The structure of a 19-residue fragment from the C-loop of the Fourth Epidermal Growth Factor-like Domain of Thrombomodulin

Marc Adler, Marian H. Seto, Danute E. Nitecki, Jiing-Huey Lin, David R. Light, and John Morser
From Berlex Bioscience, Inc., Richmond, California 94804-0099

The solution structure has been determined for a 19-residue peptide that is fully folded at room temperature. The sequence of this peptide is based on the C-loop, residues 371-389, of the fourth epidermal growth factor-like domain of thrombomodulin, a protein that acts as a cofactor for the thrombin activation of protein C. Despite its small size, the peptide forms a compact structure with almost no repeating secondary structure. The results indicate the structure is held together by hydrophobic interactions, which in turn stabilize the two β-turns in the structure. The first β-turn in the C-loop represents a conserved motif that is found in the published structures of five other epidermal growth factor-like proteins. The critical role of Phe376 in the stabilization of the first β-turn is consistent with mutagenesis data with soluble thrombomodulin. The results also show that a small subdomain of a larger protein can fold independently, and therefore it could act as an initiation site for further folding.

Thrombomodulin (TM), an endothelial cell surface glycoprotein, binds thrombin and alters its specificity away from fibrinogen cleavage and toward the activation of protein C. The activation of protein C by thrombin is accelerated 1000-fold when TM is present as a cofactor. Generation of activated protein C, which inactivates factor Va and factor VIIIa, is an important anticoagulant mechanism of the endothelial cell surface (1-3). TM is a multidomain protein, which spans the endothelial cell membrane. Full cofactor activity is present in the soluble ectodomain produced by elastase (4). Several studies performed with mutagenesis or with peptides derived from the ectodomain of human TM have defined the domains required for activity. The smallest fragment with full cofactor activity for the activation of protein C by thrombin contains the last three consecutive EGF-like domains, EGF4-6 (5, 6). Recent studies suggest that the region of thrombomodulin that binds tightly to thrombin is distinct from the region that modulates the active site of thrombin. Significantly, a construct made from the fourth and fifth EGF-like domains, EGF45, retains approximately 10% of the cofactor activity, although binding to thrombin is drastically reduced (5).

Deletion mutants near EGF4 effect $k_{cat}/K_m$ for the thrombin-TM complex with protein C, and removal of the fourth domain results in a complete loss of cofactor activity. However, deletion mutants that include the C-loop of EGF6 have a normal $k_{cat}/K_m$ for protein C but decreased affinity for thrombin, as measured by the $K_d$ for thrombin (6). A cyclic peptide based on the C-loop of EGF5 and the interdomain loop between EGF5 and EGF6 binds with high affinity to thrombin at the anion exosite of thrombin, a positively charged groove on the surface of thrombin important for TM, fibrinogen, and hirudin binding (5, 7, 8). The results suggest that EGF56 contains the high affinity binding site for thrombin and EGF4 contains residues that are absolutely required for activity.

Site-specific mutants around EGF4, which result in low activity analogs, have defined some residues important for cofactor activity in this domain (9, 10). This includes Asp345 in the interdomain region between EGF3 and EGF4, Glu357 and Tyr358 in the B-loop, and Phe376 in the C-loop of EGF4 (the numbering of the residues in this paper is consistent with the sequence of thrombomodulin given by Suzuki et al. (11)). Met388 in the interdomain region between EGF4 and EGF5 adjacent to the C-loop of EGF4 can be oxidized to the low activity methionine sulfoxide analog (12). Perhaps of more interest are EGF4 mutants, which result in an increase in cofactor activity. Replacement of Met388 by leucine results in an analog with twice the cofactor activity of wild-type TM (13).

When a second mutation is introduced in the C-loop, His381 → Gly, and combined with the Met388 → Leu mutation, the resultant TM analog has four times the activity of wild-type TM. Clearly, one way these mutants could modulate cofactor activity is by altering the conformation of TM. To test this hypothesis, a set of cyclic peptides was synthesized based on the sequence of the individual loops of TM. Each loop contained a single disulfide. NMR was used to measure structures of these peptides in solution. It was hoped that a comparison between several peptides with single site mutations would shed light on the relationship between structure and function.

Of course, these structural comparisons could only be made if the peptides were folded. In our experience, most peptides of this size do not fold in aqueous solutions. Indeed, there are relatively few structures of peptides of this length listed in the Brookhaven Protein Data Bank. There are seven solution structures for peptides of less than 30 residues (Spring, 1994).
All of these compounds contain at least two internal cross-links, which form the core of the structures. Most of the remaining residues form hairpin loops that are wrapped around a densely packed central core. During the course of this study, NMR was used to investigate the structure of nine different peptides. Each peptide contained approximately 20 residues and a single disulfide cross-link. The sequences were based on a single loop found either in TM or a homologous EGF-like protein. For most of the peptides, the spacing between the cysteines was much greater than the loops found in the small peptides of the protein bank data. Therefore, there must be a significant loss of conformational entropy during refolding. The experimental results indicated that the only peptides that formed a compact structure were based on the sequence of the C-loop of TM-EGF4 (Table I). This was an unexpected result, since the peptides contained a long stretch of 13 amino acids between the cysteines. The experimental results were used to evaluate the structural homology between TM-EGF4 and other EGF-like proteins or domains that, in turn, provided a working hypothesis that explains some of the functional data.

EXPERIMENTAL PROCEDURES

Materials—All peptides used in this study (Table I) were synthesized by solid phase synthesis methods on a Biosearch 9600 Peptide Synthesizer. The peptides were cleaved by trifluoroacetic acid containing small amounts of anisole, thioanisole, and ethanedithiol and were purified by preparative high pressure liquid chromatography on a reverse phase C18 column using 0.1% trifluoroacetic acid in an acetonitrile/water gradient. The cyclization was achieved in dilute water/dimethyl sulfoxide solution, adjusted to pH 8, in 4 days or until Ellman's reaction was arbitrarily assigned to be consistent with known values for both proteins lined by Adler and Wagner (25). Chemical shifts were somewhat arbitrarily assigned to be consistent with known values for both proteins and water. The chemical shift of water was assigned at 4.88 ppm at 3 °C and 4.66 ppm at 23 °C. Unless otherwise noted, the chemical shifts are reported for 3 °C at pH 6.8.

Structural Constraints—The HN-C′H coupling constants were measured directly from the DQF-COSY spectra without any correction for line width. Dihedral angle constraints were obtained for six φ angles (Fig. 1) using the following method. It was assumed that the observed coupling constants were accurate to ±1.5 Hz and that the maximum coupling should not exceed 10.5 Hz. A computer program was written that used the Karplus equation as parameterized by Pardi et al. (26) to calculate all values of φ that would match the coupling constant within experimental error. The maximum and minimum values of φ were then used as dihedral angle constraints. This method produces a continuous function that nearly matches the values used by Kline et al. (27) for 8.0 and 10 Hz. The method sets a range of −77° to −53° for φ angle of Glu374, which had a coupling constant of 4.9 Hz. The coupling constants of both Leu348 and Phe350, 7.4 and 7.5 Hz, respectively, were also used for calculating restraints using the same program. A restriction was added to the calculation that rejected any positive values of φ. The intraresidue HN to C′H NOE peak for both of these residues indicated an interproton distance of approximately 3 Å, which is inconsistent with a positive value of φ. The resulting constraints were −165° ≤ φ ≤ −75° for Phe350 and 1° wider for Leu348.

NOE peak intensities were quantified from a NOESY spectrum of the double mutant (H381G,M388L). The recycle delay between pulses was 3.6 s. A 100-ms mixing time was used to limit the artifacts caused by spin diffusion. The spectrum was processed using 75° shifted sine bell. The base line was flattened with a 5th order polynomial subroutine in both directions on the fully transformed data set using the Bruker processing software UXNMR. This subroutine has an automatic selection of baseline points. Correction along the F2 axis was performed in eight uneven sections to minimize the distortions caused by the dispersive water signal.

The distance constraints were calculated from the peak intensity. A correction factor was included, which controlled for variation of peak width, based on the relative width of a resonance in the F2 direction compared with the width of the peaks used for calibration. Eight well-resolved methylene pairs were used to calculate the scaling factor between the peak volumes and the target distances. The variation in intensity between these peaks was 20%. The standard equation for translating peak volumes into target distances was modified so that the distances were lengthened to compensate for any experimental uncer-
tainty. First, the volume of each peak was divided by two to compensate for variations in both peak width and intensity. Also, it was assumed the volumes of the weaker peaks were less accurate than the more intense ones. This uncertainty was handled mathematically by using ½ instead of ½ power to calculate distances from peak volumes. The final affect on the experimental data was that the target distances of 2.2, 2.5, 3.0, and 4.0 Å were lengthened to 2.5, 3.1, 3.8, and 4.6 Å respectively (an upper limit of 5.0 Å was used for all observed NOEs). The accuracy of the modified distance function was verified by examining the target values both intra- and sequential HN to C+H NOEs. The ranges of distances calculated from NOE peaks were roughly 20% larger than the expected values. The calculated distances obtained from 100-ns NOESY spectrum ranged from 2.5 to 4.7 Å. An additional 0.7 Å was added to all NOEs involving methyl groups (27). Lower bound constraints were set to the Van der Waal’s contact radii.

No stereospecific assignments were performed for the methylene protons. If a pair of NOE peaks was observed between two protons of a diastereomeric pair to a third proton, the weaker NOE peak was used to calculate the distance constraint. When only one NOE was observed, the distance constraint was referenced back to the nearest heavy atom that was equidistant from both protons. The distance constraint was lengthened by the fixed distance between the protons and the heavy atom. In general, the preliminary structures were not used to further interpret the spectra due to the inherent uncertainty involved with these techniques. However, some of the NOEs to the C+H positions of the two phenylalanines were stereoselectively assigned when the preliminary structures indicated a separation of greater than 7 Å between protons that had NOEs to same Phe C+H.

The issue spin diffusion was not explicitly addressed during the preparation of the constraints. NOEs that involved a pair of methylene protons were examined for evidence of spin diffusion. We specifically looked at pairs where one of two NOEs was very intense and the other could be a source of spin diffusion. In all cases, the inter-proton distance calculated for the weaker NOE was confirmed by an independent structural constraint.

Structure Calculations—All structures were generated using the distance geometry package, DGG (version 2.2.0), of Insight II, which was generously provided by BIOSYM, Inc. After smoothing the bond matrix using the triangle inequality, the individual structures were embedded and subject to a maximum of 10,000 steps of simulated annealing at a maximum temperature of 200 K. The resultant structures were minimized by a conjugate gradient using a maximum of 250 steps. 34 out of 100 structures were selected for making comparisons. These structures had residual penalty functions below the mean structure and had no NOE violations greater than 0.2 Å. All peptides examined had the same fold. The structures with higher residual penalty functions showed greater deviations from the mean.

RESULTS

New Mutants in the EGF4 Domain—It was previously shown (10) that there was a drastic loss of activity when alanine was substituted for the following residues in EGF4 of TM: Glu357 and Tyr358 in the B-loop and Phe376 in the C-loop. Working with a construct made from EGF4–6 of TM, TMiE4–6-M388L, it was shown that full function was restored by conservative substitution of the two aromatic residues. Thus, both Tyr358 → Phe and Phe376 → Tyr were fully active. Conservative substitution of Glu357 partially restored function. Glu357 → Gin had only 1 ± 2% of native activity, and Glu357 → Asp restored 48 ± 15% of the function (p < 0.05). Thus, the functional assay indicates there is little tolerance for variation at Glu357, and major changes are not tolerated at positions Tyr358 and Phe376.

A fourth set of substitutions was made in an attempt to increase activity. His381 is only found in human TM. Glycine is found in this position in both mouse and hamster TM. Comparison with homologous proteins indicates that there was probably a β-turn at this position. Experiments demonstrated that His381 → Gly doubled the specific activity of human TMiE4–6-M388L. The double mutant, H381G, M388L, is 400% more active than soluble TM analogs with the native human sequence. The substitution of His381 → Ala (10) had no affect. His381 → Pro actually caused a slight decrease in activity (60 ± 28%).

NMR Structural Studies—Three separate peptides were synthesized based on the sequence of TM-EGF4; the A-loop, residues 345–361; the B-loop, residues 352–371; and the C-loop, residues 371–389. Each peptide was cyclized by forming a single native disulfide bond. For the A- and the B-loops, 2-amino butyric acid was substituted for the unpaired cysteine. The NMR results indicated that the A- and B-loops did not fold into compact structures. The two-dimensional NMR spectra of these peptides showed that the chemical shifts of the protons were similar to the random coil values for the same amino acids (28). Also, the J_{iH·iH} coupling constants were all within 1 Hz of the random coil values with the single exception of 11 Hz for Val371.

Exhaustive analysis of the NOESY spectra of both the A- and B-loops revealed only eight NOEs that connected residues separated by at least one amino acid; two bridged across the disulfide bonds, one connected the HNs of Asn355 to Tyr358 and the remaining five involved residues separated by a single amino acid. Only two of the eight NOEs involved backbone-backbone interactions. Structure calculations utilizing the combined experimental constraints from both peptides failed to converge on a unique structure.

The solution structures of three other loops from EGF-like proteins were examined: the C-loop of TM-EGF5, the C-loop of transforming growth factor-α, and the B-loop of human urokinase EGF domain. Inspection of the two-dimensional TOCSY spectra indicated that there was little chemical shift dispersion of the protons beyond what was expected based on the random coil values (28). No further analysis was attempted. The results from the C-loop of TM-EGF5 were confirmed by a recent report (8). Although the peptide is unfolded in solution, it forms a unique conformation upon binding to the anion exosite of thrombin.

Two-dimensional spectra of the peptide based on the C-loop of TM-EGF4 indicated that the peptide did form a compact structure. In particular, the chemical shifts of the amide protons ranged from 7.5 to 9.3 ppm (Table II). The comparable range for the same protons in unfolded peptides would be 8.2–8.4 (28). To probe further the relationship between structure and function, a total of four peptides derived from the C-loop of EGF4 were synthesized (Table I). Although the peptides folded into compact structures, the isolated peptides had no measurable effect on modulating the activation of protein C by thrombin when tested alone as a cofactor for thrombin or as a competitive inhibitor of the action of thrombomodulin on thrombin, even at concentrations as high as 5 μM. The activity measurements shown in Table I were performed by incorporating the sequence of each of these peptides back into a truncated but fully active form of thrombomodulin containing the fourth, fifth, and sixth EGF-like domains, TMiE4–6-M388L. The activity, as a percentage of the specific activity of the TM native sequence, ranged from 46 to 10%. The most detailed structural work was performed on the double mutant, H381G, M388L, which this represents the most active sequence. The two-dimensional spectra indicated that all four peptides had the same overall fold (see below for details).

Proton Chemical Shift Assignments—The proton chemical shifts for the native sequence appear in Table II. Selected residues from the other peptides are also listed. Assignments have been made for 113 of 115 slowly exchanging and nonexchanging protons in the native sequence. Assignments of the remaining protons were probably obscured by degenerate protons within the same residue. Fig. 1 shows the sequential NOEs that were used in making the assignments.

Structure of the Peptide Based on the Double Mutant, H381G,M388L—The structure reported here is obtained from

G. Rumennik and D. R. Light, unpublished data.
TABLE II
Chemical shifts in ppm of assigned protons of C-loop of the 4th EGF domain of thrombomodulin at p H 6.8, 3 °C
Selected residues are included from the other peptides listed in Table I.

| Residue       | H1  | H2  | H3  | H4  | Others     |
|---------------|-----|-----|-----|-----|------------|
| Val371        | 3.65| 2.03| 0.85| 0.83|            |
| Cys372        | 4.62| 2.92| 2.54|     |            |
| Ala373        | 4.00| 1.31|     |     |            |
| Glu374        | 4.04| 1.89| 1.85|     | 2.18       |
| Gly375        | 3.97| 3.30|     |     |            |
| Phe376        | 4.76| 2.52|     |     |            |
| Ala377        | 4.77| 1.22|     |     |            |
| Pro378        | 4.69| 2.27| 1.60|     | 1.91, 1.86 |
| Ile379        | 4.04| 1.67| 1.24| 0.76| 2.18, 1.62 |
| Pro380        | 3.92| 1.71| 1.99| 1.81| 1.9, 1.37  |
| His381        | 4.40| 3.29| 2.92| 2.90| 1.62, 1.59 |
| Glu382        | 4.63| 1.63| 1.97| 1.91| 1.94, 1.75 |
| Pro383        | 4.26|     | 1.82| 1.75| 1.94, 1.75 |
| His384        | 4.48| 3.05| 1.26| 1.03| 1.94, 1.75 |
| Arg385        | 4.33| 1.47| 1.21| 1.03| 1.94, 1.75 |
| Cys386        | 5.05| 2.87| 2.77|     |            |
| Glu387        | 4.67| 1.90| 1.86| 1.82| 1.86, 1.59 |
| Met388        | 3.75| 1.61| 1.95| 1.95| 1.94, 1.75 |
| Phe389        | 4.28| 2.80| 2.80|     |           |
| Mso388        | 3.74| 1.73| 2.48| 2.23|           |
| Leu388        | 3.70| 1.24|     |     | 1.24       |
| Gly389        | 4.07| 3.63|     |     | 0.56, 0.34 |

a From [Met388]Mso peptide.
b From Met388 Leu peptide.
c From H381G.M388L peptide.
d From H381G.M388L peptide.

*From Met388Mso peptide. *From Met388 Leu peptide. *From H381G.M388L peptide.

**Fig. 1.** A graph that shows the sequential NOEs that were used in making the proton assignments. The right column identifies the type of NOEs. D, signifies an NOE between the X proton in the first residue to the Y proton in the second residue. The NOEs are listed for sequential residues except where noted. The widths of the black lines are proportional to the NOE intensity. For the three prolines, NOEs to the C'H were used in place of NOEs to the CN. The symbol (I) is used to denote NOEs that either could not be observed for either practical or theoretical reasons, such as potential overlap or residues that are missing a proton. The bottom line of the graph lists some of the observed J(§Hs, Hs) coupling constants. Symbols are used as follows: c indicates a J(§Hs, Hs) of 5.5 Hz, b indicates a J(§Hs, Hs) of 10.0 Hz, a indicates a J(§Hs, Hs) of 15.0 Hz. The use of the symbols a and b do not imply that the residue is part of an a-helix or a b-sheet.

The peptide based on the double mutant at 3 °C (p H 6.8). A total of 273 NOEs was assigned for the double mutant. Approximately 21 of the 48 intraresidue NOEs and 20 of the 85 sequential NOEs had upper bound distances that were greater than the distance allowed by the covalent geometry. This left a total 172 useful NOEs, which had the following distribution: 27 intraresidue, 65 sequential, 26 medium range (I to I + 2) and 45 long range (I to I + 5) NOEs. Torsion angle constraints for the angle o were derived from eight HN-C'H coupling constants. This gives an average of 9.5 useful constraints per residue. The structure itself is shown in Fig. 2A.

The root mean square deviations of the well determined backbone atoms is 0.6 Å to the average structure and 0.9 Å for the pair-wise interactions (Fig. 2B). This figure excludes N-terminal Val371 because there is almost no structural information for this residue. Root mean square deviation for all heavy atoms to average structure is 1.3 Å (1.8 Å for the pair-wise interactions.) The side chain conformation has been accurately determined for Phe376, Ile379, His384, and Gln387. Constraints on the protein backbone also determine the locations of the side chains of Ala373, Ala377, Pro380, and Pro382. Less information is available for the other side chains.

The Structure of the C-loop of EGF4 in Solution—The final structure is depicted in Fig. 2A. The molecule forms a loop-like structure that is bracketed on either side by the two well defined b-turns. The first turn, which extends through residues Ala373, Glu374, Gly375, and Phe376, is a type II b-turn. There is a well defined hydrophobic pocket surrounding residue Phe376. The phenylalanine side chain is flanked by, and has NOEs to, Cys372, Ala373, Ala377, Cys386, and Leu388 (Met388 in the native sequence). These residues limit the solvent exposure of the aromatic ring to the outer edge.

The second bend includes residues Ile379, Pro380, Gly381, and Glu382. Both type I and II b-turns are compatible with the experimental data for the double mutant. The intensity of intraresidue NOEs between the HN and C'Hs of Gly381 would clearly resolve the ambiguity if there were stereospecific assignments available for the C'Hs. Unfortunately, the stereospecific assignments could not be determined in a reliable fashion. The other three peptides, including the native sequence, all have histidine at third position of this turn, and all three exhibit a type I b-turn. This b-turn is part of a five-residue insertion, Pro380 to His384, in the sequence of this EGF-like domain (Table III).

The b-turn is stabilized by hydrophobic interactions that are centered on Ile379. Only the outer edges of the methyl groups are exposed to the solvent. The side chain of Ile379 residue is covered by the methylene side chains of residues Pro380, Glu382, Arg385, and Gln387. The close interaction between these side chains probably adds to the stability of the protein.

A third, less well defined type I b-turn appears between residues Glu382, Pro383, His385, and Arg385. The chain itself forms a right angle turn through this bend. This geometry distorts the conformation of Glu382 and weakens the hydrogen bond between the CO of Glu382 to the NH of Arg385.

The three prolines, Pro378, Pro380, and Pro382, all have trans peptide bonds. There was no detectable amount of any folded
species that contained a cis peptide bond. All three prolines are located at bends in the protein backbone. The prolines are all involved in delineating the second \( \beta \)-turn. This is part of a five-residue insertion in the sequence of the C-loop (Table III).

There are three hydrogen bonds that can be easily identified in the structure. Two of the hydrogen bonds are found in the \( \beta \)-turns: CO Ala\textsuperscript{373} to HN Phe\textsuperscript{376} and CO Ile\textsuperscript{379} to Glu\textsuperscript{382}. A third hydrogen bond is found between the CO of Ala\textsuperscript{377} to the HN of Gln\textsuperscript{387}. This hydrogen bond appears where the protein backbone crosses back upon itself (Fig. 2A). Other potential hydrogen bonds may exist in this structure but cannot be identified given the resolution of structures. Indeed, the peptide appears to have only a few internal hydrogen bonds that could contribute to the stability of the structure.

Structural Homology—Table III lists the sequences of the C-loops from the five homologous proteins or domains whose atomic coordinates were available. The sequence alignment is based upon the structural comparison depicted in Fig. 3. 10 residues were identified as playing the same role in all six peptides. They are in TM-EGF4 (the peptide used in this study), the first \( \beta \)-turn, the next two residues after it (Cys\textsuperscript{372},Pro\textsuperscript{378}), the second cysteine and its proceeding residue (Arg\textsuperscript{385} and Cys\textsuperscript{386}), and finally Met\textsuperscript{388}, which covers one side of Phe\textsuperscript{376}. The pair-wise root mean square deviation of the backbone atoms for these 10 residues is 1.0 Å. The comparable figure for the uncertainty in our own structures is 0.7 Å. Furthermore, the structural homology extends to the orientation of the peptide bonds and the C\textsuperscript{\beta}. It should be noted that the conserved bend is a type II \( \beta \)-turn and with an X\textsuperscript{-}X\textsuperscript{-}Gly-(Phe/Tyr) sequence; the first two positions are variables, the third position is glycine, and the fourth position is either phenylalanine or tyrosine. This sequence is found in many proteins that are homologous to EGF.

The sequence of the structurally conserved residues can be described as Cys-X-X-Gly-(Phe/Tyr)-X...X...X-Cys...X. The first gap contains between one and six residues; the second gap contains either one or two. The structural similarity in the last position is surprising. The charge, hydrophobicity, and size of this residue vary between the proteins listed in Table III. Also, TM-EGF4 and FXa-C, the two longest C-loops, exhibit a one-residue deletion prior to this residue. The side chain of this residue is in close contact with the conserved aromatic residue in the first \( \beta \)-turn and appears to limit the ring’s exposure to solvent. This interaction must be important to the stability of the protein, since the interaction is maintained despite the large variation of sequence in this position. In fact, there is little conservation of the sequence for six out of ten conserved positions (Table III).

All six C-loops exhibit a second chain reversal shown on the left side of Fig. 3. Four of the proteins share the same overall
Temperature Studies—A series of one-dimensional NMR spectra was obtained for the double mutant from 8 to 65 °C (data not shown). The peptide retained sufficient structure to protect the amide protons from exchange with solvent up to 40 °C. The HN resonances disappeared in a cooperative fashion above 40 °C and were nearly invisible at 50 °C. The chemical shifts of all resonances changed in a continuous manner from 8 to 65 °C. This indicates that there was a rapid exchange between the folded and unfolded conformations. Once the temperature was lowered from 65 °C back to 8 °C, the protein completely refolded and showed no sign of any degradation.

A more detailed study of temperature effects was carried out using two-dimensional NMR. A NOESY spectrum of the double mutant was collected at 23 °C and visually compared to the corresponding data obtained at 3 °C. Although the intensities of the cross-peaks were attenuated at the higher temperature, there was no evidence of any detectable change in conformation.

DQF-COSY spectra were obtained for all four peptides (Table I) at both 3 and 23 °C. The similarity of chemical shifts again indicated that the conformation remained intact. There were, however, consistent changes in C-H of residues 376–378 and 386–388 (Fig. 4). These residues form an antiparallel structure. Some of the chemical shift perturbations extended to the side chains that participate in the hydrophobic pocket surrounding residue Phe376 (Fig. 4). The elevated temperature had little effect on the aliphatic protons located near β-turns, indicating that the β-turns were stable at the higher temperature and there was no global unfolding of the peptide.

Single Site Mutations—Table I shows the sequences and relative activities of the four peptides used in this study. A preliminary examination of the chemical shift data and the 200-ms NOESY spectra indicated that all four peptides had the same structure. These results were unexpected, since there is a 40-fold difference in activity when these sequences are incorporated back into the parent protein (Table I).

A detailed comparison was made between the 200-ms NOESY spectra (3 °C) of the double mutant and the peptide with oxidized Met388 ([Met388]Mso). These peptides represent the most and least active sequences. Of the original 213 assigned for the double mutant, only two NOEs were found missing for [Met388]Mso peptide. Both NOEs involved the backbone protons of Phe389. The remaining 17 NOEs to Phe389 were found in both peptides. The [Met388]Mso peptide also had some new NOEs not found for the double mutant. Observation of the new NOEs probably stemmed from the higher sample concentration used for the [Met388]Mso peptide. NOESY spectra of the native peptide and the Met388 → Leu were also very similar to corresponding data for the double mutant. Each spectrum contains nearly the same set of NOEs for the side chains of residues 388 and 389. However, lower sample concentration precluded a more quantitative examination of the data.

Similarity in the structures of the four peptides is also demonstrated by comparing the chemical shifts. The substitution of His383 → Gly did not significantly affect (+0.08 ppm) the chemical shifts of the protons beyond a 5-Å radius of the site of the modification. The oxidation of the Met388 also had very minor effects.

It is worth noting that the spectra of the oxidized [Met388]Mso peptide indicated the presence of two closely related peptides, even though the compound was pure, as judged by high performance liquid chromatography and mass spectrometry. At 3 °C, there was measurable splitting of all the resonances of the methionine sulfoxide, Mso388, and in the backbone protons of Gln387 and Phe389. Within accuracy of the data, the intensity of both sets of peaks was the same. At 23 °C, this splitting became more pronounced and affected additional residues in the hydrophobic pocket around Phe376. The mono-oxidation of S5 of methionine introduces a chiral center at the sulfoxide. Since the peptide was made with a synthetically prepared derivative of methionine, it contained a racemic mixture of both R and S forms of methionine sulfoxide at the S5 position. These results indicate that each enantiomer has a slightly different conformation.

**DISCUSSION**

**Protein Folding and Stability**—The structure presented here (Fig. 2A) is in some ways typical for small peptides. It contains...
only a few hydrogen bonds, and its structure is dominated by \(\beta\)-turns. However, the protein is uncharacteristically flat and extended. Phe^{376} is the only residue that has close interactions with more than one other strand of the protein. The small amount of interior volume in the protein is defined by the packing of the side chains and not by the backbone. This protein lacks the tightly coiled structure that characterizes the other small proteins that appear in the protein data bank. It is hard to judge whether this flat structure will be found in other isolated peptides that contain a single disulfide loop. However, five of the six loops examined in this study failed to form a compact structure.

As discussed in the results section, the two \(\beta\)-turns are stabilized by the formation of hydrophobic pockets. It is worth noting that the side chains of two other hydrophobic residues, Val^{371} and Phe^{389}, do not interact with the hydrophobic pocket surrounding Phe^{376}, even though their backbone residues are close to this pocket (Fig. 2A). A possible explanation for these results can be found by examining the structure of the homologous proteins. If Val^{371} and Phe^{389} are compared to homologous residues in other EGF-like proteins (Table III), the corresponding amino acids do not participate in stabilizing this hydrophobic cluster. The results imply that the structural constraints that control folding of the intact protein are somehow encoded in the isolated C-loop.

Finally, the structure of this peptide has some interesting implications for protein folding. It clearly shows that a subdomain of a larger protein can act as an autonomous folding unit. The temperature shift data implies that the two \(\beta\)-turns are more stable structures and, therefore, may guide the folding of the peptide. However, this peptide is small enough such that it could find the correct structure by random search of conformational space, and folding may take place in a single cooperative step. The folded C-loop may then act as a template that guides the subsequent steps in protein folding. The results suggest that the folding of the backbone and the side chains can take place concurrently. Overall folding of the protein may consist of a series of precise events with intermediates that have a well-defined structure.

Structure and Function—When the sequences of the four peptides are incorporated back into TM, there is a 40-fold difference in activity (Table I and Ref. 13). The NMR work presented here indicates that there is no detectable difference in activity (Table I and Ref. 13). The NMR work indicates that there is no detectable difference in activity (Table I and Ref. 13). The NMR work indicates that there is no detectable difference in activity (Table I and Ref. 13). The NMR work indicates that there is no detectable difference in activity (Table I and Ref. 13). The NMR work indicates that there is no detectable difference in activity (Table I and Ref. 13).