and the core of the molecule. The lack of NOE's in that region may be accounted for by a higher mobility of this peptide, which has also been demonstrated in the case of mammal scorpion toxins (Fontecilla-Camps et al., 1981, 1988, and personal communication). This mobility is even increased in the insect toxin due to the absence of the disulfide bridge linking the C-terminus to the core of the mammal toxins (bridge 12–63). However, from Figure 9, it is clear that the general shape of the insect toxin is similar to that of mammal toxins, except at the C-terminal end. Residue 59 of AaH IT can be considered to be topologically equivalent to the C-terminus of the mammal toxins.

The 38–64 disulfide bridge constraints together with several long-range NOE's occurring between the 66–70 C-terminal peptide and the core of the molecule (Figure 6) contribute to the positioning of this peptide. This peptide forms an extended strand, antiparallel to the upper strand of the β-sheet, but

without hydrogen-bond connections. Two slowly exchanged NH protons belonging to residues 66 and 68 have been localized in this segment. As judged by analysis of structure I (Table II), obtained by distance geometry, the amide proton of residue 66 is involved in a hydrogen bond with the carbonyl group of residue 63, thus contributing to the stabilization of the C-terminal peptide; on the other hand, there is no obvious partner to the amide proton of residue 68 in structure I. The position of this C-terminal peptide with regard to the hydrophobic surface constitutes a very peculiar feature of this anti-insect scorpion toxin compared to mammal scorpion toxins, from either α- or β-type, and may be responsible for its specific mode of action.

It has been proposed that the loop connecting the second and third strand of the β-sheet might be involved in the interaction of mammal toxins with their receptor, its size possibly modulating their specificity (Fontecilla-Camps et al., 1981, 1988). This loop is actually a tight turn in the insect toxin, but it is not accessible to the solvent because it is covered by the 60–63 peptide. Residues 60–63 of AaH IT are topologically homologous to residues 40–37 of AaH II, respectively. It is thus tempting to consider the insect toxin 60–63 peptide as part of the site of interaction.

As a result of these correspondences, the 64–70 peptide of AaH IT may be considered as an addition when compared to mammal toxins. As a matter of fact, this C-terminal peptide interferes with the region that has been previously proposed to play a role in mediating the toxin-receptor interaction, with respect to the mammal directed toxins. This region has been referred to as the "conserved hydrophobic surface" (Fontecilla-Camps et al., 1980, 1988). It is rich in exposed hydrophobic residues, mainly aromatic ones, i.e., tyrosine residues 5, 42, 47, and 49 in AaH II and tyrosine residues 4, 40, 42, and 58 and tryptophan residue 47 in CsE V_3. Consequently, special care was taken in analyzing the spatial organization of the tyrosine side chains. The occurrence of numerous NOE's between side-chain protons of the aromatic residues allows a well-defined description of the relative positions of these residues. Figure 10 shows this spatial organization as determined by the DISMAN calculations and reveals a similar arrangement of Tyr5, Tyr36, Tyr43, Tyr63, and Phe45, as in AaH II and CsE V_3. The central position of Tyr43 is further highlighted by the high-field shift of its side-chain protons (see Table I), due to strong ring current effect. This aromatic cluster is solvent-exposed, as it is in mammal toxins. Furthermore, Ile68 and Ile69 side chains are close to the aromatic cluster, as demonstrated by the presence of NOE's between them. Thus, the hydrophobic characteristic of this exposed region, observed in mammal scorpion toxins, is also depicted in the insect toxin. DISMAN calculations also locate Ile56 and Leu47 side chains within this hydrophobic cluster, thereby increasing the general hydrophobicity of this surface.

Thus, we propose that the insect specificity of AaH IT is at least partially determined by the C-terminal heptapeptide protruding from the edge of the hydrophobic surface. This hypothesis extends the model of the interaction of mammal scorpion toxins with the sodium channel, in which the hydrophobic surface is the constant part and the adjacent regions are the variable parts (Fontecilla-Camps et al., 1988, and references therein), to the insect-directed scorpion toxins.

ACKNOWLEDGMENTS

We thank Dr. Fontecilla-Camps for making his unpublished results available to us and for numerous discussions. His group is also acknowledged for supplying their computing facilities to H.D. We are highly indebted to Prof. Wüthrich and Prof. Rochat for fruitful discussions.

REFERENCES

Arseniev, A. S., Kondakaov, V. I., Maiorov, V. N., & Bystrov, V. F. (1984) *FEBS Lett. 165*, 57–62.

Bax, A., & Davis, D. G. (1986) *J. Magn. Reson. 65*, 355–358.

Beneski, D. A., & Catterall, W. A. (1980) *Proc. Natl. Acad. Sci. U.S.A 77*, 639–643.

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611–626.

Braun, W., Bösch, C., Brown, L. R., Go, N., & Wüthrich, K. (1981) *Biochim. Biophys. Acta 667*, 377–396.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–525.

Carson, M., & Bugg, C. E. (1986) *J. Mol. Graphics 4*, 121–123.

Catteral, W. A. (1977) *J. Biol. Chem. 252*, 8660–8668.

Couraud, F., Jover, E., Dubois, J.-M., & Rochat, H. (1982) *Toxicon 20*, 9–16.

Darbon, H., Zlotkin, E., Kopeyan, C., Van Rietschoten, J., & Rochat, H. (1982) *Int. J. Pept. Protein Res. 20*, 320–330.

De Dianous, S., Kopeyan, C., Bahraoui, E.-M., & Rochat, H. (1987) *Toxicon 25*, 731–741.

De Lima, M.-E., Couraud, F., Lapied, B., Pelhate, M., Diniz, C. R., & Rochat, H. (1988) *Biochem. Biophys. Res. Commun. 151*, 187–192.

Fontecilla-Camps, J. C., Almassy, R. J., Suddath, F. L., Watt, D. D., & Bugg, C. E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6496–6500.

Fontecilla-Camps, J. C., Almassy, R. J., Ealick, S. E., Suddath, F. L., Watt, D. D., Feldmann, R. J., & Bugg, C. E. (1981) *Trends Biochem. Sci. 6*, 291–296.

Fontecilla-Camps, J. C., Habersetzer-Rochat, C., & Rochat, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7443–7447.

Gordon, D., Zlotkin, E., & Catterall, W. A. (1985) *Biochim. Biophys. Acta 821*, 130–136.

Grishin, E., Volkova, J. M., & Soldatova, L. N. (1982) *Bioorg. Khim. 8*, 154–164.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Jover, E., Couraud, F., & Rochat, H. (1980) *Biochem. Biophys. Res. Commun. 95*, 1607–1614.

Kharrat, R., Darbon, H., Rochat, H., & Granier, C. (1989) *Eur. J. Biochem. 181*, 381–390.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Lester, D., Lazarovici, P., Pelhate, M., & Zlotkin, E. (1982) *Biochim. Biophys. Acta 701*, 370–381.

Loret, E. P., Mansuelle, P., Rochat, H., & Granier, C. (1990) *Biochemistry 29*, 1492–1501.

Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528–533.

Müller, N., Bodenhausen, G., Wüthrich, K., & Ernst, R. R. (1985) *J. Magn. Reson. 65*, 531–534.

Pashkov, V. S., Maiorov, V. N., Bystrov, V. F., Hoang, A. N., Volkova, T. M., & Grishin, E. V. (1988) *Biophys. Chem. 31*, 121–131.

Piantini, U., Sorensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.

Possani, D. L. (1984) in *Handbook of Natural Toxins* (Tu, A. T., Ed.) Vol. 2, pp 513–550, Marcel Dekker, Inc., New York.

Rance, M., Sorensen, O. N., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Rochat, H., Rochat, C., Kupeyan, C., Miranda, F., Lissitzky, S., & Edman, P. (1970) *FEBS Lett. 10*, 349–351.

Rochat, H., Bernard, P., & Couraud, F. (1979) in *Advances in Cytopharmacology* (Ceccaldi, B., & Clement, F., Eds.)

Vol. 3, pp 325–333, Raven Press, New York.
Roussel, A., & Cambillau, C. (1989) in *Silicon Graphics Geometry Partners Directory (Fall 1989)* (Silicon Graphics, Eds.) pp 77–78, Silicon Graphics, Mountain View, CA.
Wagner, G. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 101–139.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 160*, 343–361.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Wüthrich, K. (1989) *Acc. Chem. Res. 22*, 36–44.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.
Zlotkin, E., Fraenkel, G., Miranda, F., & Lissitzky, S. (1971a) *Toxicon 9*, 1–8.
Zlotkin, E., Rochat, H., Kupeyan, C., Miranda, F., & Lissitzky, S. (1971b) *Biochimie 53*, 1073–1078.
Zlotkin, E., Kadouri, D., Gordon, D., Pelhate, M., Martin, M.-F., & Rochat, H. (1985) *Arch. Biochem. Biophys. 240*, 877–887.

Structural Microheterogeneity of a Tryptophan Residue Required for Efficient Biological Electron Transfer between Putidaredoxin and Cytochrome P-450$_{cam}$†

Patrick S. Stayton and Stephen G. Sligar*
Department of Biochemistry, University of Illinois, Urbana, Illinois 61801
Received July 26, 1990; Revised Manuscript Received November 9, 1990

ABSTRACT: The carboxy-terminal tryptophan of putidaredoxin, the Fe_2S_2·Cys_4 iron–sulfur physiological redox partner of cytochrome P-450$_{cam}$, is essential for maximal biological activity [Davies, M. D., Qin, L., Beck, J. L., Suslick, K. S., Koga, H., Horiuchi, T., & Sligar, S. G. (1990) *J. Am. Chem. Soc. 112*, 7396–7398]. This single tryptophan-containing protein thus represents an excellent system for studying the solution dynamics of a residue directly implicated in an electron-transfer pathway. Steady-state and time-resolved measurements of the tryptophan fluorescence have been conducted across the emission spectrum as a function of redox state to probe potential structural changes which might be candidates for structural gating phenomena. The steady-state emission spectrum (λ_{max} = 358 nm) and anisotropy (α = 0.04) suggest that Trp-106 is very solvent-exposed and rotating partially free of global protein constraints. The time-resolved fluorescence kinetics for both oxidized and reduced putidaredoxin are fit best with three discrete components of approximately 5, 2, and 0.3 ns. The lifetime components were assigned to physical species with iodide ion quenching experiments, where differential quenching of the longer components was observed ($k_{\tau=2}$ = 5.9 × 10^8 M^{-1} s^{-1}, $k_{\tau=5}$ = 1.3 × 10^8 M^{-1} s^{-1}). These findings suggest that the multiexponential fluorescence decay results from ground-state conformational microheterogeneity and thus demonstrate that the essential tryptophan exists in at least two distinguishable conformations. Small differences in the relative proportions of the components between redox states were observed but not cleanly resolved. It thus appears that Trp-106 exists in similar microconformations for both oxidized and reduced putidaredoxin, characterized by three lifetimes with similar relative proportions. These results do not rule out differences within the cytochrome P-450$_{cam}$ complex and indeed suggest potential mechanisms for selection of specific geometries. The absence of fluorescence lifetime wavelength dependencies demonstrates that each species displays the red-shifted, solvent-exposed, steady-state spectra. Differential solvent exposure is thus unlikely to be the primary mechanism for the asymmetric iodide quenching. Differential interactions of the tryptophan microconformations with an anionic protein surface are proposed as a mechanism for the iodide effects and supported by quenching studies with the neutral quenching agent acrylamide. The mechanism is consistent with a physical model of putidaredoxin–cytochrome P-450$_{cam}$ complex formation where the negatively charged surface is suggested to be the anionic binding surface of putidaredoxin implicated in putidaredoxin reductase and cytochrome P-450$_{cam}$ interactions [Stayton, P. S., & Sligar, S. G. (1990) *Biochemistry 29*, 7381–7386]. This model thus places the essential Trp-106 directly at the cytochrome P-450$_{cam}$ binding interface in position to mediate electron transfer.

Biological electron transfer is an extremely complex process that has attracted a great deal of theoretical and experimental effort (Marcus, 1956; Hopfield, 1974; Jortner, 1976; Marcus & Sutin, 1985; Scott et al., 1985; McLendon, 1988). The role of the protein medium in directly modulating biological electron-transfer processes has been the subject of intriguing but relatively fewer experimental and theoretical investigations. Perhaps the most direct demonstration of such a role has been the observation that an aromatic side chain is required at amino acid position 82 in cytochrome *c* for efficient electron transfer to occur with Zn protoporphyrin IX substituted cytochrome *c* peroxidase (Liang et al., 1988). Gray (Therien et al., 1990) and Wolynes (Kuki & Wolynes, 1987) have considered specific through-bond pathways in cytochrome *c* and myoglobin and mutagenesis studies with the blue copper protein azurin have also related electron-transfer kinetics with a specific electron-transfer pathway (van de Kamp et al., 1990). In addition, orientational effects on observed rates,

† This research was supported by National Institutes of Health Grant GM 31756.
* To whom correspondence should be addressed.

0006-2960/91/0430-1845$02.50/0