Hevein: NMR Assignment and Assessment of Solution-State Folding for the Agglutinin–Toxin Motif[†,‡]

Niels H. Andersen,[*,§] Bolong Cao,[§] Adela Rodríguez-Romero,[∥] and Barbarín Arreguin[∥]

Department of Chemistry, University of Washington, Seattle, Washington 98195, and Instituto de Química, Universidad Nacional Autônoma de México, Ciudad Universitaria, México D. F. 04510, Mexico

Received June 9, 1992; Revised Manuscript Received September 26, 1992

ABSTRACT: The first high-resolution solution-state structure of a member of the toxin–agglutinin folding motif with the WGA disulfide linkage is presented. The ^{1}H NMR spectrum of hevein has been 100% assigned from residue 2 through residue 43, the C-terminus, using two-dimensional correlation and NOE spectroscopy. During the course of the NOESY analysis, the three-dimensional structural features of hevein were derived, using nonstereospecific distance constraints (*with tight bounds*) for XPLOR simulated annealing followed by unconstrained relaxation in the CHARMm force field, at two levels of long-range constraint density. In addition, a large number of low-bound-only constraints, corresponding to unobserved NOE's, were used in both refinements. The first structure elucidation employed a total of 180 distance constraints (60 of which were medium or long range, $i/i+n$ with $n \geq 2$). The second refinement employed 244 (101 medium or long range) constraints: some conformation-insensitive intraresidue constraints were deleted, two misassigned long-range constraints were corrected, and 41 new $i/i+n$ ($n \geq 2$) constraints were added. The average bounds precisions of the two refinements were comparable (±0.44 Å) and significantly tighter than those that result when a universal low bound corresponding to the sum of the van der Waals radii was used. (The more conservative treatment of NOE's gave the same final structure but required a higher constraint density before assignment errors would stand out during the refinement.) Constraint density also has a significant influence on convergence and accuracy using tight constraints. The study demonstrates that convergence within an ensemble of solution structures is not a dependable criterion for either the accuracy or precision of the derived structure. The best fitting conformers from the refinement at the higher constraint density bear a greater similarity to the solid-state structure of the domains of wheat germ agglutinin (0.95 Å rmsd over residues 2–32) than to the recently reported 2.8-Å X-ray structure of hevein (1.25 Å rmsd over residues 2–32, 2.83 Å rmsd over residues 2–42). The consensus conformer from the solution data is defined to a backbone rmsd of <0.6 Å over the full sequence for which NMR data could be collected. The dominant architectural features in the solution state are (a) an antiparallel β-sheet formed by the sequence from Leu[16] to Ser[26] with a reversing loop turn at 20–22; (b) a less well aligned third strand, consisting of contacts between 37 → 39 and 20 ← 18, which places the otherwise unconstrained C-terminal loop in close contact with the core; (c) a short helix from Asp[28] to Ser[32] followed by two additional residues with α_R conformations; and (d) an unusually high proportion of non-glycine α_L units. The latter include all four Asn residues in many of the conformers generated. We propose that the specific locations of the Asn and Gly residues and the resulting local secondary structure preferences are fold determinants in this motif. Three loci of residual backbone motion (residues 13/14, 34/35, and 27/28) are evident in the NMR data. It remains to be established which of the structural and motional features observed are associated with the agglutinizing and immunological activities of hevein.

The agglutinin–toxin folding motif was first recognized (Drenth et al., 1980) as a pattern of closely spaced disulfide linkages. If the two vicinal Cys residues are uniformly numbered as 17, 18 [their sequential positions in hevein (Walujono et al., 1976)] and the first of four nearly identical domains in wheat germ agglutinin, WGA[1] (Wright, 1987), the Cys placement pattern is readily recognized. The four domains of WGA have a highly conserved structure in the crystalline state based on the 1.8-Å-resolution X-ray structure (Wright, 1987). The "short" postsynaptic snake toxins [for examples, see Low et al. (1976) and Négrerie et al. (1990)] are obviously more distantly analogous but have also been included as members of this folding motif (Drenth et al., 1980). The snake toxins are excluded from the remaining discussion. The roles of WGA and hevein in their plant sources appear to be unknown. Most of the proteins in this class are recognized toxins or allergens. In the case of hevein, which is the major protein in the bottom fraction of ultracentrifuged rubber-tree latex (Archer, 1960), such allergic reactions are a serious concern with latex condoms serving as a key method for limiting the spread of HIV.

To our knowledge no definitive solution structure elucidation is available for any protein with the disulfide linkage pattern of WGA. A preliminary account of NMR studies of the ragweed allergen Ra5G was presented (Warren et al., 1991). In that account the NMR structures did not converge well

[†] This research was supported by Office of Naval Research contract and grant funds (N00014-88-K-202 and -91-J-1393) which are hereby acknowledged.

[‡] The coordinates of three representatives of the major α_L conformer (HαLN-06, HαLN-07, and HαLW-01) and one set of coordinates (HαRN-32) for the minor α_R form have been deposited in the Brookhaven Protein Data Bank. The same record also includes models HαLN-R6 and HαLW-R1, in which the interaction of the C-terminal loop with the remainder of the structure has been optimized by a further refinement stage.

[*] Address correspondence to this author.

[§] University of Washington.

[∥] Universidad Nacional Autónoma de México.

0006-2960/93/0432-1407$04.00/0

FIGURE 1: Sequence homology between hevein, Ra5G, and the four domains of wheat germ agglutinin. Identities or conservative substitutions (including Y → F, G → N or D, Y → W, Q → K, and F → L or M as hydrophobic side chains) are shown as bold one-letter symbols. Prolines that are underlined are not conserved. Gaps in the sequence, when the Cys residues are aligned, appear as dashes. The position of the dash for deletions in longer loops is such as to preserve the greatest possible homology for the non-Cys residues. For Ra5G, the alignment continues only through the fourth Cys residue, and a divergent linkage pattern has been established (Metzler et al., 1992).

(Cα rmsd >2 Å), and the authors suggested that this may be "an example of structural resolution being limited by motional disorder". More recent studies (Metzler et al., 1992) have established that a different disulfide linkage pattern is present in that allergen, and its solution structure has been defined to a <0.7-Å backbone rmsd from residues 5–39. With the distinct linkage isomerism there is no reason for a structural analogy between hevein and Ra5G; however, given the degree of homology to WGA, one would certainly expect very extensive local structural correspondences and a similar overall fold (see Figure 1).

We now report the complete nonstereospecific assignment of the NMR spectra of hevein in 10% aqueous dioxane and two stages of conformation elucidation using NOESY-derived distance constraints (DCs). At both stages (which differ in constraint density), we have employed tighter bounds for the DCs than is typical of NMR structure elucidations of proteins, a strategy (Andersen et al., 1992) which is still exploratory. The results in the present case were confirmed by parallel refinements using a conservative conversion of NOE intensities into looser bounds. The initial set of 180 distance constraints afforded a moderate resolution picture of the aqueous folding geometry and (when tight distance constraints were employed) provided a surprisingly well-defined core structure which was distinguishably differentiated from that of WGA. At the stage of the final refinement (using 136 interresidue and 108 intraresidue distances with both high and low bounds and a large set of low-bound-only constraints), the consensus conformer (defined to a backbone rmsd of <0.6 Å over the full sequence for which NMR data could be collected) still differs from the WGA structure (0.9-Å backbone rmsd over residues 2–32) and differs to an even greater extent (2.8-Å backbone rmsd over residues 2–42) from an independent X-ray structure of hevein that appeared[2] during the course of this study.

EXPERIMENTAL AND COMPUTATIONAL METHODS

Sample Preparation. Hevein was isolated from the heavy fractions of *Hevea brasiliensis* latex as previously described (Rodríguez et al., 1986) and used without further purification. NMR samples were prepared by dissolving 18 mg of hevein in 0.5 mL of either 90/10 D_2O/dioxane-d_8 or 70/20/10 H_2O/D_2O/dioxane-d_8 in which the pH was maintained at 6.6 with 60 mM phosphate. The phosphate buffer was prepared to a measured pH of 6.6 at 100 mM in protic media and then repeatedly lyophilized using reconstitution with 99.99+% D_2O. The final reconstitution with 70/20/10 H_2O/D_2O/dioxane-d_8 reproduced the nominal pH to ±0.1. The acidic media, pH 2.4, 2.9, and 3.6, were obtained from the respective protic or deuteronic sample at pH 6.6 by titration with either CF_3-COOH or CF_3COOD.

CD Spectroscopy. Aliquots from the NMR sample were diluted to ca. 50 μM with 10 mM aqueous phosphate buffer. The concentrations of the solutions were determined by UV assuming a molar absorbtivity of ϵ_{280} = 12 660, corresponding to the expectation for the aryl side chains (Johnson, 1990). CD spectra were recorded in 1.0-mm path-length cuvettes at normal operating temperature (ca. 25 °C) using a JASCO Model J720 spectropolarimeter. The wavelength and degree ellipticity scales were calibrated using the *d*-10-camphorsulfonic acid (CSA) sample provided by the manufacturer, assuming that the CSA minimum corresponds to $\theta_{192.5}$ = −15 600 (Yang et al., 1986). Typical spectral accumulation parameters were a time constant of 0.25 s and a scan rate of 100 nm/min with a 0.2-nm step resolution over the range 178–270 nm with 16–24 scans averaged for each spectrum. The accumulated average spectra were trimmed at a dynode voltage of 650 prior to baseline subtraction and smoothing using the reverse Fourier transform procedure in the JASCO software.

NMR Spectroscopy. All NMR spectra were recorded at 500.13 MHz on Bruker AM-500 (or comparably upgraded WM-500) spectrometers. One-dimensional controls and exchange studies (at 0.5 → 7 mM) were recorded at a resolution of 0.71 Hz/point and revealed no significant concentration dependence for the δ values over the range examined. All two-dimensional spectra were acquired using 7 mM solutions in the phase-sensitive mode by using time-proportioned phase

[1] Abbreviations: CD ΔCD, circular dichroism and difference CD spectrum; DCs, distance constraints; FR, SA structures have been relaxed (without NOE distance constraints) into the nearest conformational minimum; HIV, human immunodeficiency virus; HαLN-06, HαLN-07, HαLW-R1, HαRN-32, etc., specific conformer models from the ensemble of solution structures of hevein; LBOs, low-bound-only distance constraints; MM and MD, molecular mechanics and dynamics; SA, dynamics simulated annealing procedure or structures generated from it; TFA, trifluoroacetic acid; UV, ultraviolet absorption spectrum; *Amb. a.* V and *Amb. t.* V (=Ra5G), ragweed allergens, fraction 5; d_{ij} and r_{ij}, experimental estimates of interproton distances and those observed in the actual molecule or a structural model, respectively; rmsd, root-mean-square deviation measured over the N, Cα, C′, and O atoms of specified segments of the peptide backbone; rmswv, a specified weighted root-mean-square violation measure; α_R and α_L, amino acid residue conformations centered about $\phi = -70°/\psi = -40°$ and $\phi = +60°/\psi = +30°$, respectively; θ_λ, molar ellipticity at λ(nm) in units of deg·cm²/(residue·dmol). The common abbreviations of 2D NMR—NOESY, TOCSY, ROESY, COSY, DQF-COSY, TPPI, MLEV-17, NOE, t_1, t_2, τ_m, d_{NN}, $d_{\alpha N}$, etc.—are employed without further comment. Specific NOE interactions and the corresponding derived distance constraints are identified by amino acid (residue number or one-letter symbol and residue number) and hydrogen position—HN (or N), α, β, etc.—for each proton involved; e.g., 28α/31HN represents a cross-peak or constraints for the α-proton of residue 28 and the backbone NH of residue 31. All other references to individual residues are by symbol *and* residue number—L16 corresponds to the leucine at position 16—or use standard format, Leu^{16}.

[2] Professor Soriano-Garcia provided the coordinates of the published (Rodriguez-Romero et al., 1991) structure (R = 24.1%) from the 2.8-Å-resolution data set.

Table I: NOE Distance Categories for Tight Constraints[a]

NOE size[b]	distance constraint			
	type 1	type 2	type 3	type 4
VL (>0.5)	1.95–2.5	2.00–2.45		
L (0.25–0.49)	2.25–2.8	2.15–2.75	2.10–2.80	2.0–2.8
M+ (0.13–0.249)	2.4–3.1	2.35–3.1	2.40–3.05[d]	2.3–3.2
m– (0.06–0.129)	2.70–3.4	2.70–3.35[c]	2.65–3.25[d]	2.6–3.4
s (0.025–0.059)	3.00–3.9	3.10–3.8[c]	3.05–3.9[d]	2.9–3.9
vs (0.010–0.024)	3.20–4.4	3.20–4.3[c]	3.25–4.6[d]	3.1–4.6
unobserved (≤0.02)	≥3.3	≥3.4	≥3.5	≥3.3

[a] For connectivity types, the second listed subscript is the designation of the proton further along the peptide chain: thus $d_{N\alpha}$ is an i/i constraint while $d_{\alpha N}$ is an $i/i+n$ constraint with n a positive integer. Type 1 constraints are all unambiguous d_{NMe}, $d_{N\beta}$, $d_{\beta N}(i+1)$, and $d_{NN}(i+1)$ connectivities. Type 2 constraints are $d_{\alpha\beta}(i+3)$ and $d_{\alpha N}(i+n)$, when n = 1 or 3, and $d_{\beta N}(i+2)$ connectivities. Type 3 constraints are d_{XMe} and $d_{\alpha N}(i+n)$, when n = 2 or ≥4, and $d_{N_iN_j}$, $d_{\alpha_i\alpha_j}$, and $d_{N_i\beta_j}$, when j does not equal i, i–1 or i–2, connectivities. Type 4 constraints are all other connectivities, including not previously specified intraresidue ones. [b] NOE sizes are indicated on a scale from very large (VL) to very small (vs); the range of fractional intensities included in each class is given. [c] The upper bound is extended by 0.15 Å when $n \geq 2$. [d] The upper bound is extended by 0.25 Å for $d_{\alpha N}(i+2)$ constraints.

incrementation (Marion & Wüthrich, 1983; Drobny et al., 1979) at a resolution (in ω_2) of 2.8 Hz/point in protic media and 2.2 Hz/point in D_2O-rich media. The general features of data collection and processing have been detailed previously (Andersen et al., 1987, 1989, 1992). For most of the experiments in protic media, water suppression was achieved by saturation during the 0.9–1.7-s preparatory delay (and mixing time in the case of NOESY spectra) with the decoupler cycled between six different offsets (Zagorski, 1990a) rather than a single set value. With this decoupler cycling procedure, we were able to suppress the water resonance more completely and obtained less distortions in the baselines.

For NOESY spectra (Bodenhausen et al., 1984) 544 → 740 t_1 increments were collected, each with 32–64 scans of 2K complex data points. For COSY (Marion & Wüthrich, 1983; Bodenhausen et al., 1984), RELAY-COSY (Wagner, 1983; Bax & Drobny, 1985), and DQF-COSY (Marion & Wüthrich, 1983; Rance et al., 1983) experiments, 800 → 1024 t_1 increments were collected each with 24 or 32 scans of 4K or 2K points. For TOCSY spectra (Bax et al., 1985), 754 → 896 t_1 increments were collected, each with 32 scans of 4K complex points. The RELAY experiments employed two 17-ms delays separated by a π pulse. The TOCSY experiments employed an MLEV-17 spin lock (of 18-, 43-, 60-, and 83-ms duration in D_2O media; 75 or 90 ms for protic media). Delays were introduced on both sides of the π pulse to eliminate ROE peaks formed during the spin lock (Griesinger et al., 1988). For the 180-ms ROESY spectrum (Kessler et al., 1987), 600 t_1 increments were collected, each with 80 scans of 2K complex data points. The pulse sequence was $90°–t_1–(\beta–\tau–)_{\gamma}–\pi)_n–t_2$, with all pulse generated by the decoupler; the π pulse was incorporated to remove phase anomalies and resonance offset effects caused by the soft 90° preparation pulse (Zagorski, 1990b). All 2D data sets were processed using FTNMR and FELIX (Hare Research Inc.) after transfer from the Aspect 3000 to Silicon Graphics 4D-20G workstations. For the NOESY data sets, the first points in t_1 were multiplied by 0.5 (Otting et al., 1986).

To resolve spectral overlap, data was collected at two pHs and at temperatures of 301 ± 2, 310, 315, and 320 ± 1 K at both pH 6.6 and in the more acidic media. The NOESY spectra employed mixing times from 80 to 600 ms. The τ_m = 600 ms NOESY (in D_2O) was used to clarify assignments for otherwise ambiguous cross-peaks by intraresidue relays of the NOE's to alternate sets of non-shift-coincident protons.

Backbone NH Exchange Rates. These were initially estimated as $t_{1/2}$ values for signal disappearance in 1D spectra recorded at pH ≈ 7.2 for a sample freshly dissolved in the D_2O media. Additional experiments used to confirm the assignments and to categorize the signals as fast, intermediate, or slowly exchanging followed the protocols described here: (a) the reappearance of specific NH signals (or COSY cross-peaks) for a fully exchanged sample upon dissolution in pH = 3.1 media at 311 K and (b) storage of the protic form in D_2O media for 6 days at pH = 6.6 (at ~278 K), followed by pH adjustment (→2.8) and an extensive COSY data set (collected over 14 h) for the detection of the remaining weak HN/Hα peaks.

NMR Structure Generation and Refinement Procedures. The distance constraints used were classed in six NOE size catogories; the high and low bounds ascribed to each when used as tight constraints appear in Table I. For each NOESY recorded with $\tau_m \leq 180$ ms, we assigned the largest nongeminal cross-peak an arbitrary unit intensity (on the basis of peak height) and assigned distance constraints on the basis of relative peak heights according to the scheme summarized in Table I. The relative intensities for peaks in the first three categories were only taken from spectra with $\tau_m \leq 120$ ms. At least three spectra were used to define the NOE size for each constraint, justifying the use of a larger number of upper-bound categories. If a specific connectivity fell into different classes in the spectra examined, the shortest applicable low bound and longest applicable high bound were employed. The NOESY spectra recorded with τ_m = 200–300 ms were used only to quantitate weaker (s and vs) long-range connectivities for which no obvious secondary NOE pathways exist. For constraints involving prochiral methylene, if the pair of interactions were of approximately equal intensity at the shortest mixing time, both constraints were used; otherwise, only the stronger NOE was converted into a wild-card constraint, and the upper bound was increased by 0.2 Å. For constraints involving methyl groups the upper bound was increased by 0.5 Å. With this scheme the average bounds precision was ±0.44 Å. *For the alternate "conservative" bounds treatment, the upper bounds were uniformly increased by 4% and the low bounds were decreased to 1.9 Å for categories VL, L, and M+; 2.2 Å for m–; and 2.4 Å for the smaller NOE's. These changes should accommodate spin diffusion and motional averaging effects if they are present in the structure.*

XPLOR-2.1 (Polygen Corp.) was employed for a three-stage simulated annealing (T < 1000 K during the high temperature dynamics with cooling to 300 K) protocol essentially the same as that we have published for the structure elucidation of endothelin-1 (Andersen et al., 1992) with the following exceptions: no torsion constraints were used at any stage, the dynamics step size was 2.0 fs rather than 0.5 fs, and the NOE scale factor was greater by a factor of 2 during both

the cooling and Powell minimization steps. The general procedure is thus very similar to that described by Kraulis et al. (1989) in the use of exaggerated bond length, angle, and improper torsion force constants to maintain regularized covalent geometry but differs with respect to the manner in which NOE potentials were ramped, rescaled, and converted from "soft-square" to biexponential "square well" and the ramping and final values of the van der Waals repulsion term during the course of each stage of simulated annealing (SA). The initial structures, which included Glu[1] even though no constraints were available for this residue, were seven randomly generated by XPLOR, a fully extended ($\phi = \psi = 180°$) structure (used for six runs), and a fully α-helical structure (used for four runs); two runs were based on a homology structure generated from WGA, and seven runs were initiated from a constructed model which contained the antiparallel β-sheet suggested from the qualitative analysis of the diagonal plot (vide infra). The three stages of SA consisted of the following: first stage, predisposing acyclic structures for disulfide closure; second stage, high-temperature constrained dynamics of cyclized structures; and third stage, further SA with added low-bound-only constraints (vide infra). After each stage of SA, structures were cooled to 300 K in 25° steps each of 100-fs duration and then subjected to a brief Powell minimization. Hydrogen bond potentials were excluded throughout the procedure. We refer to the cooled structures at the third stage as the "NMR" or "SA" structures. For some starting structures, particularly those that were helical for the entire length of the sequence, it was necessary to repeat the third stage of the annealing protocol one or two additional times in order to achieve an acceptable fit to the distance constraints. The residual violations of the constraints were assessed in two ways: 1) from the E_{NOE} values reported by the XPLOR program after each Powell minimization (during which k_{NOE} values were set to 50 kcal·mol^{-1}·Å^{-2} for sequential and long-range connectivities and half of that value for intraresidue and low-bound-only connectivities) and (2) using functions available within the penalty function scan (PFS) program of DISCON (Andersen et al., 1991). Of the latter, the fraction of bounds violated by >0.2 Å, the sum of absolute violations (or average violation in each constraint class), and a weighted violations measure defined as

$$\text{rmswv} = \text{rms weighted violations} = \left[\frac{1}{n-1}\sum_k\left(\frac{\text{viol}}{\text{range}}\right)_k^2\right]^{1/2} \quad (1)$$

were the most useful in the present study.

Constraints were of the form d, d_-, d_+, where d is the estimated distance and d_- and d_+ are the allowed excursions to the low and high side of the estimate (see Table I). A violation is defined as any excursion of r_{ij} outside of the constraint range, $d - d_- \rightarrow d + d_+$; and range, in the equation for rmswv, is given by $d_- + d_+$. For constraints with no upper bound, only the violation of the lower bound was used for both the calculation of the rmsd and all fit measures based on violations (the meaningless range value is replaced by the average range for all constraints with both high and low bounds). The latter, designated as low-bound-only constraints (LBOs), are analogous to the antidistance constraints (ARCs) used by Ernst (Brüschweiler et al., 1991) and more recently by Kopple (Bean et al., 1992).

The SA structures were further relaxed by a 500-step steepest descent minimization without NOE constraints in the default CHARMm force field (Brooks et al., 1983), excluding the H-bond term, of the commercial package from Polygen Corp. The resulting structures, which display no violations of standard geometry (Table II), are designated as freely relaxed (FR) structures and were used in the structural comparisons and figures unless specified to the contrary.

Table II: First-Generation Structures and Geometry Violation Statistics

	SA structures	FR structures[d]
XPLOR values		
bond rmsd[a]	0.007 ± 0.001	0.019 ± <0.0004
angle rmsd[b]	2.30 ± 0.03	3.34 ± 0.07
improper rmsd[b]	1.12 ± 0.08	6.98 ± 0.55
E_{repel} (r = 0.8)[c]	30.5 ± 10.2	10.7 ± 4.5
CHARMm E_{L-J}[c]	67.4 ± 31.4	−185 ± 7.0

[a] Given in Å units. [b] Given in degrees. [c] Given in kcal/mol (1 kcal/mol = 4.18 kJ/mol). [d] At least 50% of the deviations reported by the XPLOR program are due the substantial differences in parameters between the CHARMm force field and the greatly simplified parameter set used in XPLOR.

Structure Comparisons and Conformer Definition by a Second Stage of Refinement. (A) General Methods. The backbone rms measures of structural convergence were performed on the FR structures using the routines within INSIGHT (Biosym Technologies Inc.), which were also used to generate the structural figures. INSIGHT was also used to search structures for potential hydrogen bonds and for solvent-inaccessible amide NH atoms. The results of the former search were further trimmed to meet the following more restrictive criteria—N–H···O = 1.96 ± 0.25 Å and N···O = 2.95 ± 0.30 Å. An NH is deemed inaccessible if it does not contribute to the solvent Connolly surface obtained with the default 1.4-Å probe. Likewise, side-chain groups were judged as fully internalized or partially exposed on the basis of their absence or partial presence in the 3.4- and 2.4-Å-probe Connolly surfaces, respectively.

(B) Conformer Clusters and Second-Generation Refinement. The first-generation FR structures were grouped in conformer clusters on the basis of the backbone dihedral similarities over segments of 4–10 residue length (with particular attention directed to key loci such as residues 8, 14–15, 26–27, and 35), the occurrence of a common set of hydrogen bonds with expectation geometry, and a similar pattern of constraint violation. The best fitting FR structure in each cluster was rerefined by simulated annealing, this time including (if NH exchange data indicate significant sequestration from water) the specific H-bonds that define the cluster as distance constraints. Each H-bond was defined by two distances, as given above, and these constraints were given the same weight as the key NOE constraints (which included only those interresidue connectivities involving at least one spin site from within the residue 16 → 34 core). The remaining interresidue NOE constraints were given 53% of this weighting. Intraresidue constraints were given an even lower weighting (33%). By the time these refinements were performed, additional NOE's had been confirmed producing constraints for 16 intraresidue, 9 sequential, and 41 long-range connectivities. Two previously used constraints (19α/33δ* and 19β*/33δ*) had to be eliminated since they were found to be ambiguous or incorrectly attributed. The complete final set of constraints is given in the supplementary material. Three rounds of SA refinement were used; in the last round all H-bond and torsion constraints were eliminated so that the final structures reflect only the NOE constraints. The final SA structures were relaxed as previously (but including the H-bond potential). The resulting structures appear as conformer models in Figure 9. Deviations from ideal geometry

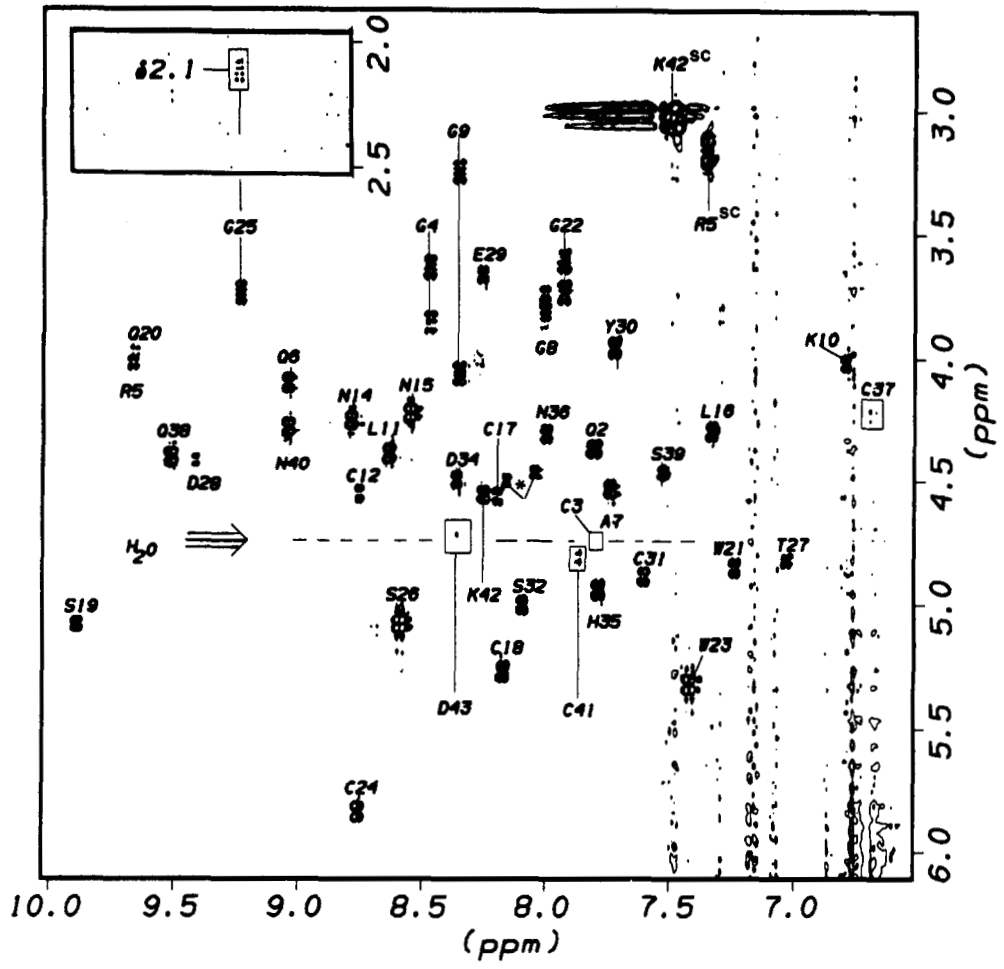

FIGURE 2: pH = 2.4 COSY spectrum (HN/α fingerprint region) of hevein at 310 K. The insert shows the extreme upfield Gly α. Completely or partially bleached α-methines are enclosed in boxes. The asterisk indicates impurities or isoforms.

for this final set of structures were fully comparable to those given in Table II for FR structures. The CHARMm energies ranged from −1442 to −1047 kcal/mol, of which those lower than −1290 kcal/mol were retained. The $E_{L\text{-}J}$ values ranged from −190 to −146; −178 ± 7 kcal/mol for the retained structures.

All computations, structure refinements, and comparisons were carried out on Silicon Graphics 4D-20G or 4D-25TG workstations.

RESULTS AND DISCUSSION

Due to the numerous multiple occurrences of identical (or very similar) residues and the extensive resonance overlap in the β,β' and HN spectral ranges, the standard (Wüthrich, 1986) two-stage (spin system correlation followed by sequential) NOESY assignment procedure failed. A hybrid, iterative "main-chain-directed"/spin-system correlation strategy was developed.

The assigned pH = 2.4 COSY appears as Figure 2. Forty correlations (including five as pairs of peaks, due to Gly residues) are expected; 38 are observed in the COSY shown: the C3 and D43 cross-peaks are absent due to α-H bleaching; they can, however, be observed at other temperatures as the water resonance moves. At pH = 6.6, two backbone NHs (Q2 and D28), a few $-CONH_2$–E/Z cross-peaks, and the Arg and Lys side-chain NHs are lost due to cross-saturation during the ca. 1.5-s water suppression pulse. NOESY spectra for each pH were recorded at several temperatures (see supplementary material, Figure 6) in order to differentiate the signals that appear at any particular NH chemical shift.

The pH changes were useful for the assignment of the obvious amino acids that have titratable side chains. This can be illustrated (Figure 3) by the differentiation of the numerous Asp and Asn residues: the D_2O COSY spectra were particularly useful for this, since the α-methines of these residues tend to be close to the H_2O signal. TOCSY and RELAY experiments provided a few key intraresidue correlations that had not be assigned (or were ambiguous) in the NOESY spectra. TOCSY experiments using an MLEV-17 spin lock gave correlations out the full length of side chains in most cases in both protic (90 ms) and deuteronic (60- or 83-ms spin-lock time) media. The resulting HN/γ, HN/δ, and α/δ correlations were particularly significant for (a) assigning well-dispersed backbone–NH resonances associated with α-methines that are nearly shift coincident or partially obscured near the water resonance and (b) distinguishing intra- and interresidue correlations in the corresponding NOESY spectra. Figure 4 illustrates this with some rather remarkable chemical shifts that were observed for Q20. The downfield portion of the α-methine region shows the excellent differentiation of α/β cross-peaks based on the size of the passive $\alpha\beta$ scalar coupling constant (Driscoll et al., 1989); β-hydrogens are indicated as either β_a/β_g (for their anti or gauche relationship to Hα) or as β/β' with the primed species as the upfield member of the shift-resolved methylene. In concert with the NOE intensities, these should provide stereospecific assignments for at least 60% of the prochiral β-methylenes.

Sequential Assignment. Since the full sequential assignment proved unusually difficult, we will not present the complete process here; rather, the key steps and illustrative

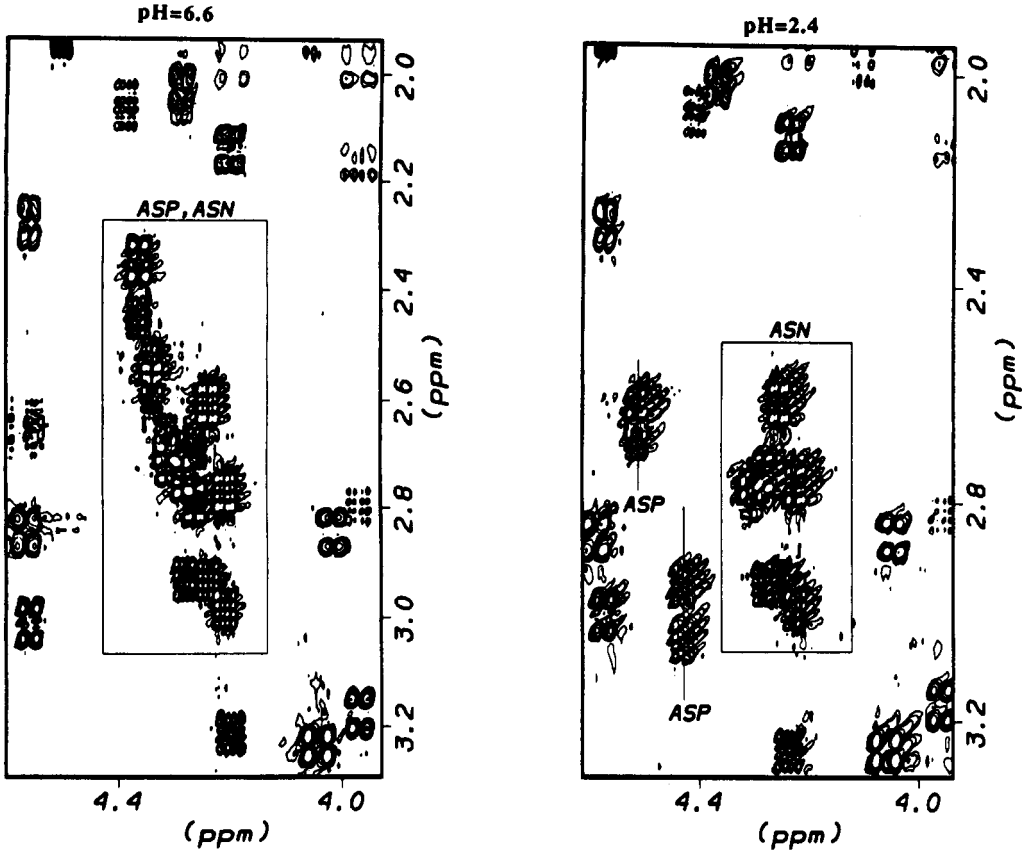

FIGURE 3: Effect of pH upon the $\alpha/\beta,\beta'$ cross-peak region containing the Asn and Asp connectivities in the COSY spectra recorded at 310 K. At pH 2.4, the Asp[43] α resonance is shifted downfield (δ = 4.69) of the segment displayed.

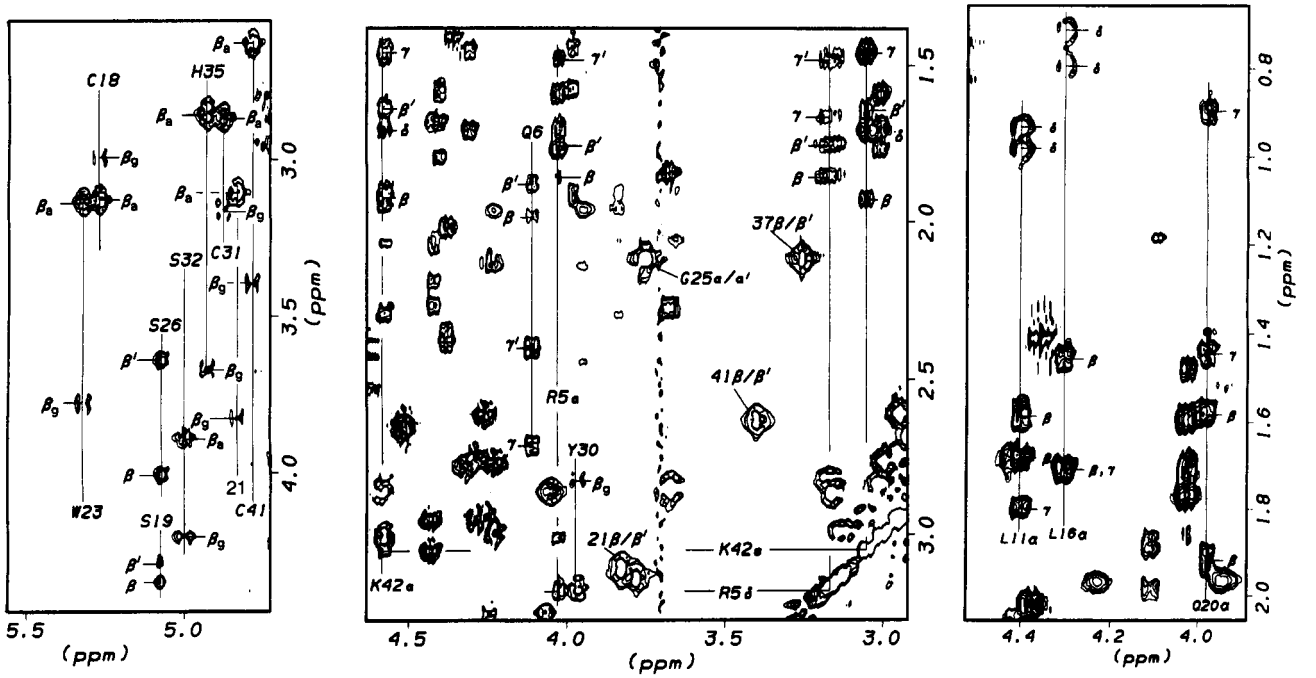

FIGURE 4: Portions of the TOCSY spectrum recorded at pH = 2.4 and 310 K with a 60-ms MLEV-17 spin lock. Panel A (left) shows the distinct cross-peak multiplicity patterns that allow for the assignment of antiperiplanar and gauche α/β interactions. Panels B and C (center and right) show the delineation of long side chains and illustrate a "normal" (Q6 in panel B) and unusual (Q20 in panel C) set of shifts for glutamine side-chain protons.

examples will be highlighted. The only unique residues of hevein are Ala[7] and Thr[27]. Other possible entry points for sequential assignment were provided by the aryl residues, which in every case gave δ/β and/or δ/α NOE's in D_2O media and a selection of δ/HN NOE's in protic media (supplementary material figures). The Tyr unit was located in a unique string, EYCSP, by $i/i+1$ connectivities. The WGW triplet was readily identified and could be connected to the first feature which we assigned: a very intense cross-strand α/α NOE which was identified as $18\alpha/24\alpha$ in this way. The key in- and cross-strand NOE's are outlined in Figure 5. Figure 5 also shows a key tertiary contact, 19α/38N, which is not forced

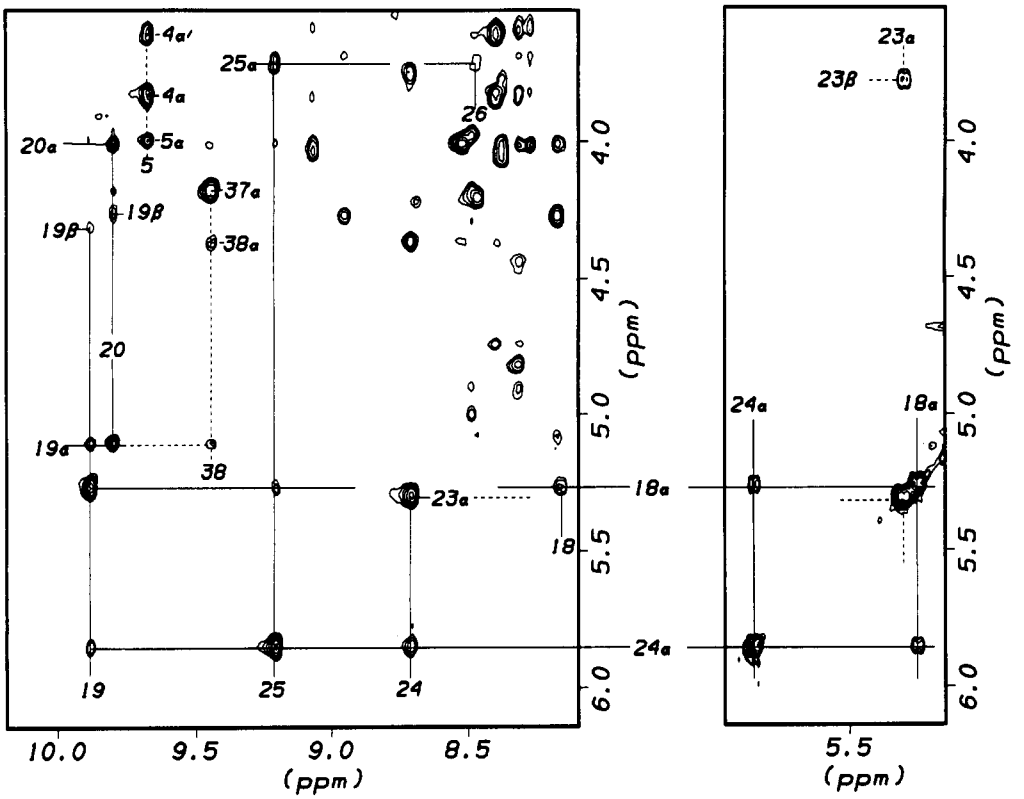

FIGURE 5: NOESY features associated with a short stretch of antiparallel β-sheet: pH = 6.6; left panel at 320 K, τ_m = 300 ms; right panel at 310 K, τ_m = 180 ms. A tertiary contact between residues 19 and 38 is also annotated.

by a nearby disulfide linkage. Other NOE's identified later in the assignment process (40/17, 38/18, 37/17, 36/20, 34/21) suggest an irregularly aligned third strand in the sheet.

Two examples of NOESY sequencing based on the unique Ala[7] appear in the supplementary material figures. At pH 2.4, a complete and unambiguous string of $d_{\alpha N}$ and/or $d_{\beta N}$ connectivities spanned from residues 3 to 12. This included Leu[11], and thus Leu[16] was assigned by default. This provided the remaining sequencing of the N-terminal region. The D_2O NOESY spectra showed two Pro spin systems which could be differentiated by the clear 12α/Pro[13]-δ NOE (supplementary material tissue). The near shift correspondence of the Ser[32]-β,β′ and the Pro[33]-δ,δ′ resonances proved to be confounding through most of the early assignment work.

The sequential connectivities are summarized in Figure 6, and the full listing of chemical shifts appears in Table III. At pH 2.4, the assignment is 100% complete, including all side-chain protons. However, no resonances corresponding to an N-terminal amino acid[3] could be identified with certainity at either pH. The pH behavior of the residues ascribed to the 2-position and to the C-terminal and C-penultimate positions is in accord with expectations for these locations.

NOESY-Based Derivation of the Aqueous Solution Fold. Only two secondary structure features could be surmised from qualitative features of the sequential NOE's: the previously mentioned antiparallel β-sheet structure and a very short

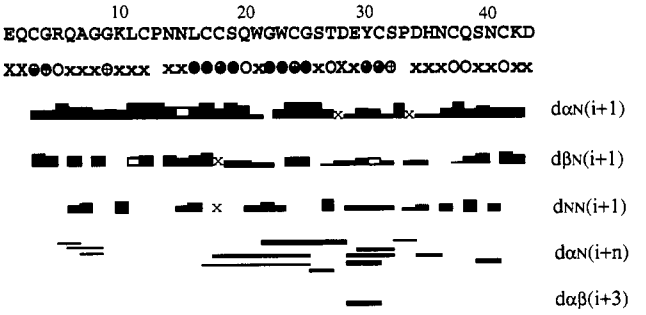

FIGURE 6: Sequential connectivity scheme and NH exchange data. The first row shows the exchange behavior: X indicates signals that disappear due to cross relaxation (during 1.7 s at pH = 6.6), x indicates signals which are significantly restored within 4 min after dissolution of the fully D-exchanged sample in H_2O at pH = 3.1 and T = 311 K; O indicates NH signals which appear during a subsequent 40-min period; a cross-filled O indicates NH signals not in the preceding categories but which display NH/αH COSY cross-peaks in a spectrum collected from 2 to 6 h after dissolution; a top-filled O indicates NH/αH cross-peaks which do not appear in COSY spectra until at least 8 h after dissolution (pH = 3.1, T = 311 K); a side-filled O indicates NH signals that are largely undiminished in intensity after 2 h at pH = 7.2 in D_2O solution at 315 K; a bottom-filled O indicates NH/αH cross-peaks which are still detectable (at pH = 2.8) after 6 days in D_2O at pH = 6.6 followed by a lyophilization and D_2O reconstitution and pH adjustment step. With this scheme any NH with a three-quarter or fully filled symbol can be viewed as slowly exchanging and is probably persistently H-bonded. The remaining rows show the sequential connectivities with intensities indicated by bar thickness: open bars indicate a potentially ambiguous assignment, an x indicates that the connectivity cannot be quantitated or confirmed due to shift coincidences. The connectivity categories are indicated in the right margin.

helical segment (Asp[28] → Ser[32]) indicated by a string of d_{NN} connectivities and a set of $i/i+3$ interactions. The latter were unambiguous in the NOESY spectra recorded at both pH values but were somewhat more pronounced at pH = 2.4. The short helix and antiparallel sheet, and a number of additional

[3] From the few remaining unassigned resonances, a weak case could be made for the N-terminus being partially Ala: δ (Hα, Hβ), 4.36, 1.41, pH 2.4; 4.08, 1.31, pH 6.6. At pH 6.6 one additional spin system was evident in the TOCSY spectra. This spin system (with very unusual chemical shifts—α, 2.36; ββ′, 1.85 and 1.53; γγ′, 2.47 and 1.74—if it is attributed to a Glu residue) displays NOE connectivities to the aryl side-chain protons of Trp[24]; however, in the absence of connections to 2HN (which cross-saturates rapidly at this pH) we view this as a provisional assignment.

Table III: Hevein Chemical Shifts

residue	NH	αH	βH[c]	others[d]
		(A) Conditions:	10% Dioxane, pH = 2.4, 310 K[a,b]	
Gln[2]	7.825	4.37	2.02	γH 2.37; ε2 7.64, ε1 7.095
Cys[3]	7.89	4.715	2.96, 2.555	
Gly[4]	8.49	3.86, 3.63		
Arg[5]	9.69	4.025	1.86, 1.75	γH 1.66, 1.465; δH 3.16; δNH 7.355
Gln[6]	9.05	4.105	1.965, 1.875	γH 2.710, 2.40; ε2 7.67, ε1 6.695
Ala[7]	7.745	4.55	0.775	
Gly[8]	8.03	3.845, 3.765		
Gly[9]	8.37	4.06, 3.25		
Lys[10]	6.805	4.025	1.755, 1.74	γH ~1.59; δH ~1.71; εH ~3.00; εNH 7.53
Leu[11]	8.655	4.40	1.68, 1.59	γ 1.80; δH 0.985, 0.935
Cys[12]	8.775	4.575	3.005, 2.85	
Pro[13]		4.57	2.29, 1.915	γH 2.065, 1.965; δH 3.83, 3.66
Asn[14]	8.81	4.255	2.945, 2.61	γNH 7.41, 6.885
Asn[15]	8.57	4.23	2.99, 2.765	δ2 7.34, δ1 6.78
Leu[16]	7.35	4.305	1.695, 1.445	γ 1.71; δH 0.795, 0.72
Cys[17]	8.215	4.59	4.05, 2.865	
Cys[18]	8.19	5.265	*3.13*, 2.99	
Ser[19]	9.91	5.08	4.35, 4.285	
Gln[20]	9.67	3.975	1.91, 1.575	γH 1.425, **0.86**; ε2 6.885, ε1 6.76
Trp[21]	7.26	4.84	3.83, *3.11*	δ1 7.305; ε3 7.715; ζ2 7.43, ζ3 7.165; η3 7.18; ε1 10.11
Gly[22]	7.95	3.735, 3.605		
Trp[23]	7.44	5.32	3.785, *3.14*	δ1 7.085; ε3 7.915; ζ2 7.515, ζ3 7.16; η3 7.24; ε1 10.25
Cys[24]	8.785	5.84	*3.005*, 2.84	
Gly[25]	9.25	3.745, **2.10**		
Ser[26]	8.605	5.07	4.005, 3.635	
Thr[27]	7.05	4.805	4.845	γ 1.37
Asp[28]	9.43	4.430	3.05, 2.945	
Glu[29]	8.285	3.675	1.84	γH 2.27
Tyr[30]	7.74	3.96	*3.17*, 2.82	δH 7.495; εH 6.78
Cys[31]	7.63	4.875	3.16, *2.865*	
Ser[32]	8.11	4.995	4.205, *3.885*	
Pro[33]		4.725	2.45, 1.96	γH 2.14, 1.96; δH 4.23, 3.945
Asp[34]	8.385	4.515	2.68, 2.62	
His[35]	7.80	4.93	3.67, *2.845*	δ2 7.175; ε1 8.705
Asn[36]	8.02	4.315	2.785	δ2 7.145, δ1 6.82
Cys[37]	6.70	4.235	3.25, 2.115	
Gln[38]	9.53	4.415	2.07, 1.685	γH 2.265, 2.19; ε1 8.30, ε2 8.205
Ser[39]	7.55	4.475	4.055, 3.695	
Asn[40]	9.06	4.286	2.95, 2.745	δ2 7.38, δ1 6.79
Cys[41]	7.89	4.785	3.395, *2.625*	
Lys[42]	8.25	4.575	1.915, 1.63	γH ~1.46; δH ~1.705 εH ~3.04; εNH 7.505
Asp[43]	8.37	4.70	2.95, 2.86	
		(B) Conditions:	10% Dioxane, pH = 6.6, 310 K[a,b]	
Gln[2]		4.30	2.045, 2.05	γH 2.465, 2.43; ε2 7.715, ε1 7.08
Cys[3]	7.85	4.71	2.96, 2.54	
Gly[4]	8.405	3.845, 3.62		
Arg[5]	9.735	4.045	1.85, 1.725	γH 1.655, 1.42; δH ~3.135
Gln[6]	9.055	4.07	1.935, 1.895	γH 2.845, 2.405; ε2 7.785, ε1 6.63
Ala[7]	7.735	4.535	0.745	
Gly[8]	7.975	3.83, 3.745		
Gly[9]	8.40	4.055, 3.24		
Lys[10]	6.785	4.02	~1.77	γH ~1.605; δH ~1.70; εH ~2.99
Leu[11]	8.615	4.405	1.680, 1.585	γ 1.795; δH 0.985, 0.935
Cys[12]	8.785	4.57	3.015, 2.86	
Pro[13]		4.57	2.285, 1.90	γH 2.065, 1.965; δH 3.84, 3.67
Asn[14]	8.78	4.25	2.94, 2.615	γNH 7.485, 6.795
Asn[15]	8.515	4.22	2.995, 2.78	δ2 7.35, δ1 6.735
Leu[16]	7.295	4.295	1.69, 1.465	γ 1.71; δH 0.795, 0.72
Cys[17]	8.175	4.58	4.03, 2.86	
Cys[18]	8.17	5.265	*3.13*, 2.98	
Ser[19]	9.945	5.125	4.385, 4.305	
Gln[20]	9.87	4.045	1.865	γH 1.62, **0.445**; ε2 7.280, ε1 6.935
Trp[21]	7.305	4.95	3.855, *3.12*	δ1 7.335; ε3 7.71; ζ2 7.45, ζ3 7.17; η3 7.205; ε1 10.175
Gly[22]	7.935	3.75, 3.645		
Trp[23]	7.395	5.315	3.79, *3.12*	δ1 7.05; ε3 7.91; ζ2 7.505, ζ3 7.16; η3 7.245; ε1 10.30
Cys[24]	8.745	5.86	*3.01*, 2.845	
Gly[25]	9.24	3.73, **2.075**		
Ser[26]	8.53	5.075	3.995, 3.64	
Thr[27]	7.045	4.785	4.84	γH 1.34
Asp[28]		4.325	2.71, 2.67	
Glu[29]	8.745	3.55	1.775, 1.715	γH 1.98, 1.895
Tyr[30]	7.885	3.98	*3.19*, 2.805	δH 7.485; εH 6.76
Cys[31]	7.595	4.885	3.12, *2.92*	
Ser[32]	8.51	4.99	4.21, 3.985	

Table III (Continued)

residue	NH	αH	βH[c]	others[d]
		(B) Conditions: 10% Dioxane, pH = 6.6, 310 K[a,b]		
Pro[33]		4.75	2.44, 2.025	γH 2.18, 1.995; δH 4.21, 3.98
Asp[34]	8.28	4.375	2.46, 2.355	
His[35]	7.60	4.715	3.525, *2.755*	δ2 6.915; ε1 7.95
Asn[36]	7.97	4.275	2.80, 2.69	δ2 7.115, δ1 6.78
Cys[37]	6.715	4.21	3.22, *2.15*	
Gln[38]	9.49	4.395	2.07, 1.675	γH 2.295, 2.19; ε1 8.43, ε2 8.34
Ser[39]	7.52	4.47	4.06, 3.69	
Asn[40]	9.015	4.295	2.94, 2.755	δ2 7.41, δ1 6.865
Cys[41]	7.81	4.81	3.425, *2.655*	
Lys[42]	8.35	4.515	1.935, 1.665	γH ~1.49; δH ~1.71 εH ~3.035
Asp[43]	7.795	4.35	2.600, 2.540	

[a] The chemical shifts are reported to the nearest 0.005 ppm relative to internal (trimethylsilyl)propionate. [b] The extreme upfield chemical shifts shown in bold face are readily rationalized as aryl ring shielding effects (see Figure 10). [c] For β protons, that one which is antiparallel to its vicinal α neighbor (based on the distinct TOCSY peak pattern shown by it and its geminal partner and/or from its COSY multiplet pattern) is italicized; those that display a gauche relationship ($J_{\alpha\beta} \leq 5$ Hz) to the α proton (based on similar criteria) are underlined. [d] For the side-chain NH_2 protons of Asn (δH) and Gln (εH), that one which is trans to the amide carbonyl oxygen is designated as δ2 (or ε2) and displays diagnostically larger NOE's with at least one of the β (or γ) protons (Montelione et al., 1992). As generally observed, the trans NH is downfield for those residues which display normal shifts.

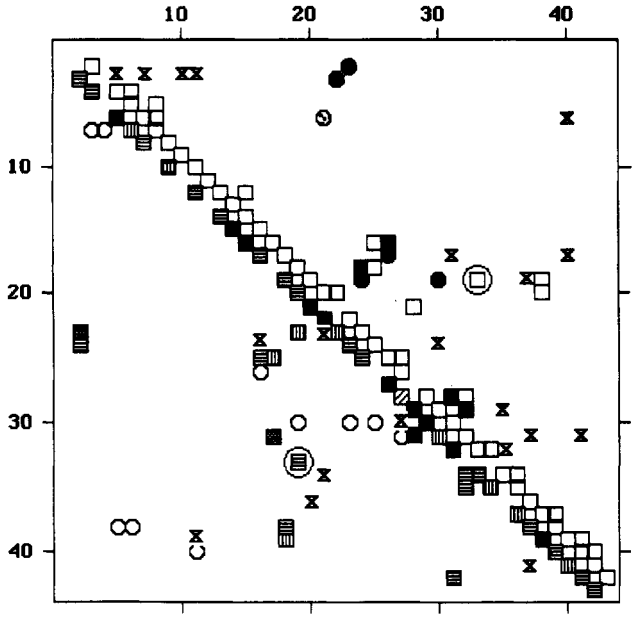

FIGURE 7: Diagonal plot of the NOE's observed in hevein. Both axes represent the amino acid sequence. The plot is not symmetrical; different connectivities are shown. Above the diagonal: □ indicates an NOE between $H_{\alpha i}$ and H_{Ni+n} (n is a positive integer), ■ indicates either an NOE between H_α's or an NOE between $H_{\alpha i}$ and H_{Ni+3}, ● indicates an NOE between H_{Ni} and $H_{\alpha i+n}$ (n is a positive integer), and ⊙ indicates an NOE between H_α and side-chain protons. Below the diagonal: ■ indicates either that NOE's between H_{Ni} and H_{Ni+n} and $H_{\beta i}$ and H_{Ni+n} both were observed or that an NOE between $H_{\alpha i}$ and $H_{\beta i+3}$ was observed, a vertically hatched ▣ indicates an NOE between H_{Ni} and H_{Ni+n} only, a horizontally hatched □ indicates an NOE between $H_{\beta i}$ and H_{Ni+n} only, a cross-hatched □ indicates a sequential NOE between H_N and side-chain protons, a stippled ▣ indicates an NOE between $H_{\beta i}$ and $H_{\alpha i+n}$, and ○ indicates NOE's either between H_N and side-chain protons or between side-chain protons. Connectivities appearing as an x are long-range NOEs which were found late in the analysis and were used only in the final conformer refinements (vide infra). The two enriched boxes correspond to the 19/33 connectivities that were used in the initial refinements and were later found to be suspect.

tertiary contacts, are evident in a diagonal plot of the NOE data (Figure 7).

For an initial elucidation of the folding of hevein, we selected 180 conformationally significantly NOE connectivities (58 intraresidue, 62 $i/i+1$, and 60 medium to long range) which were converted into tight distance constraints (as per Table I). XPLOR-2.1 (Polygen) (Brunger, 1990) was used to generate an ensemble of NMR structures by the following sequence: initial noncyclic structure—(SA, patch disulfides) → initial cyclic structure—$(SA)_1$ → stage 2 SA structure—$(SA)_{1-3}$ → SA structure. The simulated annealing structures were relaxed by an unconstrained steepest descent CHARMm minimization → FR (fully relaxed) structures. Twenty-six runs were initiated (from widely different noncyclic starting models), and no structures were rejected on the basis of either constraint violations or high CHARMm energies. The stage 2 SA structures displayed from 20 to 36 violations in excess of 0.2 Å with sums of absolute violations ranging from 11.6 to 26.4 Å, corresponding to an rmswv measure of 0.18–0.57 Å. The structures with the poorest fits were produced from the totally unrealistic full-length helix model. At stage 2, six of the better fitting SA structures were surveyed for interresidue distances less than 3.4 Å, and the NOESY spectra were carefully reexamined—unambiguously absent cross-peaks were then converted to low-bound-only constraints (LBOs), $d_{ij} \geq 3.6$ Å. In like manner all instances of unambiguously unobserved backbone $i/i+2$, $i/i+3$, and $i/i+4$ peaks were also converted to LBOs. A total of 310 constraints were added in this manner. (*As it turned out, these had little influence on the subsequent refinement.*) After the SA procedure was repeated one additional time, all of the pair interactions representing unobserved NOE's were eliminated, and a set of converged NMR structures was obtained for 21 of the 26 runs with the highest sum of violations corresponding to 7.00 Å; the rmswv (see eq 1) over the original 180 constraints was 0.13 Å. The SA procedure was repeated (one or two times) for those structures that still displayed rmswv >0.15 Å. The final set of 26 SA structures displayed an rmswv of 0.10 ● 0.02 Å.

The SA structures were relaxed into the nearest conformational minimum by steepest descent CHARMm minimization (without NOE constraints). The rmswv measure

[4] Due to our use of DCs that are considerably tighter than those used in most XPLOR-based protein structure refinements, this violation measure needs to be put in perspective and compared to the commonly reported rms violation measure (from program XPLOR) which would be observed with both tight and loose DCs. For a typical first-generation structure (rmswv = 0.13 Å) generated by SA against tight bounds, XPLOR reports an rms violation of 0.08 Å (versus our tight constraints) or ≤0.04 Å when broader, "conservative", constraint ranges are used for the XPLOR statistics. Likewise, for a fully relaxed second generation structure, the rmswv measure is 0.11 and 0.03 Å, respectively, against the tight and loose constraints; XPLOR reports "rms diff" values of 0.08 and 0.04 Å, respectively, against the same constraints.

FIGURE 8: Views of structural features of the first-generation hevein structure ensemble. The disordered C-terminus appears in the lower left-hand corner of each panel. Panels A and B show the entire ensemble of 26 NMR structures obtained using tight constraints and reveal that the overall fold is defined (pairwise 3 → 41 backbone rmsd = 1.44 Å); in panel B, the better defined core (pairwise 15 → 34 backbone rmsd = 0.65 Å) is shown with the more "disordered" termini in gray tone. Panel C shows, in the same perspective employed for panel B, the consensus (17 of 30 conformers) from the ensemble obtained using a "conservative" conversion of the NOE intensities to broader constraints (pairwise 15 → 34 backbone rmsd = 0.80 Å).

Table IV: Changes in Constraint Violations during the Course of the First-Generation Structure Refinement

		penalty functions (180 constraints)		
refinement stage	E_{NOE}[a]	rmswv[b]	frac >0.2 Å	Σ|viol|[b]
initial structures				
all (av, SE)	>10⁶	11.7 ± 10.5	0.41 ± 0.12	767
helical only		11.6	0.50	727
homology		1.74	0.29	87
simulated annealing				
second stage (⟨worst 15%⟩)	636	0.55	0.20	26.2
second stage (⟨best 15%⟩)	458	0.20	0.12	11.6
final stage				
best	21	0.06	0.01	2.65
mean	49 ± 27	0.10 ± 0.02	0.04	5.48 ± 1.10
worst	132	0.13	0.08	7.55
FR (no NOE constraints)				
best	148	0.18	0.09	7.99
mean	425 ± 141	0.27 ± 0.04	0.14	14.18 ± 2.77
worst	776	0.37	0.16	17.15

	82 core nonintraresidue constraints[c]		
refinement stage	rmswv[b]	frac >0.2 Å	Σ|viol|[b]
SA structures			
best	0.04	0.00	0.78
mean	0.09 ± 0.03	0.03	2.28 ± 0.77
FR structures			
best	0.13	0.07	3.01
mean	0.27 ± 0.07	0.17	7.38 ± 1.92

[a] Units of kcal/mol. [b] Violation measures are given in angstroms; rmswv is defined by eq 1. [c] This includes *all* nonintraresidue constraints involving at least one proton that is within the core region, defined as residues 14–35.

increased to 0.27 ± 0.04 Å for the fully relaxed (FR) structures.[4] The 310 added LBO constraints contributed a maximum sum of violation of only 0.19 Å for the FR structures; thus the summary of violations (Table IV) includes only the structurally significant constraints (*in the lower portion of the table only 82 interresidue constraints involving at least one spin site from within the structure-conserved core region, residues 14–35, are included in the penalty functions*). The CHARMm energies for the FR structures ranged from −1046 to −859 kcal/mol, equal to −24.3 to −20.0 kcal/(residue·mol), consistent with values obtained for unstrained structures using this force field (Andersen et al., 1992). The $E_{L\text{-}J}$ term obtained, −4.2 kcal/(mol·residue), is consistent with the values observed in NMR structures of larger, rigid protein systems, for example, interleukin-1β, −3.7 kcal/(mol·residue) (Clore et al., 1991). While 20 of the 26 FR structures could be viewed as superior with regard to both constraint violation measures and CHARMm energies, we chose not to eliminate any potential conformers in our presentation of the initial conclusions concerning the solution-state fold of hevein.

Features in the First-Generation Structures of Hevein. Views of the consensus features in our ensemble of FR structures appear in Figure 8. Panel A reveals that all structures represent the same overall folding pattern with a backbone atom rmsd (excluding the extreme residue at each terminus) of 1.44 Å. Panel B shows that the degree of structural convergence is greater in a core region from residues 15 through 34 (backbone rmsd = 0.65 Å). By a stricter definition of convergence the core is actually limited to residues 16–32. A short helical segment (residues 28–32) is well-defined in essentially all of the structures. As will be shown, this depiction obscures additional elements of conserved structure within the noncore segments. The structural convergence obtained at this stage is not an artifact of the use of tight bounds which might not compensate for the effects of segmental motion. Panel C shows that the same fold is obtained when all LBOs are eliminated and the low bounds are set at the values used when NOE intensities are converted to distances in the conservative fashion.

One of the most unusual features of the WGA domains is the occurrence of two adjacent non-glycine residues with an α_L conformation: $\phi_{14} = +57 \pm 7°$, $\psi_{14} = +31 \pm 10°$, $\phi_{15} = +64 \pm 7°$, $\psi_{15} = +22 \pm 12°$ (for eight determinations, two protomers of each of four domains) (Wright, 1987). These correspond to an Asn-Asn pair in all WGA domains and in hevein. In our hevein structures, the Asn-Asn unit is found predominantly in the form that is analogous to the WGA structures; the values for the 11 best fitting structures (rmsd = 0.36 Å from 13 → 16) are $\phi_{14} = +62 \pm 10°$, $\psi_{14} = +10 \pm 16°$, $\phi_{15} = +66 \pm 13°$, $\psi_{15} = +18 \pm 20°$. However, another conformational cluster (seven structures with an rmsd of 0.43 Å over the same range of residues) was also obtained with an α_R conformation: $\phi_{14} = -44 \pm 13°$, $\psi_{14} = -23 \pm 17°$, $\phi_{15} = -54 \pm 19°$, $\psi_{15} = -59 \pm 10°$. This may be a locus of conformational heterogeneity for the solution-state of hevein.

The first-generation ensemble of NMR structures generated for hevein was obtained without using any hydrogen-bonding information. The CHARMm minimization, used to relax the structures, also excluded the $E_{H\text{-bond}}$ term. Thus a comparison of the structures and the NH exchange data should provide a test of the NOE-derived structures. As shown in Figure 6, at least nine of the backbone NHs (residued 16–19, 22, 23, 25, and 30–31) display extremely slow exchange ($t_{1/2}$ > 12 h at pH 3). An additional 11 residues display $t_{1/2} > 0.8$ h, with the values for residues 3, 4, 24, and 32 approaching those observed for persistently H-bonded structures. Of the 13 NHs thus implicated as highly sequestered, nine (3, 4, 18, 19, 22, 25, 30–32) are listed in Wright's (1987) tabulation of homologous main-chain hydrogen bonds in the four WGA domains. The backbone NHs in residued 16, 18, 23, 24, 27, and 39 were water-inaccessible, as judged by a Connolly surface generated with a 1.4-Å probe (using INSIGHT).

Further Refinement, Second-Generation Structures. The solution-state structure of hevein had been defined to surprisingly high precision in the first-generation structures using only a semiquantitative set of nonstereospecific NOE distance constraints. The solution that emerged presented a well-defined fold, of which the dominant feature is a buckled antiparallel sheet, characterized by a close 18α/24α contact surrounded by the expected persistent H-bonds (17 → 25/25 → 17 and 19 → 23). Buckling at one end of the sheet could be seen by the appearance of a turn-associated 22 → 19 H-bond, which is also present in the WGA domain structures. The other end of the β-sheet structure was indicated by the much reduced intensity of the 16α/26α NOE, which was barely detectable in the NOESY spectra. This through-space interaction was, however, confirmed in a ROESY spectrum and displayed significantly greater intensity in the rotating frame experiment; its reduced intensity in the NOESY may be attributable to motional disorder—ROESY cross-peak intensities are less sensitive to changes in the effective correlation time. The only other section of regular secondary structure was a short α-helix from Asp^{28} to Ser^{32}. By the usual criteria, this represented a more than acceptable degree of both consistency with experimental data and convergence for an NMR structure at this level of refinement. As it turns out, the structural conclusions are flawed even though they display excellent convergence in the core region.

A further round of refinement was undertaken on the basis of three developments: (1) further analysis of the NOESY data yielded in excess of 40 new long-range constraints, some of which were significantly violated by our FR structures;[5] (2) a moderate-resolution X-ray structure of hevein became available for comparison (Rodriguez-Romero et al., 1991)[6] but the X-ray structure did not agree with the conclusions reached by the NMR study, and (3) a set of potentially key constraints between residues 19 and 33 were shown to be either invalid or suspect.[7] Our aims in this additional refinement stage were to ascertain the effect of the now suspect constraints on the refinement process, to incorporate the new constraints, to establish whether the solution- and solid-state structures of hevein are distinguishably different from each other and from domain C of WGA, and to define individual conformers (Andersen et al., 1992; Lai et al., 1993) rather than an "average structure" with regions of motional disorder.

The best 20 structures in our FR ensemble represent only five distinct conformers based on the criteria outlined under Experimental and Computational Methods. The better fitting representatives of each (a total of nine structures) were refined through three cycles of SA against the new constraint table including, as regularizing distance constraints (only during the first two cycles of SA), the previously recognized H-bonds that characterize each distinct conformer cluster. The X-ray structure[2] was also used as starting structures for XPLOR SA refinement based on the NOE constraints. A total of 43 structures were generated. The additional long-range constraints that became available late in the assignment process eliminated a number of conformers from consideration, and ten of the structures were eliminated by their high CHARMm energies. The remaining conformer structures are collected in Figure 9 together with depictions of the core structure from the previous round of refinement and regions from the X-ray structures of WGA and hevein. (*Similar results were obtained using expanded low bounds for this constraint set.*) The fit measure shown for each structure, or ensemble, is the average violation of the tight constraints *after relaxation.* The rmswv values ranged from 0.10 to 0.15 Å and had increased by less than 10% during the unconstrained relaxation. The fit measures were taken over 206 constraints [including, within the core, 37 sequential ($i/i+1$) and 82 $i/i+n$ ($n > 2$)] and 44 "key" LBOs. Conformer models that were rejected displayed average violations greater than 0.06 Å and rmswv values greater than 0.18 Å.[4]

Panel A of Figure 9 shows the most extreme conformers that were judged to be consistent with the NMR data. The largest ensemble of structures (16/23, $E_T = -1366 \pm 47$ kcal/mol) is represented in panel A by five structures shown in green and is designated as the "hevein α_L-NMR conformer" (HαLN). The next most frequently obtained conformer, HαRN (3/23, $E_T = -1329 \pm 58$), which is shown in yellow, is the 15-α_R conformer. Given that the HαLN species typically violate less than 10/250 constraints by in excess of 0.3 Å, and those violations all involve wild-carded δ/ε-aryl protons or

[5] A total of 60 new constraints were added as they became available from the continuing NOESY analysis. These replaced 30 conformation-insensitive intraresidue constraints in the original set. As this increase in constraint density occurred, structural convergence decreased in our ensemble of structures and a number of large violations (>0.8 Å) began to appear.

[6] The coordinates and conclusions reached in the parallel X-ray and NMR investigations were intentionally not shared in order to maintain each as an independent structure determination.

[7] The basis for this confusion in the assignment is the subject of a supplementary material figure.

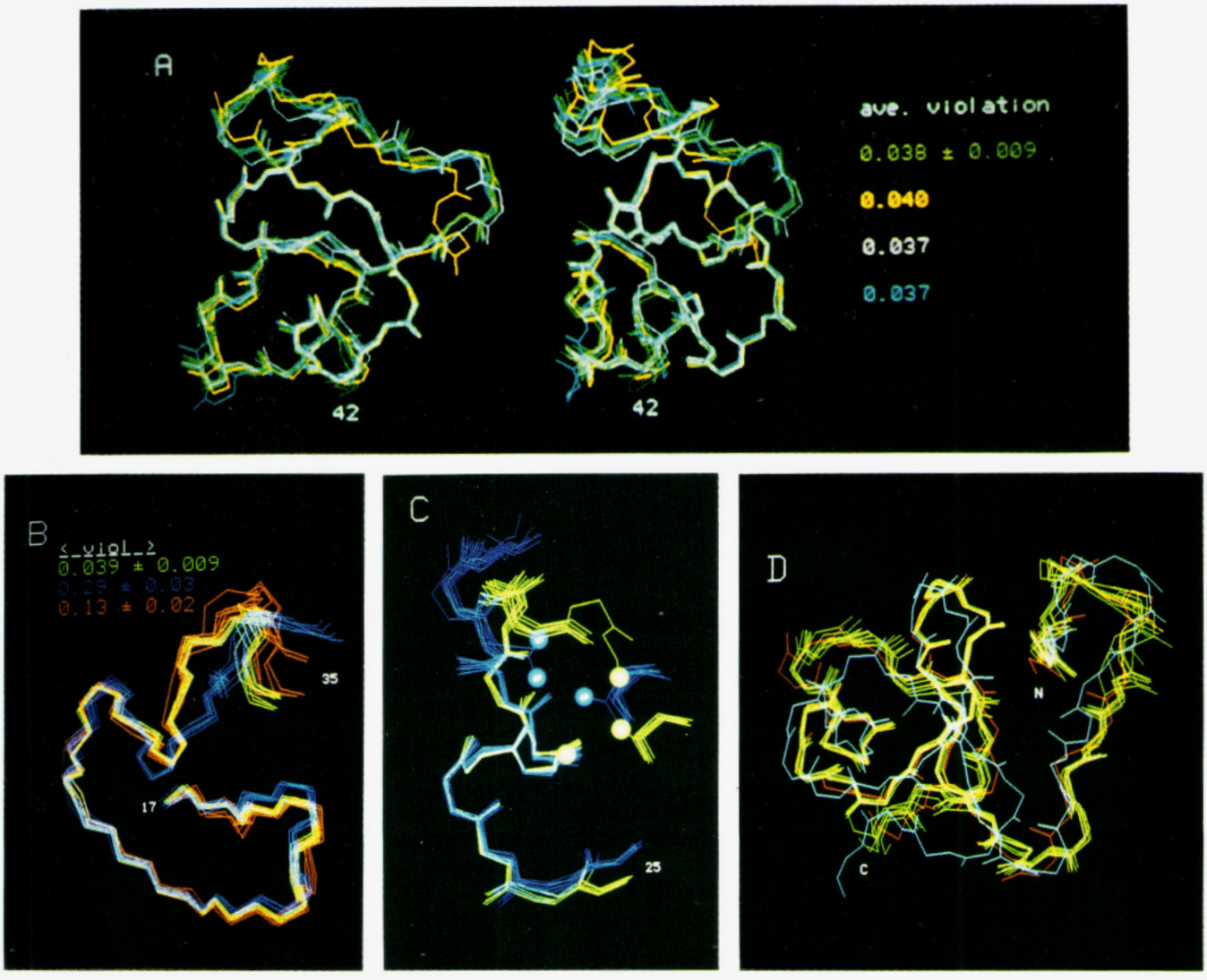

FIGURE 9: Second-generation conformer models for hevein in the solution state: comparisons with WGA-like structures, the FR structures from the first-stage refinement (from Figure 8) which included two unsubstantiated long-range distance constraints and had a lesser density of long-range connectivities, and the reported (Rodriguez-Romero et al., 1991) low-resolution solid-state structure. In panels A, B, and D the structures are overlaid to a least-squares fit over the backbone atoms of residues 16–32. The average violations shown (panels A and B) for each structure, or ensemble, are taken over 250 constraints, which included 37 sequential ($i/i+1$) constraints within the core region and 82 $i/i+n$ ($n \geq 2$) constraints involving the core region. In the case of panel B, two of the structures had been refined with an incomplete set of constraints lacking 51 (or 48) of the constraints that were located in the later NOESY analyses; these constraints were, however, used in evaluating the average violation measure. Panel A displays residues 2–42 of four second-generation conformers for the solution state of hevein that survived our fit and energy criteria; the right-hand view is a 38° rotation of the left-hand view about the vertical axis. The major conformer (HαLN, green structures) is represented by five structures (backbone rmsd = 0.42 ± 0.02; the values for the entire ensemble of this conformer were 0.58 ± 0.13). The backbone rmsd comparisons between the other conformers and HαLN for residues 2–42 (and 16–32) were as follows: for HαRN shown in yellow, 1.18 ± 0.13 (0.30 ± 0.03); for the pink conformer, 0.77 ± 0.13 (0.30 ± 0.08); for the cyan conformer, 0.98 ± 0.08 (0.27 ± 0.05). Panel B compares the major conformer from panel A (in green), that in figure 9 (blue), and those (red) obtained with a corrected abbreviated constraint set (see text). Residues 17–35 are displayed. The least-squares fitting was over all backbone atoms of residues 16–32; residue 16 and the carbonyl oxygens were deleted for clarity. Panel C shows the helical domain in the first- and second-generation ensembles. The disconnected segment is Ser[19]. In the first-regeneration set 19-Hα, 33-Hδ1, and 33-Hδ2 are shown as blue balls. In the second-generation set 19-, 30-, and 37-Hα's appear as green balls. Panel D shows residues 2–42 for the X-ray structure of hevein (cyan) which displayed a large average violation (1.12 Å) of the NOE-derived distance constraints, the HαLN conformer derived by NMR (green), and the best-fitting ($\langle|\text{viol}|\rangle$ = 0.08 Å) WGA-like structure (red) that could be generated. For clarity, only a single representative of the WGA-like set is included; the convergence within the group is excellent—2 → 32 backbone rmsd = 0.19 ± 0.04 Å.

methyl groups,[8] it becomes necessary to examine whether the minor conformers are required to rationalize the NOE data: is there any basis for viewing them as populated conformers rather than abberant solutions? A careful analysis of the differences in violations between HαLN and HαRN reveals a self-consistent set of ten long-range NOE distances that appear as negative violations (model distances shorter than the experimental constraint) in the α_R conformer but as positive violations in HαLN and a set of four related NOE distances with the opposite trend. This is precisely what would be expected (Brüschweiler et al., 1991) if the α_R conformer is a minor conformer in the solution-state equilibrium occurring under the experimental conditions examined. The relative reduced intensity for the 16α/26α cross-peak in NOESY spectra (versus the ROESY) may also reflect conformational changes at these loci. The two remaining conformer structures in panel A illustrate the largest differences that were observed at other loci (Gly[9], Pro[13], and His[35]); their inclusion in the

[8] We employ $\langle r^{-6} \rangle$ averaging to calculate model r_{ij} values, for comparison to the constraints, and in the XPLOR SA refinements for all categories of wild-carded proton sets. There is no evidence of hindered rotation for the ring of Tyr[30].

ensemble, however, does not provide a significant improvement in fit to any long-range NOE constraints. There is, to date, no experimental basis for viewing them as populated conformers of hevein in the solution state.

Panel B provides a basis for discussing the differences between the structures at this and the previous stage of refinement; the region from Cys^{17} to His^{35} as it appears in the final refinement (green) and the preceding one (blue) is illustrated. Excellent convergence was observed in each case; pairwise backbone (16–32) rmsd values for the ten best structures were 0.19 ± 0.03 Å for the final refinement and 0.40 ± 0.04 Å for the penultimate series. The structures are, however, quite distinct, displaying an average pairwise backbone (16–32) rmsd of 1.08 ± 0.06 Å between structures from the first- and second-generation ensembles. The difference can be shown to be largely due to the two "incorrect" constraints in the earlier refinement. The changes in constraints were $19\alpha/33\delta^* \rightarrow 19\alpha/30\alpha$ and $19\alpha/37\alpha$, $19\beta^*/33\delta^* \rightarrow 19\beta^*/30\alpha$. We carried out nine SA refinements in which only these three changes were made in the constraints; none of the additional 65 constraints found during subsequent NOESY data analysis were added. The resulting structures appear in red in panel B of Figure 9 and match the final set quite well. It is thus apparent that the extent of structural error that can be introduced by a few questionable long-range constraints is large and that structural convergence is not always a dependable criterion for the correctness of an ensemble of NOE-derived structures. The added long-range constraints did, however, serve to improve the convergence of the final ensemble of structures.

Panel C [which shows the residue 28–35 span and Ser^{19}, the latter being included to illustrate the specific locations of the altered constraints, $19\alpha/33\delta^*$ (blue balls) $\rightarrow$ $19\alpha/30\alpha$ and $19\alpha/37\alpha$ (green balls)] illustrates the helical domain of hevein. This span is now well-determined, displaying backbone torsion rmsds of less than ±14° throughout. With the exception of ψ_{32} (= +140 ± 7°), the entire span is α_R configured: $\langle\phi_{28-35}\rangle = -74 \pm 19°$, $\langle\psi_{28-35}\rangle = -24 \pm 21°$. This helical region, limited to Asn^{28} to Ser^{32} in the first-generation structures, can thus be described as an eight-residue helix with a single kink in it.[9] At several points the backbone torsion angles suggest an internally hydrated α-helix, $\phi \approx -95°$ and $\psi \approx +10°$ (Karle & Balaram, 1990).

Turning to comparisons (panel D of Figure 9) of the NMR structure (HαLN) to WGA and the existing X-ray structure, the criterion used to trim the SA conformers would have eliminated both as models for the solution-state structure of hevein. In order to test the possibility of hevein having a structure similar to WGA, we incorporated additional constraints so as to maintain a WGA-like fold while retaining the full set of NOE constraints. The resulting structures were indeed WGA-like: backbone (2–32) rmsd, 0.46 ± 0.02 Å, versus domain C of WGA, 0.96 ± 0.05 Å, versus HαLN—the rmsd between HαLN and WGA-C is 0.95 ± 0.04 Å. As a group these WGA-like structures displayed a mean violation of 0.09 ± 0.01 Å from the NOE-derived distances. The best fitting of these conformers is shown (red structure) in panel D. In excess of 60 XPLOR SA runs with torsion constraints that maintain a ≤0.50-Å backbone (2–32) rmsd with WGA were carried out; in every case the rmswv values for WGA-like structures exceeded 0.27 (versus 0.11–0.13 for all members of the HαLN ensemble). The solution-state structure of hevein is thus readily distinguishable from the solid-state structure of WGA over any homologous region—the core only or from residues 2–32, the full span over which a similar fold would be predicted. *However, the solution-state structures derived in the present work resemble the solid-state structure of WGA more than that recently reported for hevein itself.*

The crystal structure of hevein displays, over the $Leu^{16} \rightarrow Ser^{32}$ core region, a backbone rmsd of 1.25 ± 0.05 Å versus HαLN. This 2.8-Å solid-state structure does not have a well-formed antiparallel sheet in the core region, displaying a Σ-(viol)^2 of 31.7 Å^2 over 19 cross-strand NOE distances (versus 0.32 ± 0.20 Å^2 for HαLN). Violations of NOE constraints include ten instances of X-ray model distances of >5.0 Å for which indisputable backbone-H/backbone-H interstrand NOE's are observed. Large deviations are also observed for a number of loop/loop and loop/core interactions. Over the 2–42 backbone the X-ray structure has a 2.83 ± 0.10 Å rmsd with the HαLN ensemble. The solution structure also provides better explanations for the unusual chemical shifts observed in hevein. As previously noted, the side-chain shifts of Q20 are remarkable, and the methylene protons of Gly^{25} are extremely shift divergent, with one of them appearing at an unusual upfield position, 2.08 ppm. Figure 10 provides a rationale for these observations and demonstrates that the X-ray structure does not rationalize the shifts as well. A rationale for a key inter-side-chain NOE in this region is also absent in the solid-state structure. The differences between the solution- and solid-state structure models for hevein will be explored in collaboration with Dr. Soriano.

CONCLUSIONS

The present study has provided another demonstration of the importance of obtaining the highest possible density of long-range constraints in attempts to define protein tertiary structure by NMR and a vivid illustration of the need for absolute certainty in assignments of any long-range constraints employed. In the case of hevein, two incorrect assignments (out of a total of 60 long-range constraints for a 43-residue system) did not prevent convergence; *rather, they improved the convergence to an incorrect structure.* As the additional 41 long-range constraints, extracted from further NOESY analysis, were added, an inconsistency was apparent—the ensemble of structures from SA refinements began to show larger NOE constraint violations and a lessened degree of convergence as measured by the backbone rmsd over the core region of the structure. Structural convergence within a set of NMR structures cannot be taken as an indication of the accuracy of the structure refinement nor as validation of the procedure used. When two long-range NOE attributions were corrected, only two conformers were required to rationalize all of the NOE's observed to within experimental error (±≤0.25 at $d_{ij} \leq 3.0$ Å, ±0.5 at $d_{ij} = 3.8$ Å). Neither of these solution conformers corresponds to either the recently reported X-ray structure or the WGA domain structure, but they do resemble the latter. Thus the present study is one of the rare instances in which NMR has provided a notably different structure for a small protein than that which was predicted on the basis of homology modeling or derived by X-ray crystallography.

It should also be noted that the degree of structural convergence observed within the entire ensemble of solution structures (<1.2 Å pairwise backbone rmsd from residues 2–42, <0.4 Å over the residue 16–32 core region) is better

[9] All attempts (using constrained dynamics followed by relaxation in the CHARMm force field) to create models lacking this kink produced structures that had not only significantly increased violations but also higher CHARMm energies (by >20 kcal/mol) in the absence of NOE constraints.

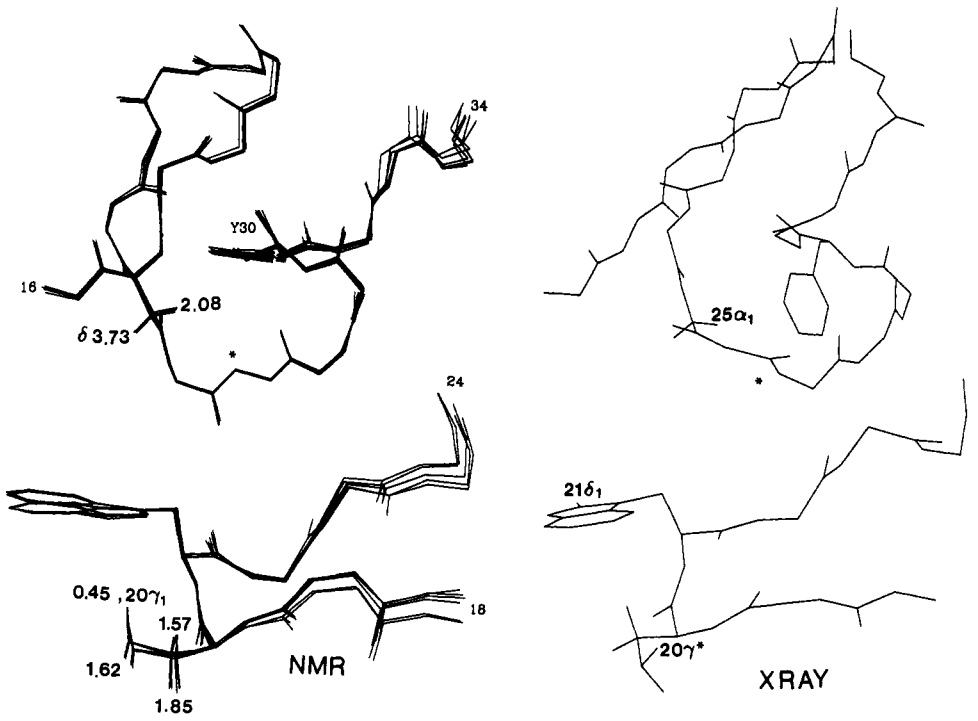

FIGURE 10: To the left, views of the shielding of the Gly[25] $H\alpha_1$ by the aryl ring diamagnetic effect of Tyr[30] and the differential shielding of the Gln^{2-} side-chain protons by the indole ring of Trp[21] as seen in the major conformer of the second-generation NMR structure. The unusual chemical shifts placed by the protons on the structural depictions are those observed at pH 6.6. The location of 27HN (which appears far upfield, at 7.05 ppm) is indicated by an asterisk. The upfield 20γ proton displays a moderately weak NOE to $21\delta_1$ and ϵ_1 which correlates with the 3.4 ± 0.1 Å $20\gamma_1/21\delta_1$ distance in the model. To the right, the corresponding segments from the solid-state structure are shown, in one case rotated to yield a better view of the Tyr side-chain orientation. $25\alpha_1$ is not predicted to be as highly shielded and 27HN would not be shielded at all in this model. Both the Gln^{20} and Trp^{21} side chains appear in different conformations which explain neither the shifts nor the inter-side-chain NOE ($20\gamma^*/21\delta_1$ distance = 7.4 Å).

than that expected (Clore & Gronenborn, 1991, 1992). Within the core region the constraint density (counting all interresidue constraints twice) varies from 8 to 23 per residue; outside of the core the count varies from 2 to 13 per residue. At a total of 244 constraints for 43 residues, in the absence of torsion constraints and *pro-R/S* assignments, this represents a first-stage NMR structure based on the criteria of Clore and Gronenborn (1991), and the expected structural accuracy is ≥1.5 Å backbone rmsd. The higher accuracy obtained in this study is the result of tighter distance bounds, in particular, abandoning the universal 2-Å low bound. The risks and rewards associated with this strategy will be fully documented in a subsequent paper.

CD spectroscopy provides some insights into one of the distinctions between the present NMR study, the X-ray structure, and the WGA analogy. The far-UV CD traces for hevein throughout the pH range from 2.4 to 8.0 are shown in Figure 11 and bear a considerable resemblance to those of WGA. The previous CD comparisons of hevein and WGA (Rodríguez-Romero et al., 1989), which did not penetrate to 180 nm, establish that the CD spectral features are absent upon cleavage of the disulfide linkages. The difference CD, between the most acidic conditions and the others, corresponds to the expectation spectrum for a very short helix (Yang et al., 1986; Harris et al., 1992). If the five-residue span (28–32) with $i/i+3$ connectivities is the sole contributor to this ΔCD feature, the decrease in θ_{222} at pH 2.4 corresponds to a 41% increase[10] in helix population at the most acidic conditions examined, based on the formula for θ_{222} given by Scholtz et al. (1991). The NMR data also provide evidence for a pH-induced conformational change in or near the helical segment. When the amide NH chemical shifts at pH 2.4 and 6.6 are compared, we note an insignificant downfield shift (+0.03 ± 0.04 ppm) upon acidification for all residues outside of the 28 → 35 region. Within this region much larger shifts are observed: Glu^{29}, –0.46 ppm; Tyr^{30}, –0.14 ppm; Ser^{32}, –0.39 ppm; Asp^{34}, +0.10 ppm; His^{35}, +0.21 ppm. At pH 6.6, the Asp^{28} HN is absent (presumably due to rapid cross-saturation), and two distinct NH chemical shift gradients are observed over this span: δ (residue number) 8.75 (29), 7.89 (30), 7.59 (31), *and* 8.51 (32); Pro, 8.28 (34) and 7.60 (35). At pH 2.4 the Asp^{28} HN appears at an extremely downfield position, and a single NH chemical shift trend, consistent with a helix dipole effect (Wishart et al., 1991), is observed from 28 → 31—9.43 (28), 8.29 (29), 7.74 (30), 7.63 (31)—and 8.11 ppm (at residue 32) breaks the trend. The CD data are consistent with the significant population of a short helix at all pH conditions examined. At pH <4 the helix population or extent increases. While $i/i+3$ connectivities over residues 28 → 32 are observed at both pH 2.4 and pH 6.6, a detailed quantitative analysis gives better agreement for the pH 2.4 data. With regard to the absence of this feature in the solid-state structure, it should be noted that the crystals used for X-ray structure were grown at pH 8.06 (Rodriguez-Romero et al., 1991).

The solution structure of hevein derived in the present study is remarkable in several respects. Unlike most small globular proteins which have been examined, or the scorpion toxins [as exemplified by charybdotoxin (Bontems et al., 1991)], hydrophobic cluster formation does not appear to be the major fold determinant. The only residues that are judged to be entirely solvent-inaccessible are $Cys^{3,18}$ and the residues

[10] Given the ϕ/ψ values observed in the full ensemble of NMR structures (vide supra), this estimate of population change may be high; the α_R conformations at 33–35 and at 14–15 (in a minor conformer), as well as type I β turns (Perczel & Fasman, 1992), may also contribute CD signals resembling a short helix.

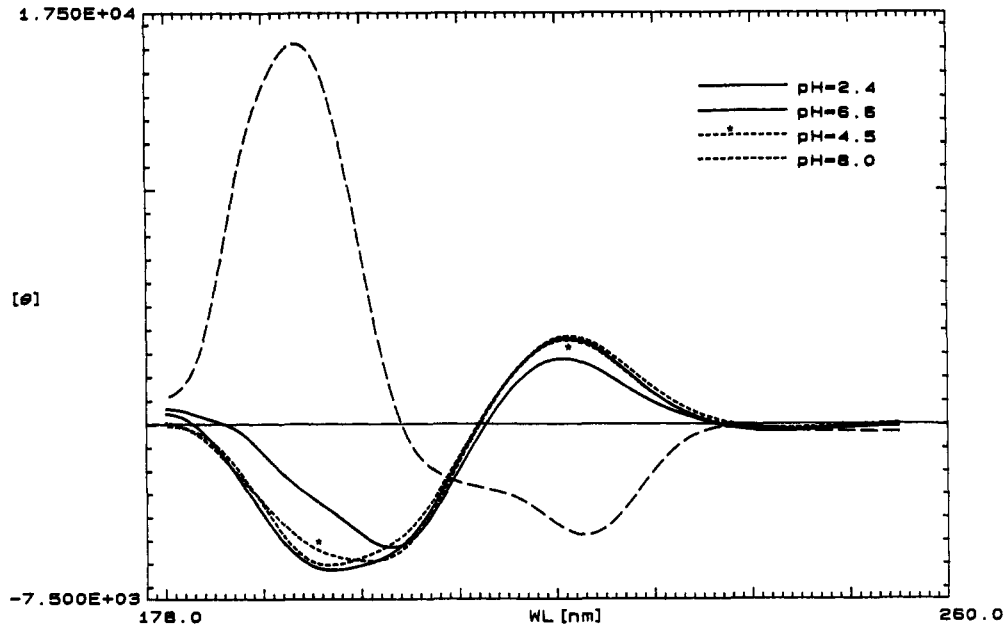

FIGURE 11: CD spectra (residue molar ellipticity versus wavelength) of hevein at pH 2.4, 4.5, 6.6, and 8.0 (at 50 μM protein and 10 mM phosphate throughout). The pH 2.4 and 6.6 traces are solid lines; those at pH 4.5 and 8.0 are broken lines and are labeled as to the pH. The heavy dashed-line trace corresponds to the ΔCD (pH = 2.4 – pH = 6.6), with the residue molar ellipticity scale based on the assumption that the entire change occurs within an eight-residue span.

adjacent to this disulfide linkage (Gly^4 and Ser^{19}) and Gly^{22}. None of the hydrophobic side chains (Leu, Pro, Trp, and Tyr) were fully internalized. One of the methyls of each leucine is partially shielded, the other being fully exposed. Both Trp and Pro residues are solvent exposed. All of the polar side chains (Asp, Glu, Lys, and Arg) are fully exposed.

Even though there is no covalent connection between the core and the C-terminus, both WGA and hevein (in the solution and solid state) display a rather similar disposition of this flexible loop relative to the core. This is remarkable given that the connecting region between the core and the conformationally independent C-terminal disulfide loop is the point of greatest sequence divergence in the structures—*n* in the following comparison is 3 for hevein and 1 for WGA.

$$\text{X-}\underline{\text{C}}\text{G-X}_3\text{-G(G/N)-X}_2\text{-}\underline{\text{C}}\text{PNN} \qquad \text{S-S}$$
$$\text{(L/H/Y)CCS-X}_2\text{-GXC-X}_4\text{-(Y/F)C(G/S)-X}_n\text{-(G/N)CQ-}$$

The sequence comparison reveals Cys residues flanked by Gly or Asn units which impart unusual flexibility and allow for local α_L conformations: a survey of all conformer structures which we generated reveals that at least five of these seven residues are in this rarer local conformation; the conformer with all Asn residues in the α_L conformation[11] is consistent with the NOE's observed. We propose that these are required, in what becomes the core, in order to *predispose the system to a fold that is subsequently fixed by the closure of the disulfides*. The flexible segment between Cys^{31} and the small disulfide loop in the C-terminus would, presumably, allow for a variety of spatial dispositions of the C-loop relative to the core. An examination of the WGA structure and HαLN reveals the following common features: the 37-Cβ-S-S-Cβ-41 (hevein numbering) unit is fully internalized, shielded from the solvent surface by the 38–40 backbone, Cys^{31}, and the 17–19 backbone. Two H-bonds stitch the loop to the core ($NH_{(18)}\cdots O_{(39)}$ and $NH_{(38)}\cdots O_{(18)}$), the latter displaying excellent geometry and a 2.87-Å N–O distance. In both WGA and hevein, the conserved $Gln^{(36/38)}$ is in close contact with both the core and the same portion of the N-terminal loop, $N\text{–}H_{(5)}\cdots OE1$. In WGA the *trans*-amide NH is bonded to $OC_{(19)}$. In the NMR structures of hevein, the H-bond acceptor is either the 19 or 20 carbonyl. Structures refined to optimize these interactions while retaining a fit to the NOE constraints indicate that the $HE22_{(38)}\cdots OC_{(19)}$ bond, as also seen in WGA, is the better model with regard to both constraint violations and CHARMm energy. Gln^{38} displays very unusual chemical shifts. For all of the other side-chain $CONH_2$ groups the NHs cis and trans to the carbonyl oxygen are found at δ 6.83 ± 0.13 and 7.40 ± 0.25, respectively; for Gln^{38} they appear at 8.30 and 8.20 ppm. We propose that the interactions described provide the driving force for the common fold preference for the otherwise unconstrained C-terminus.

The NOE-derived structures obtained in this study should prove useful for defining the molecular basis of the agglutinizing reactions recently found for hevein (Rodriguez-Romero et al., 1991). These activities, and possible allergic responses, are almost certainly associated with the exposed loops, the structures of which could not be defined if the broader bounds typical of protein structure elucidation (for example, 1.9–2.7, 2.0–3.5, 2.4–4.5, and 2.7–5.5 Å) were employed (Andersen and Cao, studies to be published separately).

The agglutinin–toxin folding motif represents an excellent system for studies of the significance of local secondary structure preferences as folding determinants (Dyson et al., 1988a,b; 1990; Wright et al., 1988). As stereospecific assignments become available and both sequential and long-range constraints are defined more tightly by the use of the DISCON algorithm (Andersen et al., 1992; Lai et al., 1993), the persistently structured regions and the nature of any remaining conformational averaging should be defined to greater precision. The resulting dynamic structure hypothesis, and the resonance assignments from the present study, will serve as the basis for studies of the refolding of hevein. Even

[11] A quick inspection of a short mixing time NOESY provides a nearly foolproof initial assignment of α_L sites. They are diagnosed by an intraresidue $d_{N\alpha}$ peak that is as large as the larger sequential $d_{\alpha N}$ peaks in the spectrum and the appearance of at least one of the corresponding $d_{NN(\pm 1)}$ cross-peaks.

at the present stage of refinement, the derived solution structure for hevein differs significantly from that seen in a moderate-resolution solid-state structure (Rodriguez-Romero et al., 1991). Further study, including a higher resolution X-ray data set (Soriano & Rodriguez, research in progress), will be required in order to ascertain which of these differences represent actual changes in conformational preference and fold in the states examined.

ACKNOWLEDGMENT

Chin-pan Chen provided the XPLOR procedures which were modified for use on this problem. We thank Scott M. Harris for help in recording CD spectra. Professor M. Soriano-Garcia (UNAM, México) provided coordinates from the 2.8-Å X-ray data for hevein.

SUPPLEMENTARY MATERIAL AVAILABLE

Six figures showing d_{NN} connectivities at pH 6.6, aryl-CH connectivities in D_2O at pH 6.6, representative annotated HN/Hα and Hβ segments of the NOESY at 310 K, illustrations of the assignment of both Pro residues and protons coincident with the Pro^{33} δ signals and the appearance of diagnostic intra-$d_{N\alpha}$ peaks for α_L residues, and a comparison of NOESY spectra recorded at three temperatures and an annotated listing of the complete set of distance constraints used in the final SA runs (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Andersen, N. H., Eaton, H. L., & Nguyen, K. T. (1987) *Magn. Reson. Chem. 25*, 1025–1034.

Andersen, N. H., Eaton, H. L., & Lai, X. (1989) *Magn. Reson. Chem. 27*, 515–528.

Andersen, N. H., Lai, X., & Marschner, T. (1991) NOESYSIM/DISCON Documentation, University of Washington, Seattle, WA.

Andersen, N. H., Chen, C., Marschner, T. M., Krystek, S. R., Jr., & Bassolino, D. A. (1992) *Biochemistry 31*, 1280–1295.

Archer, B. L. (1960) *Biochem. J. 75*, 236–240.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bax, A., & Drobny, G. (1985) *J. Magn. Reson. 61*, 306–320.

Bean, J. W., Kopple, K. D., & Peishoff, C. E. (1992) *J. Am. Chem. Soc. 114*, 3328–3334.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Bontems, F., Roumestand, C., Gilquin, B., Menez, A., & Toma, F. (1991) *Science 254*, 1521–1523.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Saminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brünger, A. T. (1990) *XPLOR Version 2.1 Manual*, Yale University, New Haven, CT.

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

Clore, G. M., & Gronenborn, A. M. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 29–63.

Clore, G. M., & Gronenborn, A. M. (1992) *33rd Experimental NMR Conference (March 31), Book of Abstracts*, pp 44, Pacific Grove, CA.

Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991) *Biochemistry 30*, 2315–2323.

Drenth, J., Low, B. W., Richardson, J. S., & Wright, C. S. (1980) *J. Biol. Chem. 255*, 2652–2655.

Driscoll, P. C., Clore, G. M., Beress, L., & Gronenborn, A. M. (1989) *Biochemistry 28*, 2178–2187.

Drobny, G., Pines, A., Sinton, S., Weitekamp, D. P., & Wemmer, D. (1979) *Faraday Symp. Chem. Soc. No. 13*, 49–55.

Dyson, H. J., Rance, R., Houghten, R. A., Wright, P. E., & Lerner, P. A. (1988a) *J. Mol. Biol. 201*, 161–200.

Dyson, H. J., Rance, R., Houghten, R. A., Wright, P. E., & Lerner, P. A. (1988b) *J. Mol. Biol. 201*, 201–217.

Dyson, H. J., Satterthwait, A. C., Lerner, R. A., & Wright, P. E. (1990) *Biochemistry 29*, 7828–7837.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Harris, S. M., Cao, B., Lee, Y. G., & Andersen, N. H. (1992) *Biopolymers* (submitted for publication).

Johnson, W. C., Jr. (1990) *Proteins: Struct., Funct., Genet. 7*, 205–214.

Karle, I. L., & Balaram, P. (1990) *Biochemistry 29*, 6747–6756.

Kessler, H., Griesinger, C., Kerssebaum, R., Wagner, K., & Ernst, R. R. (1987) *J. Am. Chem. Soc. 109*, 607–609.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241–7257.

Lai, X., Chen, C., & Andersen, N. H. (1993) *J. Magn. Reson.* (in press).

Low, B. W., Preston, H. S., Sato, A., Rosen, L. S., Searl, J. E., Rudko, A. D., & Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2991–2994.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Metzler, W. J., Valentine, K., Roebber, M., Friedrichs, M. S., & Mueller, L. (1992) *Biochemistry 31*, 5117–5127.

Montelione, G. T., Wüthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., & Scheraga, H. A. (1992) *Biochemistry 31*, 236–249.

Négrerie, M., Grof, P., Bouet, F., Ménez, A., & Aslanian, D. (1990) *Biochemistry 29*, 8258–8265.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.

Perczel, A., & Fasman, G. D. (1992) *Protein Sci. 1*, 378–395.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Rodríguez, A., Tablero, B., Barragán, B., Lara, P., Rangel, M., Arreguín, B., Possani, L., & Soriano-Garcia, M. (1986) *J. Cryst. Growth 76*, 710–714.

Rodríguez-Romero, A., Arreguín, B., & Hernández-Arana, A. (1989) *Biochim. Biophys. Acta 998*, 21–24.

Rodríguez-Romero, A., Ravichandran, K. G., & Soriano-Garcia, M. (1991) *FEBS Lett. 291*, 307–309.

Scholtz, J. M., Quian, H., York, E. J., Stewart, J. M., & Baldwin, R. L. (1991) *Biopolymers 31*, 1463–1470.

Wagner, G. (1983) *J. Magn. Reson. 55*, 151–156.

Walujono, L., Sholma, R. A., Beintema, J. J., Mariono, A., & Hahn, A. M. (1976) in *Atlas of Protein Sequences and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, p 308, National Biomedical Press, Washington, DC.

Warren, G. L., Beshah, K., Goodfriend, L., Petsko, G. A., & Neuringer, L. J. (1991) *J. Cell. Biochem., Suppl. 15G*, 90.

Wishart, D. S. Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.

Wright, C. S. (1987) *J. Mol. Biol. 194*, 501–529.

Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.

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

Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–269.

Zagorski, M. G. (1990a) *J. Magn. Reson. 86*, 400–405.

Zagorski, M. G. (1990b) *J. Magn. Reson. 89*, 608–614.