# Solution Structure of the Fifth Repeat of Factor H: A Second Example of the Complement Control Protein Module†

P. N. Barlow,*,‡ D. G. Norman,‡ A. Steinkasserer,§ T. J. Horne,‡ J. Pearce,‡ P. C. Driscoll,‡ R. B. Sim,§ and I. D. Campbell‡

*Department of Biochemistry and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.*

*Received October 18, 1991; Revised Manuscript Received January 8, 1992*

ABSTRACT: Modules which share the same consensus sequence are assumed to have common structural features, at the secondary and tertiary level. In order to test the extent of such similarities, it is necessary to examine the structures of several examples from each module family. Recently, the first three-dimensional structure of a complement control protein (CCP) module (the 16th repeat of human factor H, H16) was determined using a combination of two-dimensional NMR and simulated annealing [Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B., & Campbell, I. D. (1991) *J. Mol. Biol. 219*, 717-725]. Using the same techniques, the three-dimensional structure of a second CCP module (the 5th repeat of human factor H, H5) has now been determined. The primary sequence of H5 contains 17 residues which are identical and in equivalent position to those in H16. Thirteen of these 17 are part of the consensus sequence. The similarities between the secondary structure of H5 and that of H16 are extensive. This implies that the consensus sequence dictates a particular secondary structure. The tertiary structure of H5, a compact hydrophobic core wrapped in $\beta$-strand and sheet, bears much overall resemblance to that of H16. However, there is a deletion in the first strand of H5, and an insertion in a loop, resulting in slightly shorter overall length. This is associated with a rearrangement of residues within the hydrophobic core. The side chain of the highly conserved Tyr29, which occupies a central position within the core of H16, lies on the periphery of the core of H5.

Protein modules are autonomously folding polypeptide units of 40-100 amino acids in length, generally encoded by discrete exons. Each module retains its own global fold when its sequence is incorporated into that of a "mosaic" protein constructed of many modules. Modules may be divided into families on the basis of limited consensus sequences (Patthy, 1991). It seems plausible that the consensus sequence will confer extensive similarities in structure (Baron et al., 1990a), but more examples are required to test this hypothesis and to assess the validity of homology modeling (Sali et al., 1990) for module families. We now describe the three-dimensional solution structure of a second example of the complement control protein (CCP)[1] module (also known as the short consensus repeat or SCR) and address the question, To what extent do modules which share a consensus primary structure (i.e., lie within the same family) share a common secondary and tertiary structure?

The CCP module is about 60 amino acids in length and has characteristic highly conserved residues including four almost invariant cysteines which form intradomain disulfide bridges, Cys1-Cys3 and Cys2-Cys4 (Figure 1). Between 2 and 30 CCP modules are present in each of 12 complement proteins (Day et al., 1989; Sim & Perkins, 1989). For example, complement factor H is an abundant plasma glycoprotein (155 kDa) containing about 12% carbohydrate (Ripoche et al., 1988). Its primary sequence consists of 20 contiguous CCP modules. A group of functionally related proteins, including complement receptors 1 and 2, C4b-binding protein (C4bp), membrane cofactor protein, and decay accelerating factor, are also composed almost entirely of CCP modules arranged in contiguous fashion. All of these proteins interact specifically with the complement proteins C3b and/or C4b. The specific protein-protein interactions involved are restricted to a small proportion of the total number of CCP modules (Klickstein et al., 1988; Krych et al., 1991). Hence the module appears to be a molecular framework which may be tailored to perform specific recognition/binding functions in some cases or to play a purely structural role in others. CCP modules are being identified in a growing number of noncomplement extracellular proteins (Day et al., 1989), including the blood clotting factor XIII B subunit, $\beta_2$-glycoprotein I, interleukin-2 receptor, proteoglycan core protein, thyroid peroxidase, and several cell-adhesion proteins. In total, over 140 CCP modules have been identified in 24 different proteins.

Solution-state two-dimensional $^1$H-NMR spectroscopy has allowed determination of the tertiary structure of the 16th module of human complement factor H (H16) (Barlow et al., 1991; Norman et al., 1991), expressed as a recombinant gene product in isolation from the rest of the molecule. The primary sequence of H16 is highly representative of the CCP family, containing all the consensus residues (Figure 1). The three-dimensional structure of H16 is based on a $\beta$-sandwich arrangement, with one face comprised of three $\beta$-strands hy-

† This work is a contribution from the Oxford Centre for Molecular Science, which is supported by the SERC and MRC. T.J.H. was supported by a MRC postdoctoral training fellowship. P.C.D. holds a Royal Society University Research Fellowship.

‡ Department of Biochemistry.

§ MRC Immunochemistry Unit, Department of Biochemistry.

[1] Abbreviations: CCP, complement control protein; H5 and H16, the fifth and the sixteenth repeats of human factor H of complement, respectively; DQF, double quantum filtered; COSY, correlation spectroscopy; HOHAHA, homonuclear Hartmann-Hahn; NOE, nuclear Overhauser effect; NOESY nuclear Overhauser effect spectroscopy; PE, primitive exclusive; rms, root mean square; MIN-AV, minimized and averaged.

0006-2960/92/0431-3626$03.00/0 

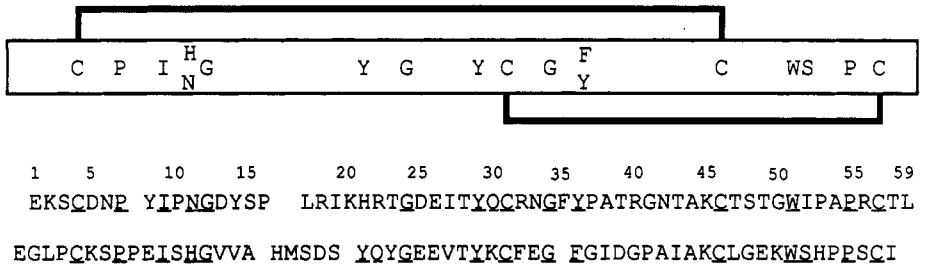


FIGURE 1: Sequences of H5 and H16. Consensus residues are underlined. The consensus sequence is boxed; consensus residues are spaced according to their position in the sequence. Disulfide bridges are show as thick black lines. The residues of H5 were numbered as indicated.

drogen bonded to form a triple-stranded region at its center and two separate β-strands forming the other face. Both faces of the sandwich contribute highly conserved hydrophobic side chains to form a compact core. The regions connecting the β-strands are comprised both of tight turns and of less well defined loops.

The 5th repeat of factor H (H5) was chosen as a second example for structural investigation. H5 contains 14 consensus residues but has less than 30% identity with H16 (Figure 1). Hence it provides the basis for a useful comparison in the context of seeking a consensus secondary and tertiary structure. In addition, it may be directly involved in the C3b-binding function of factor H. A proteolytic fragment comprising the first $5^1/_2$ modules of factor H possesses C3b-binding and factor I cofactor activity (Alsenz et al., 1985). Studies with monoclonal antibodies localized the C3b-binding site to regions within modules 2–5 (Alsenz et al., 1985). One such monoclonal antibody, MRC OX-24 (Sim et al., 1983), which inhibits binding of factor H to C3b, binds to H5 (Steinkasserer et al., 1991). Thus, elucidation of the tertiary structure of H5 may eventually provide some evidence for the way in which modules are tailored to perform specific functions.

## Materials and Methods

*Expression and Purification of H5 and Numbering Convention.* The fifth repeat from human complement factor H was expressed in a yeast secretion system (Baron et al., 1990b). Yeast cells were transformed with the shuttle plasmid pMA91 containing coding sequence for a phosphoglycerate kinase promoter (Mellor et al., 1983; Kingsman & Kingsman, 1985) and the yeast α-factor leader peptide (Kurjan & Herskowitz, 1982) fused in frame with the H5 exon to direct the expressed protein into the culture medium.

Plasmid DNA from the clone B 38-1 (Ripoche et al., 1988) was used as a template for the PCR amplification of H5. The following synthetic oligonucleotides were used:

5′–GAA AAA TCA TGT GAT AAT CCT TAT ATT CC
3′–CGG ATC CTA CAA GGT ACA TCT CGG AGC

The PCR product was digested with the restriction enzyme *Bam*H1 and subcloned into the vector pKNa containing a unique polylinker with *Stu*1 and *Bam*H1. The correct nucleotide sequence for H5 was checked using the double-strand sequencing method (Pharmacia Kit T7, Milton Keynes).

Construction of the expression plasmid and transformation of yeast cells were carried out as described previously for H16 (Barlow et al., 1991). Cells were grown and protein was harvested as before (Barlow et al., 1991).

H5 was purified by cation-exchange chromatography (gradient of 0.0–0.6 M KCl in 10 mM phosphate buffer, pH 5.3, on a Mono-S column, Pharmacia) followed by reverse-phase chromatography (25–38% $CH_3CN$, 0.1% trifluoroacetic acid in water, C8 column, Beckman). The purified material was subjected to N-terminal sequence analysis using automated Edman degradation (Applied Biosystems 470A protein gas-phase sequencer with on-line PTH analysis), capillary electrophoresis (Applied Biosystems 270A Capillary Electrophoresis System), and mass spectroscopy (VG Bio-Tech, VG Bio-Q). These techniques confirmed that the desired material had been purified to >98%.

The residues of H16, as a recombinant protein, were numbered sequentially starting with the first N-terminal residue (Barlow et al., 1991). Lacking an absolute knowledge of alignment, it was decided to adopt a similar convention for H5 (Figure 1). Consequently, equivalent residues may have different numbers, e.g., Trp52 in H16 is equivalent to Trp51 of H5.

*NMR Analysis.* Purified CCP5 was dissolved in 0.5 mL of either $D_2O$ or 90% $H_2O$/10% $D_2O$ to give an approximate concentration of 3 mM. Solution state $^1$H-NMR spectra were recorded on either a Bruker AM 500 or 600 spectrometer at 300 K and at pH 4.7. Double-quantum-filtered correlation spectroscopy (DQF-COSY), nuclear Overhauser effect spectroscopy (NOESY) (with mixing times of 100 or 250 ms), and homonuclear Hartmann–Hahn (HOHAHA) spectroscopy were performed with both $D_2O$ and $H_2O$ samples, while the primitive exclusive (PE)-COSY was with a $D_2O$ sample. Details of experimental methods are as described previously (Barlow et al., 1991).

Amide hydrogens which were relatively slow to exchange with solvent were identified by adding $D_2O$ to freeze-dried, protonated H5 and recording 1-D spectra at intervals over 1 h, followed by a DQF-COSY over 12 h.

*Assignment.* Assignment of protons to cross-peaks was achieved using a 250-ms mixing time NOESY spectrum in combination with COSY and HOHAHA data. Two factors contributed to a relatively rapid assignment process. It quickly emerged that H5 has a secondary structure very similar to that of H16. For example, the strong $\alpha$–$\alpha$ connectivities observed between positions 14 and 29 (both tyrosines in H5) and between 28 and 43 (both threonines in H5) were readily identifiable. Additionally, there was a general consistency in the chemical shifts of the backbone protons of many residues in equivalent positions in the two modules. This aided the connection of spin systems via sequential assignment.

*Structural Restraints.* Proton-proton distance restraints were determined from nuclear Overhauser enhancements (NOEs) recorded in NOESY experiments using a mixing time of 100 ms. NOEs were quantified by peak integration within the program FELIX version 1.1 (Hare, 1990) and converted to distance restraints as previously described (Norman et al., 1991). Where possible, $^3J_{\alpha N}$ coupling constants were calculated from a high-resolution DQF-COSY spectrum using an iterative least-squares fitting program. The $\alpha\beta$ coupling constants were measured directly from a PE-COSY spectrum. Stereospecific assignments and $\phi$ and $\chi_1$ angles were determined using the program STEREOSEARCH (Nilges et al., 1990). Hydrogen bonds and disulfide connections were included in

the structure calculation as described previously (Norman et al., 1991).

*Initial Structure Generation and Simulated Annealing.* All structure refinement was carried out using the program XPLOR (Brünger, 1990). Initial structures were generated using random $\phi$ and $\psi$ angles, with side chains in an extended conformation and idealized covalent geometry. Simulated annealing was carried out for a total time of 5.02 ps using a protocol described previously (Brünger, 1990), similar to that used in the hybrid distance geometry simulated annealing protocol (Nilges et al., 1988). The structures resulting from the simulated annealing were either retained or discarded on the basis of a distance restraint energy cutoff of 357 kJ $mol^{-1}$, giving a pass rate of approximately 50%. A total of 49 structures were finally energy-minimized, using a Powell conjugate gradient algorithm operating under a full force field for 200 cycles. Final distance restraint energies of 4200 kJ $mol^{-1}$ $nm^{-2}$ and angle restraint energies of 210 kJ $rad^{-2}$ were used. An average structure was produced by superposition of all structures onto the structure with the lowest overall energy using only backbone atoms, followed by coordinate averaging. The average structure was subjected to restrained energy minimization to produce a structure with good covalent geometry [the minimized average (MIN-AV) structure].

The 49 final structures were analyzed within XPLOR (Brünger, 1990) for agreement with experimental distance restraints, potential energies, rms coordinate deviation, and $\phi$ and $\psi$ angle restraints.

## Results

*Expression and Purification.* H5 was expressed in a yeast system and purified using standard techniques. The purity, as tested by capillary electrophoresis, was shown to be >98%. The yield of purified material was only 1 mg per 10 L of yeast culture. This yield is typical for CCP modules in this expression vector and compares unfavorably with that obtained for other module types (Baron et al., 1990). The molecular mass was determined to be 6633 ± 2 daltons by electrospray MS comapred to a theoretical molecular mass of 6632.4 daltons. This indicates that no side-chain modification has occurred. Correct expression was also checked by N-terminal sequencing, which confirmed that the first 40 residues were correct.

*NMR analysis.* To check for the presence of a single folded form under the conditions employed, NOESY spectra recorded in water were examined. The spectra were dominated by a pattern of cross-peaks which closely resembled that seen for H16, indicative of a predominantly homogeneous fold. However the presence of a few low broad resonances (Figure 2a) indicated some local conformational flexibility. These decreased in intensity with increasing pH (from 3.5 to 5.7) but did not disappear completely.

It was possible to assign nearly all of the protons in H5 (see Table I) using standard techniques (Wüthrich, 1986). Figure 2a illustrates the connection of spin systems for two segments of the protein. The broad resonances, and a small number of sharper cross-peaks, remained unaccounted for. There are six prolines in H5. Four of these were sequentially assigned from $\beta_i$–$\delta_{i+1}$ and $\alpha_i$–$\delta\eta_{i+1}$ connectivities, while two, Pro16 and Pro53, exhibited strong $\alpha_i$–$\alpha_{i+1}$ connectivities (see Figure 2b) indicative of a cis conformation for the XXX–Pro peptide bond (Wüthrich, 1986). Difficulty was encountered in tracing the side chains of Arg56 and of the three lysine residues beyond the $\beta$-protons, presumably because these residues are surface exposed and conformationally mobile. The hydroxyl proton of Tyr29 was assigned in H5. This type of proton is normally

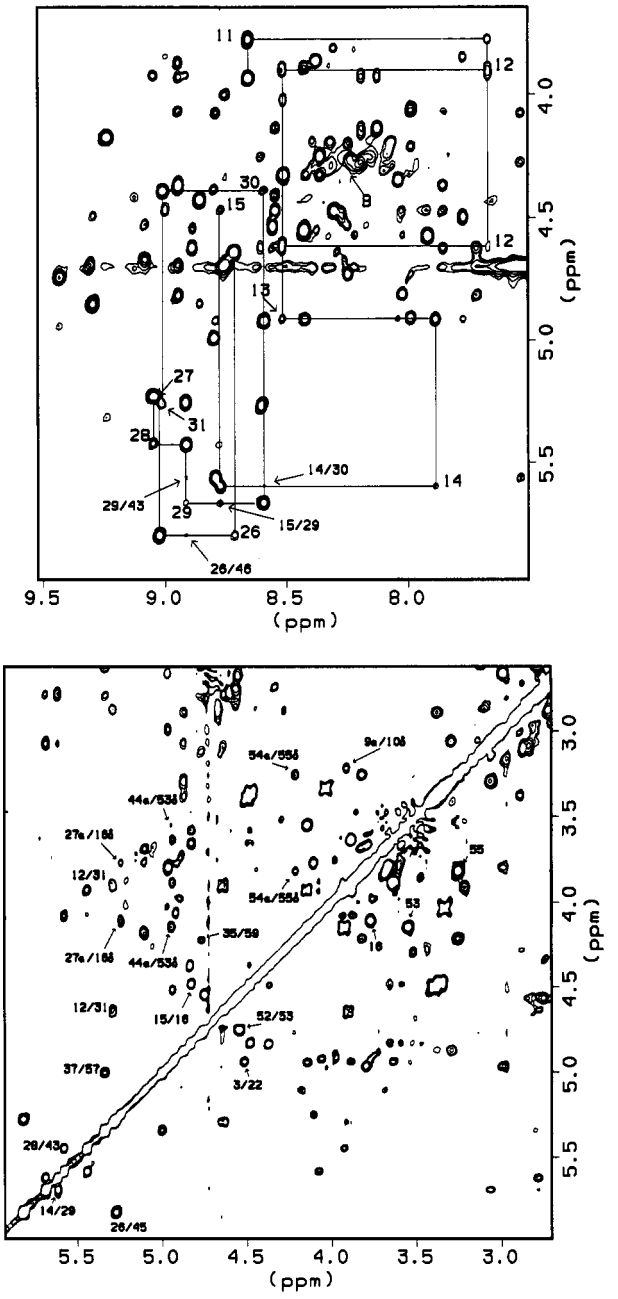


FIGURE 2: Portions of a NOESY spectrum of H5 in 90% $H_2O$/10% $D_2O$ at pH 4.7 and 300 K with a mixing time of 100 ms. (a, top) Region shown is that containing some of the cross-peaks between the backbone NH (*x*-axis) and the $\alpha$H resonances. Two sets of sequential NOE connectivities are indicated by unbroken lines with the labels at the intraresidue $\alpha$CH–NH cross-peaks. Some long-range $\alpha CH_i$–$NH_j$ cross-peaks are also labeled. The label "B" indicates the broad resonances referred to in the text, some of which were not assigned. (b, bottom) This region contains cross-peaks arising from $\alpha$CH and proline $\delta$CH resonances. The $\alpha_i$–$\alpha_j$ cross-peaks are labeled *i*/*j*. Note that in two cases these are sequential and correspond to the cis XXX–Pro peptide bonds (16 and 53) discussed in the text. An example of the more common $\alpha_{(i-1)}$–$\delta_i$ connectivity typical of a trans XXX–Pro peptide bond (55) is also indicated. Some intraresidue proline $\delta_i$–$\delta_i$ cross-peaks are labeled with the residue number. Finally the two pairs of $\alpha_i$–proline $\delta_j$ cross-peaks are labeled.

in rapid exchange with solvent and is not observed. Its presence here is indicative of its involvement in the formation of a hydrogen bond.

*Secondary Structure.* Figure 3a is a representation of the sequential NOEs observed for H5 and of the $^3J_{\alpha N}$ coupling constant data. A qualitative inspection of the data indicates that CCP5 consists predominantly of $\beta$-strands with turns in

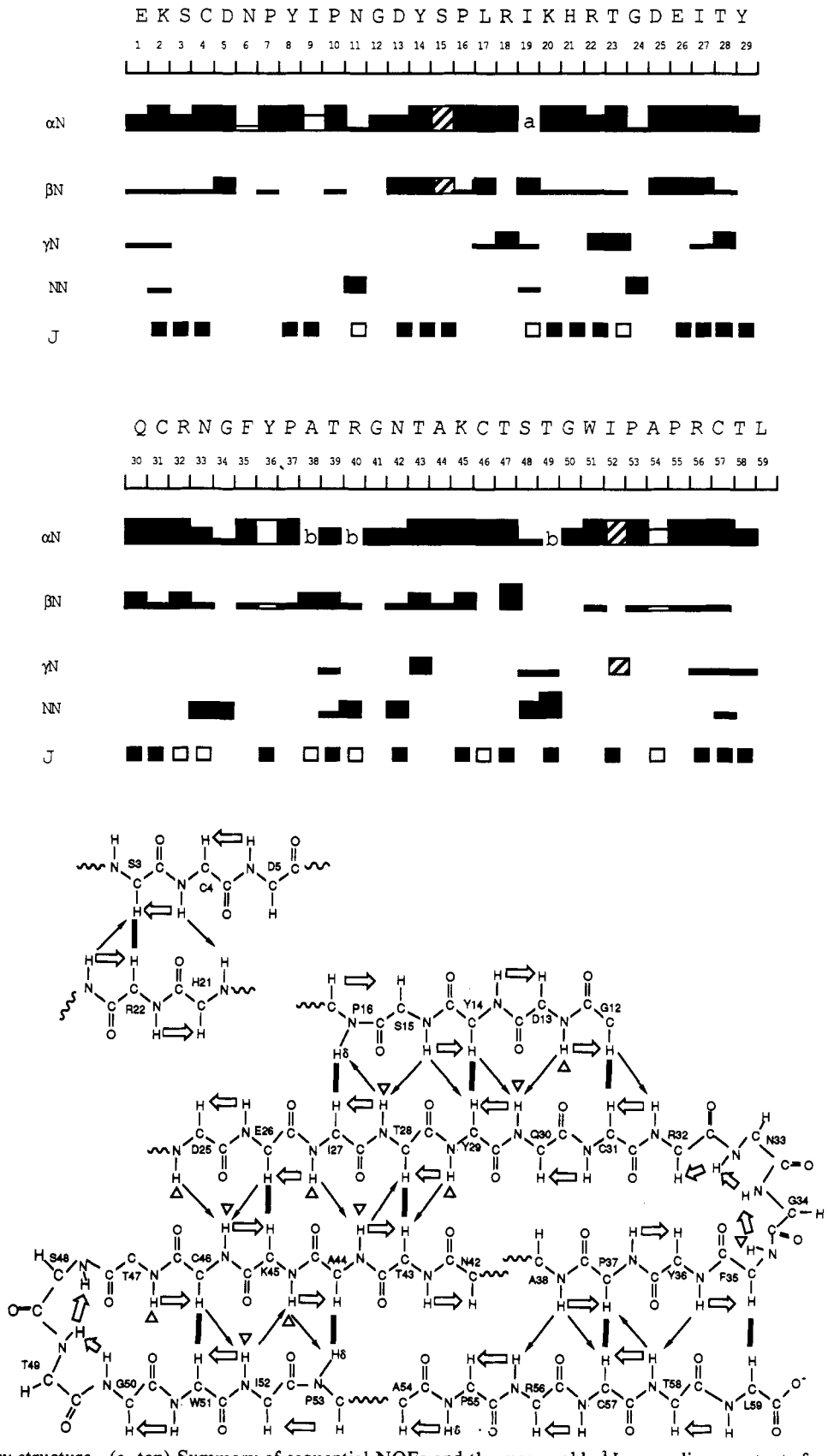


FIGURE 3: Secondary structure. (a, top) Summary of sequential NOEs and the measurable $^3J_{\alpha N}$ coupling constants for H5 at pH 4.7 and 300 K. The bar heights are indicative of strong, medium, or weak NOE intensities. Standard notation is used for the NOE classes. A white bar is used where the NOE arises from the δCH of proline, and hatched bars are employed to indicate that the NOE involves the αCH of proline. (a) Cross-peak detected at 250 ms mixing time. (b) NOE intensity cannot be estimated due to overlap with other cross-peaks. The symbols used for $^3J_{\alpha N}$ are as follows: black squares, $^3J_{\alpha N} > 7$ Hz; white squares, $^3J_{\alpha N} < 7$ Hz. (b, bottom) Secondary structural elements of H5 deduced from a qualitative inspection of the NOE and slowly exchanging amide proton data. The peptide backbone is labeled at the α-carbons. Thick black lines represent cross-sheet $\alpha_i$–$\alpha_j$ (or in two cases $\alpha_i$–proline $\delta_j$) NOEs. Thin straight black arrows indicate cross-sheet connectivities involving amide protons. Thick white arrows correspond to sequential NOEs. The triangular symbols indicate those amide protons which are slowly exchanging. The undulating lines represent portions of the backbone which have been omitted from the diagram for the purposes of clarity.

Table I: Assignment of Resonances in H5 at 300 K and pH 4.7[a]

| | | NH | αH | βH | γH | δH | other |
|---|---|---|---|---|---|---|---|
| 1 | E | | 4.02 | 2.06/2.10 | 2.30 | | |
| 2 | L | 8.76 | 4.34 | 1.36/1.64 | | | |
| 3 | S | 8.05 | 4.92 | 3.63/3.75 | | | |
| 4 | C | 8.43 | 4.53 | 2.67/1.92 | | | |
| 5 | D | 8.51 | 4.63 | 2.50/2.77 | | | |
| 6 | N | 8.89 | 4.54 | 2.57/2.67 | $\gamma NH_2$ 7.29/7.75 | | |
| 7 | P | | 3.86 | 0.49/0.60 | −0.31/0.24 | 2.54/2.77 | |
| 8 | Y | 8.38 | 4.55 | 2.81/2.85 | | | 2,6H 7.04; 3,5H 6.62 |
| 9 | I | 8.43 | 3.89 | 1.48 | $CH_2$ 0.23/0.82; $CH_3$ 0.09 | 0.15 | |
| 10 | P | | 3.94 | 1.53/1.99 | 1.49/1.80 | 2.15/3.22 | |
| 11 | N | 8.66 | 3.78 | 1.50/2.12 | $\gamma NH_2$ 6.66/6.90 | | |
| 12 | G | 7.67 | 3.90/4.62 | | | | |
| 13 | D | 8.52 | 4.92 | 2.40/2.57 | | | |
| 14 | Y | 7.89 | 5.60 | 2.45/2.77 | | | 2,6H 6.59; 3,5H 6.67 |
| 15 | S | 8.78 | 4.48 | 3.58/3.66 | | | |
| 16 | P | | 4.81 | 2.03/2.29 | 1.93/2.09 | 3.77/4.10 | |
| 17 | L | 8.03 | 4.58 | 1.54/1.57 | 1.40 | 0.60/0.77 | |
| 18 | R | 7.93 | 4.48 | −0.02 | 1.33/1.42 | 3.06/3.29 | εNH 7.44 |
| 19 | I | 8.32 | 3.82 | 1.75 | $CH_2$ 1.19/1.46; $CH_3$ 0.87 | 0.82 | |
| 20 | K | 7.19 | 4.43 | 1.50/1.57 | 1.31 | | |
| 21 | H | 8.87 | 4.85 | 2.88/3.36 | | | 2H 7.26; 4H 8.27 |
| 22 | R | 9.10 | 4.50 | 1.48 | 1.54/1.57 | 3.09 | εNH 7.43 |
| 23 | T | 7.78 | 3.46 | 3.85 | 0.90 | | |
| 24 | G | 9.01 | 3.36/4.47 | | | | |
| 25 | D | 8.30 | 4.64 | 2.86/3.08 | | | |
| 26 | E | 8.72 | 5.80 | 1.98/2.13 | 2.27/2.47 | | |
| 27 | I | 9.03 | 5.23 | 1.77 | $CH_2$ 0.10/0.96; $CH_3$ 0.63 | −0.33 | |
| 28 | T | 9.06 | 5.42 | 3.92 | 1.11 | | |
| 29 | Y | 8.92 | 5.67 | 2.28/3.06 | | | 2,6H 6.50; 3,5H 6.46; 4H 9.37 |
| 30 | Q | 8.60 | 4.39 | 1.81/1.94 | 2.12 | | |
| 31 | C | 9.02 | 5.26 | 2.43/2.86 | | | |
| 32 | R | 8.61 | 4.25 | 1.52/1.80 | 1.39/1.50 | 2.88/3.01 | εNH 7.47 |
| 33 | N | 8.37 | 4.33 | 2.74 | $\gamma NH_2$ 6.66/6.90 | | |
| 34 | G | 8.52 | 3.33/4.02 | | | | |
| 35 | F | 8.00 | 4.73 | 2.60/3.09 | | | 2,6H 6.75; 3,5H 7.11; 4H 7.08 |
| 36 | Y | 9.44 | 4.94 | 2.65/2.98 | | | 2,6H 7.05; 3,5H 6.64 |
| 37 | P | | 4.99 | 2.06/2.18 | 1.80 | 3.79 | |
| 38 | A | 8.81 | 4.39 | 1.40 | | | |
| 39 | T | 7.31 | 4.40 | 4.47 | 1.16 | | |
| 40 | R | 8.55 | 4.14 | 1.81 | 1.58/1.67 | 2.98/3.17 | |
| 41 | G | 7.13 | 3.93/4.13 | | | | |
| 42 | N | 8.20 | 4.28 | 2.45/2.88 | $\gamma NH_2$ 6.59/7.37 | | |
| 43 | T | 7.54 | 5.56 | 4.07 | 1.17 | | |
| 44 | A | 8.80 | 4.92 | 1.47 | | | |
| 45 | K | 8.59 | 5.25 | 1.27/1.73 | | | |
| 46 | C | 8.72 | 3.49 | 1.43/2.41 | | | |
| 47 | T | 7.72 | 4.81 | 4.37 | 1.10 | | |
| 48 | S | 8.95 | 4.07 | 3.87/3.93 | | | |
| 49 | T | 7.86 | 4.46 | 4.37 | 1.00 | | |
| 50 | G | 7.32 | 3.40/4.49 | | | | |
| 51 | W | 8.28 | 4.67 | 2.72/2.88 | | | 2H 7.12; 4H 6.81; 5H 6.84; 6H 6.52; 7H 6.43; NH 10.08 |
| 52 | I | 9.09 | 4.52 | 1.76 | $CH_2$ 1.10/1.30; $CH_3$ 0.81 | 0.68 | |
| 53 | P | | 4.73 | 2.00/2.43 | 1.97/2.01 | 3.55/4.14 | |
| 54 | A | 8.25 | 4.20 | 1.25 | | | |
| 55 | P | | 3.53 | 0.22/0.45 | 0.80/1.09 | 3.26/3.82 | |
| 56 | R | 7.28 | 4.17 | 1.58/1.62 | 1.34/1.45 | | |
| 57 | C | 9.25 | 5.32 | 2.60/2.78 | | | |
| 58 | T | 9.57 | 4.91 | 4.06 | 1.03 | | |
| 59 | L | 7.99 | 4.21 | 1.43 | 1.37 | 0.40/0.70 | |

[a] Chemical shifts are with respect to an external standard of dioxane (3.75 ppm).

the vicinity of Asn11, Gly24, Asn33, and Ser48. The region Arg39–Asn42 exhibits a pattern of $N_i$–$N_{i+1}$ NOEs which is not consistent with regular secondary structure. There is no evidence for any α-helical regions (Wüthrich, 1986).

The long-range NOEs observed in the case of H5 (see Figure 3b) are indicative of five extended strands separated by four turns. Gly12–Pro16, Asp25–Cys31, Asn42–Thr47, and Gly50–Pro53 appear to form a four-stranded antiparallel β-sheet, and this is consistent with the location of amides that appear to be in slow exchange. It is noteworthy that the two *cis*-prolines, 16 and 53, exhibit $\delta_i$–$\alpha_j$ connectivities across the sheet instead of the $\alpha_i$–$\alpha_j$ observed for trans residues. Pro16 and Pro53 terminate the first and the last strands in the four-stranded sheet, respectively. Residues Ser3–Cys4 appear to pair with His21–Arg22 in a short section of double-stranded sheet toward the N terminus, and Phe33–Pro37 and Cys57–Leu59 might form a double-stranded sheet at the C terminus. However, these regions do not contain slowly exchanging amides. Residues Thr39–Arg42 must form a looped out segment in order to be accommodated within the register of the β-sheets, an inference which is supported by the presence of several medium-strong $N_i$–$N_{i+1}$ NOEs in this region.

*Tertiary Structure.* A total of 407 distance restraints [226 long-range ($|i-j| > 4$) and 181 short-range ($|i-j| < 5$)] were determined from NOE cross-peak integration. Stereospecific assignments, along with 24 φ angle and 10 $\chi_1$ angle restraints,

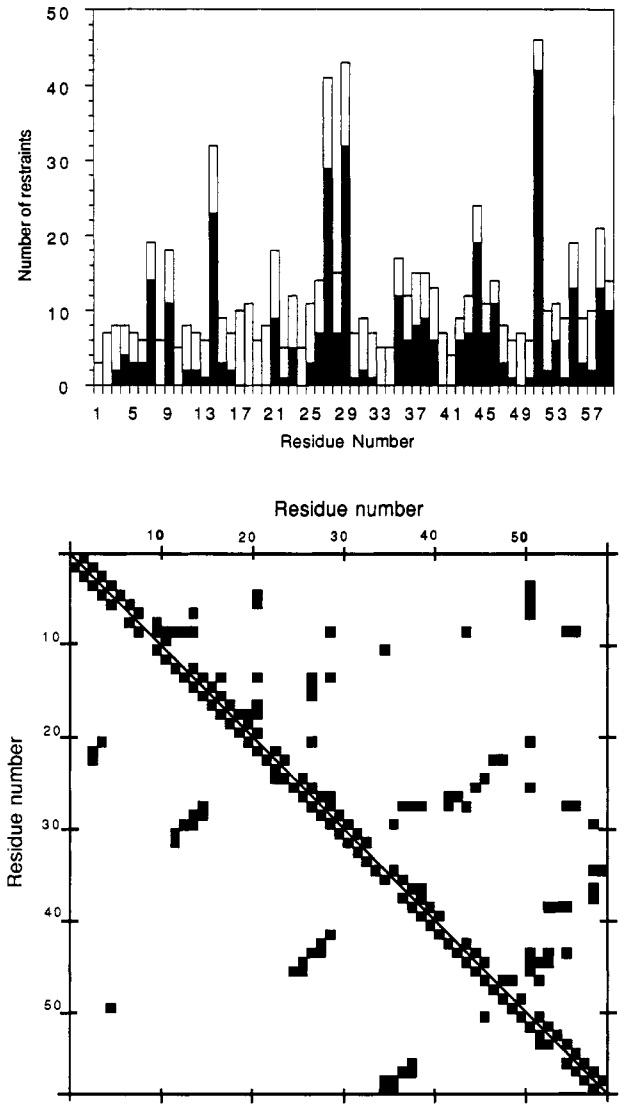


FIGURE 4: Distribution of restraints. (a, top) Sequential. Filled bars represent long-range ($|i-j| > 4$) restraints, and open bars represent short-range ($|i-j| < 5$)] restraints. (b, bottom) Spatial. Filled squares indicate that one or more distance restraints were observed between the residues horizontally and vertically. The bottom left of the diagram shows the restraints involving only main-chain protons, Hα and HN. The top right of the diagram shows the restraints involving at least one side-chain proton.

were obtained using the program STEREOSEARCH (Nilges et al., 1990). A further 24 distance restraints were used to define 12 predicted hydrogen bonds. Figure 4a shows the sequential distribution of long-range and short-range distance restraints. The number of short-range restraints is relatively constant throughout the sequence (5–10 per residue), while the long-range restraints have a patchy distribution with clusters (20–44 per residue) at the hydrophobic residues Tyr14, Ile27, Tyr29, Ala44, and Trp52. The unusually high numbers of NOEs involving these residues are consistent with the formation of a compact hydrophobic core. No long-range restraints are observed for residues 17–20. Figure 4b shows the spatial distribution of distance restraints arising from both main-chain (below the diagonal) and side-chain protons. There is a good spread of restraints, involving pairs of residues far apart from one another in the sequence, consistent with a compact globular protein.

Table II shows the distribution of potential energies observed in the 49 final structures. Violations of covalent geometry were within acceptable limits. The final structures showed a mean

Table II: Mean, Range, and Standard Deviation of Potential Energies in Final Minimized Structures and of the Average Minimized Structure[a]

| | mean | min | max | SD | AV–MIN |
|---|---|---|---|---|---|
| $F_{TOT}$ | 1213.2 | 1047.3 | 1397.3 | 82.9 | 2422.8 |
| $F_{NOE}$ | 50.4 | 38.3 | 60.6 | 5.1 | 41.2 |
| $F_{VDW}$ | −967.5 | −1051.8 | −910.6 | 28.1 | −859.3 |
| $F_{BOND}$ | 78.9 | 67.7 | 94.6 | 5.7 | 90.7 |
| $F_{ANGLE}$ | 852.6 | 798.4 | 931.9 | 29.6 | 1996.8 |
| $F_{DIHED}$ | 1261.6 | 1171.3 | 1378.6 | 49.8 | 1219.9 |
| $F_{IMPR}$ | 28.4 | 21.2 | 36.3 | 3.1 | 42.4 |
| $F_{ELEC}$ | −49.7 | −55.2 | −42.4 | 2.7 | −49.9 |
| $F_{HBOND}$ | −76.8 | −96.9 | −55.3 | 10.5 | −79.9 |
| $F_{CDIH}$ | 35.4 | 16.9 | 63.6 | 12.8 | 20.8 |

[a] The potential energies are $F_{TOT}$, total potential energy; $F_{NOE}$, energy due to distance restraint violations (42 kJ nm$^{-1}$); $F_{VDW}$, energy due to van der Waals interactions calculated from a Lennard-Jones equation; $F_{BONd}$, $F_{ANGLE}$, $F_{DIHED}$, and $F_{IMPR}$, potential energies due to covalent geometry and interaction; $F_{ELEC}$, electrostatic potential energy calculated using a constant dielectric of 80; $F_{HBOND}$, potential energy due to an explicit hydrogen-bond term, only backbone HN and O atoms are included; $F_{CIDH}$, energy due to restrained angles (84 kJ rad$^{-2}$). All energies are described in kJ mol$^{-1}$.

pair-wise coordinate difference of 0.275 nm (SD, 0.051 nm) for backbone atoms (Cα, C, N) and 0.523 nm (SD, 0.061 nm) for all atoms excluding hydrogens. The average rms (root-mean-square) coordinate differences for the 49 final structures, relative to the MIN–AV structure, was 0.21 nm (SD, 0.03 nm) for backbone atoms, and 0.40 nm (SD, 0.04 nm) for all non-hydrogen atoms. There was a mean rms deviation of 0.004 nm for distance restraints, with no distance-restraint violations greater than 0.047 nm. Angle restraints showed a mean rms deviation of 4°. The final converged structures are presented as a backbone atom overlay in Figure 5a. Residues contributing to the hydrophobic core and the disulfide bonds are shown superimposed on the MIN-AV structure backbone in Figure 5b. The N-terminal residues show considerable divergence, probably reflecting true disorder in this region of the isolated module. The 3rd, 4th, and 5th strands show excellent convergence, as do the residues within the core. The first and second strands are more disordered, particularly the region containing residues 17–20.

Residues 9–12 form a type II turn, residues 31–34 form another type II turn, and residues 47–50 form a slightly distorted type I turn. Residues 22–24 form a loose turn. Figure 6 shows a Ramchandran-type plot of the observed $\phi$ and $\psi$ angles in the MIN–AV structure. The non-glycine residues which show positive $\phi$ angles are Asn11 and Asn33. Both these asparagine residues lie within type II turns and thus may be considered conservative replacements for glycines (Wilmot & Thornton, 1988). The equivalent residue to Asn11 in H16, His13, also appeared to display a positive $\phi$ angle. This position is almost always occupied by a histidine or an asparagine in CCP modules.

Overall, the tertiary structure of H5 is appropriate to the role of a molecular scaffold. It is globular with a hydrophobic core wrapped in β-strands. The disulfide bonds are located as far apart as possible, and the C and N termini are disposed at opposite poles of the long axis of the molecule. The β-strands run parallel and antiparallel to the long axis with the first and third turns located close to the C terminus and the second and fourth turns near the N terminus. Most of the hydrophobic residues contribute to the core, exceptions being Tyr36 and the highly conserved Phe35.

## DISCUSSION

*Chemical Shifts.* H5 has been expressed in isolation, purified, and subjected to solution-state $^1$H-NMR spectroscopy.

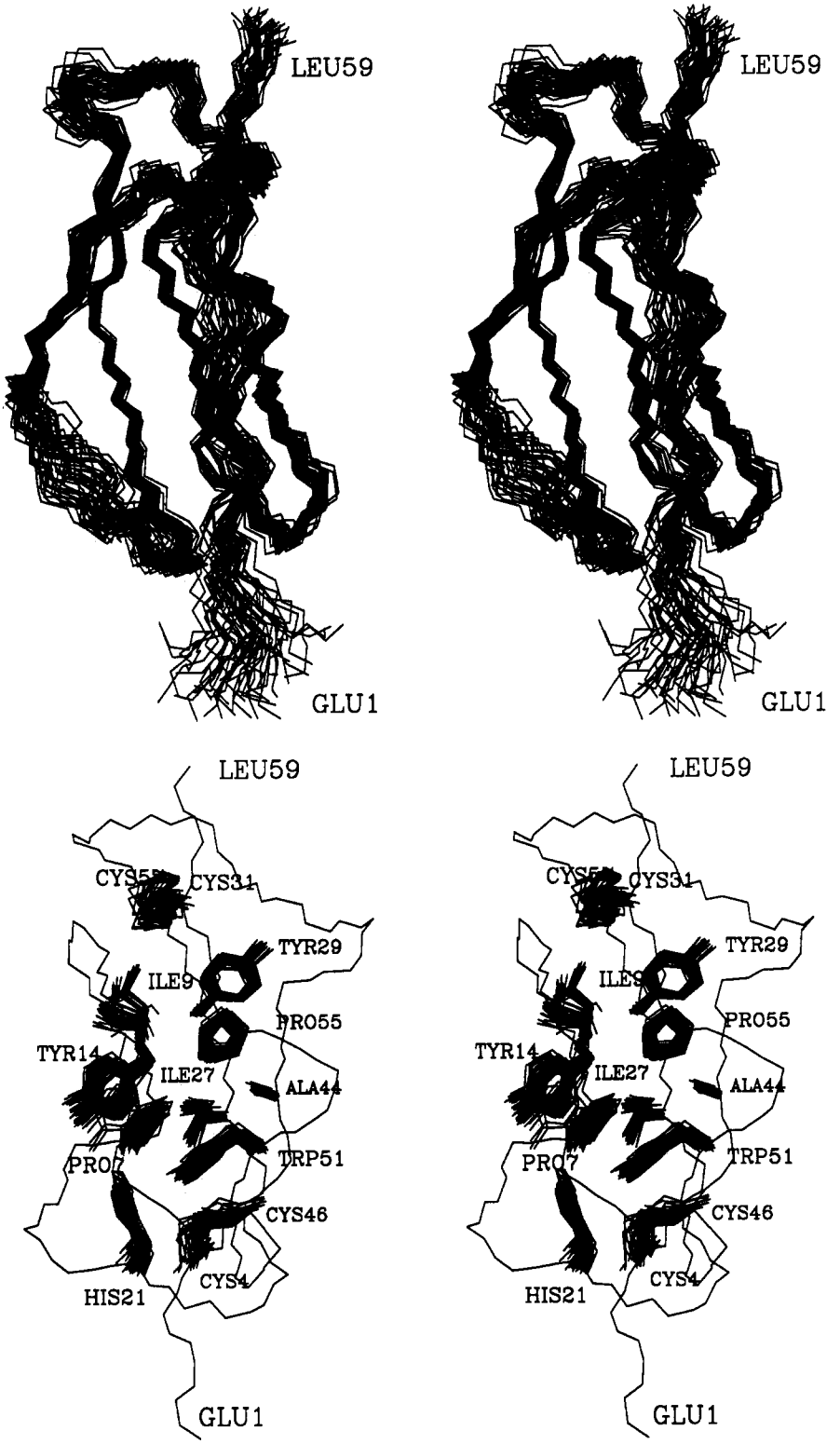


FIGURE 5: Calculated structures for H5. (a, top) Superimposition of C, Cα, and N (except in residues 1–3) shown as an overlay of all backbone atoms. (b, bottom) Superimposition of the side chains of core residues plus disulfides drawn on a backbone trace of the MIN-AV structure. Stereopictures are drawn for "cross-eye" viewing.

The sequential assignment procedure for H5 was facilitated by reference to the spectra of H16 (Barlow et al., 1991). This was possible because of the generally high degree of consistency in the upfield or downfield shifts exhibited by resonances of equivalent protons, in the two modules, for residues within the third, fourth, and fifth strands. These are well ordered in both H5 and H16 and constitute a region of high similarity in tertiary structure between the two modules. This conservation of chemical shift will expedite NMR-based structural studies on other modules of this type.

*Secondary Structure.* All four turns found in H5 are duplicated in H16 (Barlow et al., 1991). There is an extra "patch" of inferred β-sheet in H5 involving residues 3–4 and 21–22; however the lack of detectable slowly exchanging amides in this region does not indicate hydrogen bonds. There is slightly more extensive sheet formation in H5 giving rise to the four-stranded antiparallel β-sheet instead of the triple-stranded sheet observed in H16. Finally, there is some

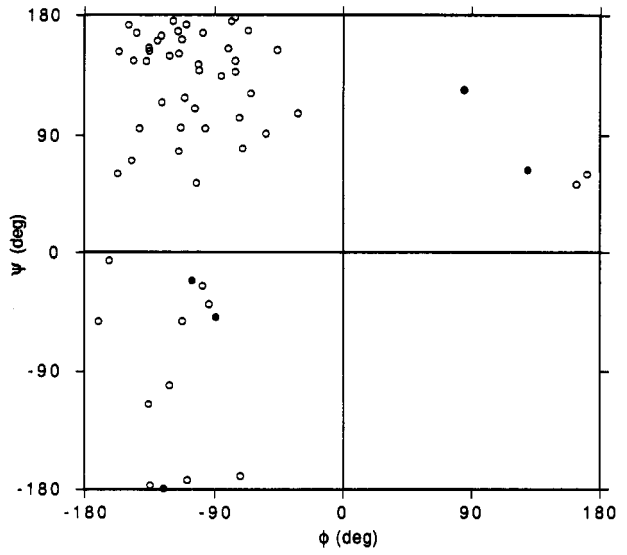


FIGURE 6: Ramchandran plot. Shown are the average $\phi,\psi$ angles observed in the final minimized structures. Open circles correspond to all residues except glycine; filled circles represent glycines.

NOE-based evidence for continuation of the $\beta$-sheet at the C terminus of H5, which is one residue longer, as expressed, than H16. This observation leads to the speculation that intermodular "linkers" (i.e., the regions between the fourth cysteine of one module and the first cysteine of the next) may be closely associated with the body of the module and contribute little flexibility to the intact protein. Overall, there is a close correspondence of secondary structure between the two modules. Hence, the 13 consensus residues, accounting for 81% of the identity seen between H5 and H16, appear to determine a consensus secondary structure. Thus all other CCP modules (i.e., modules containing the consensus sequence) can be expected to have a very similar secondary structure.

It is interesting that an alignment based on the 250 PAMS mutation data matrix (Dayhoff et al., 1983; Barton & Steinberg, 1987) treats Tyr36 of H5 as an insertion and hence aligns Ile39 and Asp40 of H16 with Ala38 and Thr39 of H5 (see Figure 1). This is not borne out by the secondary structure; the equivalent residues to Ile39 and Asp40 of H16 are Pro37 and Ala38 of H5, while Phe35 and Tyr36 of H5 replace Tyr37 and Gly38 of H16. The insertion could easily be positioned in the looped out region between Ala38 and Asn42 of H5. A modified alignment then leaves the length of turn 3 intact in all but one module of factor H, H13, which lacks three residues in the region. This use of structural data to improve alignments could be applied to the entire family of CCP modules. Such an alignment might provide a basis for the numbering of residues across the family.

*Tertiary Structure.* At the teritary level there are several interesting difference between H5 and H16 in an otherwise similar, scaffold-like structure. While both molecules are characterized by a compact hydrophobic core wrapped in $\beta$-sheet, the side chains making up the core of H5 exhibit a significantly different packing arrangement from those in H16. In particular, the highly conserved Tyr29 (equivalent to Tyr31 in H16) has undergone a 90° clockwise rotation about the C$\alpha$–C$\beta$ bond with respect to its orientation in H16, as illustrated in Figure 7. Its new position is within the loop containing residues 39–42. The hydroxyl proton of Tyr29 is slow to exchange with solvent and appears to be forming a hydrogen bond to the $\gamma$-oxygen of Thr39. This rearrangement of what seemed to be a critical element of the core of H16 may be

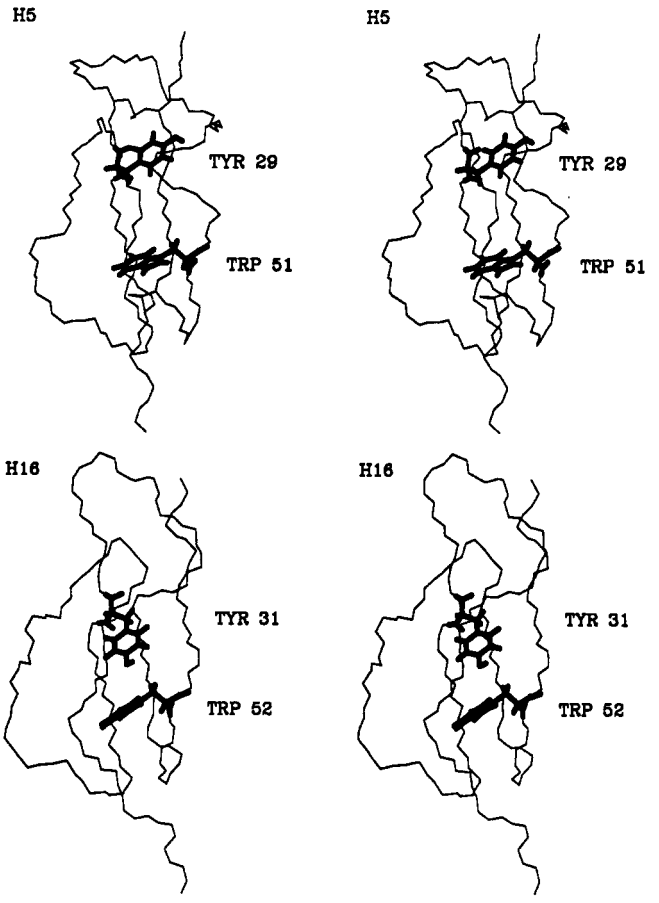


FIGURE 7: Position of Tyr29 in H5 is compared to that of its equivalent (Tyr31) in H16. Stereoviews; the tryptophan side chain is shown for reference.

explained by the deletion of one amino acid in the first strand of H5, relative to H16, and an insertion in the loop region made up by residues 39–42. The deletion shortens the core (the distance between the disulfides is 1.7 nm compared to 2.0 nm for H16), and Tyr29 is thus more easily accommodated within the extended loop containing the insertion. On the other hand, His23 has adopted almost exactly the same orientation, with respect to the core, as the equivalent highly conserved tyrosine in H16. The side-chain volume of histidine is 40 Å$^3$ smaller than that of tyrosine and it is much more polar. Therefore, it violates the "rules" for allowable substitutions in hydrophobic cores, as defined by Bordo and Argos (1990).

In the case of H16, a putative functional site was identified in the region between residues 17 and 22 (Norman et al., 1991). This region, a site of very low sequence conservation among CCP modules, exhibited very few long-range NOEs and appeared to possess a high level of local mobility. In H5, the presence of *cis*-Pro at position 16, together with the inferred $\beta$-sheet involving residues 21 and 22, results in an increased definition of this region. However, there are no long-range NOEs involving residues 17–20, and this region is still poorly defined compared to much of the rest of the molecule. It therefore remains a good candidate for a variable surface region that is not involved in the integrity of the scaffold-like structure and that might provide a range of specific protein–protein interaction sites on CCP modules.

H5 and H16 each have a patch of exposed residues, both hydrophobic and charged, toward the C terminus (Asn11, Phe35, and Tyr36 in H5; His13, Phe34 and Phe37 in H16) which may be buried in the interface between modules in the intact protein. If the intermodular interface is extensive, modules must be orientated in a defined way with respect to

their neighbors. This agrees well with the growing evidence that particular ligand-binding sites may be spread out over two or more modules (Klickstein et al., 1988; Krych et al., 1991; Martin et al., 1991).

In conclusion, the determination of the three-dimensional structure of a second CCP module has confirmed that the consensus sequence confers extensive topological similarities. This is encouraging, since it implies that knowledge-based prediction of protein structures, such as described by Kaden et al. (1990), will be applicable to CCP modules. Eventually, this should allow accurate modeling of all 140 members of the family. Of particular interest will be the growing number of CCP modules for which site-directed mutagenesis data have been correlated with function (Krych et al., 1991: Martin et al., 1991; Molina et al., 1991). Indeed, our identification of the region involving residues 17–20 as a putative ligand-binding site already provides a basis for further mutagenesis work. However, the unexpected differences in the packing of the core residues emphasizes the need for caution in homology modeling from a limited database. Efforts to expand this database are on-going in this laboratory.

ACKNOWLEDGMENTS

We are grateful to Dr. A. J. Day for useful discussions, to Mr. A. Benham for assistance in expression of H5, and to Dr. J. M. Ahearn for allowing us access to unpublished data. Mr. A. C. Willis carried out protein sequence and capillary electrophoresis. Mass spectrometry was undertaken by Dr. R. Aplin.

REFERENCES

Alsenz, J., Schulz, T. F., Lambris, J. D., Sim, R. B., & Dierich, M. P. (1985) *Biochem. J. 323*, 841–850.

Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B., & Campbell, I. D. (1991) *Biochemistry 30*, 997–1004.

Baron, M., Norman, D. G., & Campbell, I. D. (1990a) *Trends Biochem. Sci. 16*, 642–646.

Baron, M., Kingsman, A., Kingsman, S. M., & Campbell, I. D. (1990b) in *Protein Production in Biotechnology* (Harris, T. J. R., Ed.) pp 49–60, Elsevier, London.

Barton, G. J., & Sternberg, M. J. E. (1987) *J. Mol. Biol. 198*, 327–337.

Bordo, D., & Argos, P. (1990) *J. Mol. Biol. 211*, 975–988.

Brünger, A. T. (1990) XPLOR *v2.2*, Yale University, New Haven, CT.

Day, A. J., Campbell, R. D., & Reid, K. B. M. (1989) in *Progress in Immunology VII* (Melchers, F., et al., Eds.) pp 209–212, Springer-Verlag, Heidelberg.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Hare, D. R. (1990) FELIX *1.1 manual*, Hare Research, Inc., Woodville, CA.

Kaden, F., Koch, I., & Selbig, J. (1990) *J. Theor. Biol. 147*, 85–100.

Kingsman, A. J., & Kingsman, S. M. (1985) *Biotech. Genet. Eng. Rev. 3*, 377–416.

Klickstein, L. B., Bartow, T., Miletic, V., Rabson, L. D., Smith, J. A., & Fearon, D. T. (1988) *J. Exp. Med. 168*, 1699–1717.

Krych, M., Hourcade, D., & Atkinson, J. P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4353–4357.

Kurjan, J., & Herskowitz, I. (1982) *Cell 30*, 933–943.

Martin, D. R., Yuryev, A., Kalli, K. R., Fearon, D. T., & Ahearn, J. M. (1991) *J. Exp. Med. 174*, 1299–1311.

Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J., & Kingsman, S. M. (1983) *Gene 24*, 1–14.

Molina, H., Brenner, C., Jacobi, S., Gorka, J., Carel, J.-C., Kinoshita, T., & Holers, V. M. (1991) *J. Biol. Chem. 266*, 12173–12179.

Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B., & Campbell, I. D. (1991) *J. Mol. Biol. 219*, 717–725.

Nilges, M., Gronenborn, A. M., & Clore, G. M. (1988) *FEBS Lett. 229*, 317–324.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1990) *Biopolymers 29*, 813.

Patthy, L. (1991) *Curr. Opin. Struct. Biol. 1*, 351–361.

Ripoche, J., Day, A. J., Harris, T. J. R., & Sim, R. B. (1988) *Biochem. J. 249*, 593–602.

Sali, A., Overington, J. P., Johnson, M. S., & Blundell, T. L. (1990) *Trends Biochem. Sci. 15*, 235–240.

Sim, E., Palmer, M. S., Puklavec, M., & Sim, R. B. (1983) *Biosci. Rep. 3*, 1119–1131.

Sim, R. B., & Perkins, S. J. (1989) *Curr. Top. Microbiol. Immunol. 153*, 209–222.

Steinkasserer, A., Barlow, P., Norman, D. G., Kertesz, Z., Campbell, I. D., Day, A. J., & Sim, R. B. (1991) *Complement Inflammation 8*, 225–226.

Wilmot, C. M., & Thornton, J. M. (1988) *J. Mol. Biol. 203*, 221–232.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, p 17, Wiley, New York.