Protein Science (1995), *4*:1683–1695. Cambridge University Press. Printed in the USA.

Synthesis, activity, and preliminary structure of the fourth EGF-like domain of thrombomodulin

DAVID P. MEININGER, MICHAEL J. HUNTER, AND ELIZABETH A. KOMIVES
Department of Chemistry and Biochemistry, University of California–San Diego, La Jolla, California 92093-0601

(RECEIVED February 3, 1995; ACCEPTED June 16, 1995)

Abstract

The fourth EGF-like domain of thrombomodulin (TM4), residues E346–F389 in the TM sequence, has been synthesized. Refolding of the synthetic product under redox conditions gave a single major product. The disulfide bonding pattern of the folded, oxidized domain was (1–3, 2–4, 5–6), which is the same as that found in EGF protein. TM4 was tested for TM anticoagulant activity because deletion and substitution mutagenesis experiments have shown that the fourth EGF-like domain of TM is essential for TM cofactor activity. TM4 showed no TM-like activity in two assay systems, both for inhibition of fibrin clot formation, and for cofactor activity in thrombin activation of protein C.

A preliminary structure of TM4 was determined by 2D ^{1}H NMR from 519 NOE-derived distance constraints. Distance geometry calculations yielded a single convergent structure. The structure resembles the structure of EGF and other known EGF-like domains but has some key differences. The central two-stranded β-sheet is conserved despite the differences in the number of amino acids in the loops. The C-terminal loop formed by the disulfide bond between C372 and C386 in TM4 is five amino acids longer than the analogous loop between C33 and C42 of EGF protein. This loop appears to have a different fold in TM4 than in EGF protein. The loop forms the two outside strands of a broken, irregular tri-stranded β-sheet, and amino acids H384–F389 lie between the two strands forming the middle strand of the sheet. Thus, although the C-terminus of EGF protein forms one of the outside strands of a tri-stranded antiparallel sheet, the C-terminus of TM4 forms the inside strand of an irregular tri-stranded parallel–antiparallel sheet. The residues D349, E357, and E374, which were shown to be critical for cofactor activity by alanine scanning mutagenesis, all lie in a patch near the C-terminal loop, and are solvent accessible. The other critical residues, Y358 and F376, are largely buried and appear to play essential structural rather than functional roles.

Keywords: anticoagulant; fibrinogen; NMR; peptide synthesis; protein C; thrombin

Thrombomodulin (TM) is an endothelial cell surface glycoprotein that binds tightly to thrombin. TM has a direct anticoagulant effect by inhibiting the thrombin cleavage of fibrinogen to form a fibrin clot. More importantly, TM is the essential cofactor that promotes thrombin cleavage of protein C (Esmon, 1989). Activated protein C inactivates factors Va and VIIIa and thus shuts down coagulation and maintains hemostasis.

Although TM is a 70-kDa membrane-bound protein, the fragment of TM comprised of only EGF-like domains four and five has almost full TM anticoagulant activity (Hayashi et al., 1990; White et al., 1995). The fifth domain is the major site of thrombin binding and is a potent inhibitor of fibrin clot formation, but it is devoid of cofactor activity (Hunter & Komives, 1995). Thus, the fourth EGF-like domain of TM is required for cofactor activity (Lentz et al., 1993; Nagashima et al., 1993). The fragment of TM comprised of the fourth and fifth EGF-like domains binds to thrombin with an apparent K_d of 25 nM, whereas the fifth domain alone binds thrombin with a K_d of 2 μM. Because the fragment containing the fourth domain binds much more tightly to thrombin, it seems likely that the fourth domain does interact directly with thrombin (Hunter & Komives, 1995; White et al., 1995). Whether the fourth domain alone binds to thrombin has not been determined.

The structures of EGF (Montelione et al., 1987; Hommel et al., 1992), TGFα (Moy et al., 1993), the N-terminal EGF domain (EGFI) of coagulation factor IX (Baron et al., 1992), and EGFI of coagulation factor X (Selander-Sunnerhagen et al., 1992) have been determined by NMR spectroscopy. The structures of the EGF domain from E-selectin (Graves et al., 1994) and the EGFI and EGFII domain from coagulation factor X (Padmanabhan et al., 1993) have been determined by X-ray crystallography. Despite the divergence of sequence and func-

Reprint requests to: Elizabeth A. Komives, Department of Chemistry and Biochemistry, University of California–San Diego, 9500 Gilman Drive, La Jolla, California 92093-0601; e-mail: ekomives@ucsd.edu.

tion, all of the EGF-like domains listed have the same disulfide bonding pattern, with bonds between the first and third cysteines, the second and fourth cysteines, and the fifth and sixth cysteines (Savage et al., 1973; Hojrup & Magnusson, 1987; Huang et al., 1991). The structures of the aforementioned EGF-like domains are all similar (Campbell & Bork, 1993 and references therein). The main structural motif is a two-stranded antiparallel β-sheet, the midpoint of which is near the third and fourth cysteines. The amino acids between the third and fourth cysteines form a major loop at one end of the β-sheet and the amino acids between the fifth and sixth cysteines form a major loop at the other end of the sheet. The short loop between the second and third cysteines lies close to the major C-terminal loop. In general, the N-terminus and the C-terminus are quite flexible. How these structural features function in the protein-protein interactions and calcium binding of these different domains is just beginning to be understood.

A comparison of the sequences of various EGF-like domains shows that the number of amino acids between the cysteines is somewhat variable. TM4 and EGFII from coagulation factor X have similar numbers of amino acids between the cysteines, but they differ from the other EGF-like domains for which structures are known. The loop between the second and third cysteines is three amino acids shorter in TM4 and two amino acids shorter in factor X EGFII than in the other domains. Even more striking is the five-amino acid insertion in the loop between the fifth and sixth cysteines of TM4. The factor X EGFII has a four-amino acid insertion in this same loop, and it has been reported to have a structure that is similar to EGF. The fifth EGF-like domain of TM has a three-amino acid insertion, but the C-terminal loop of this domain also adopts a classic EGF-like tri-stranded antiparallel β-sheet structure when bound to thrombin (Srinivasan et al., 1994). Thus, despite the differences in lengths of the loops, the structures of all the EGF-like domains determined so far are similar.

The experiments presented here were designed to probe both the function and the structure of the fourth EGF-like domain of TM. Because deletion of the fourth domain abolishes TM cofactor activity, it seemed important to establish whether the fourth domain by itself would have cofactor activity. Knowledge of the structure of TM4 would resolve the question of whether the amino acids shown to be essential by alanine scanning mutagenesis play a structural or functional role. Finally, it was of interest to determine the structural affects of the five-amino acid insertion in the C-terminal loop of TM4.

Results

Synthesis and disulfide bond formation in TM4

Initial attempts to synthesize the 44-amino acid TM4 peptide resulted in poor overall yields and no isolable product of the correct molecular weight. Use of hydroxyazabenzotriazole in place of hydroxybenzotriazole as the activator as well as use of the preactivated pentafluorophenyl esters resulted in much improved yields. HPLC purification of the reduced, crude peptide gave one major product that was shown to have the correct molecular weight (5,073 g/mol). The purified, reduced peptide was then oxidized in a mixture of reduced and oxidized glutathione affording one major product (Fig. 1). The amino acid sequence of the domain is given in Table 1 with the residues of the synthetic domain numbered 1–44. The corresponding residue numbers (346–389) from full-length human TM are also given (Wen et al., 1987).

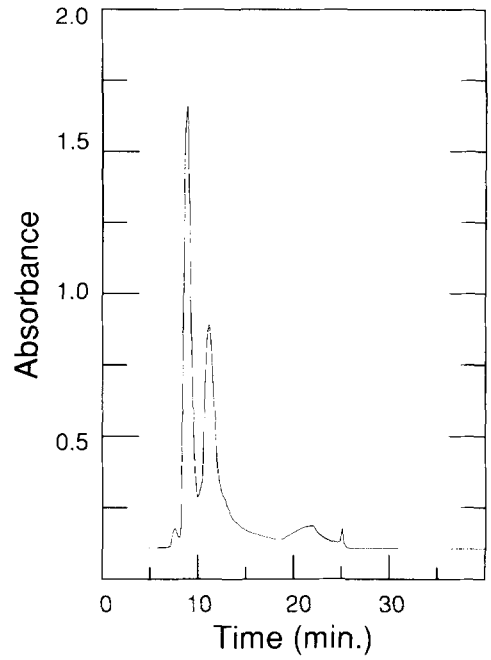

Fig. 1. Reverse-phase HPLC purification of the oxidized TM4 peptide. The oxidized, folded product was isolated by loading the oxidation solution in two 500-mL portions onto an HPLC column directly through the "A" buffer pump. The column was a Waters DeltaPak C18 column (19 × 300 mm) and the gradient consisted of 0.1% TFA to 50% acetonitrile over 1 h. The major product eluted approximately 8 min after the gradient was held at 30% acetonitrile.

The partial reduction technique of Gray (1993a,b) was used to map the disulfide bonds in TM4. This procedure involves the partial reduction of the peptide at low pH with Tris-carboxyethylphosphine. Partially reduced peptides are then purified, alkylated, and repurified. In the case of TM4, all eight possible partially reduced peptides were isolated and purified. N-terminal sequence analysis of the alkylated peptides gave the disulfide bonding pattern (1-3, 2-4, 5-6), which is the same as that found in the EGF protein (Fig. 2). Confirmation of the disulfide bonds was also obtained from $d_{\beta\beta}$ NOE crosspeaks that were observed between residues 6 and 15, 11 and 25, and 27 and 41.

Activities of TM4

TM has been shown to have two anticoagulant activities. TM directly inhibits thrombin cleavage of fibrinogen, thus preventing clot formation. TM also acts as a cofactor for thrombin, stimulating the thrombin cleavage of protein C, which subsequently shuts down the coagulation pathway. TM4 was assayed for both direct anticoagulant activity and TM cofactor activity. No significant inhibition of clot formation was observed at TM4 concentrations up to 175 μM. TM4 was assayed both as an inhibitor of TM cofactor activity in the protein C activation assay, and as a replacement for TM in this assay. Inhibition of TM cofactor activity was used as a second measure of whether TM4 could bind to thrombin, but no inhibition was observed

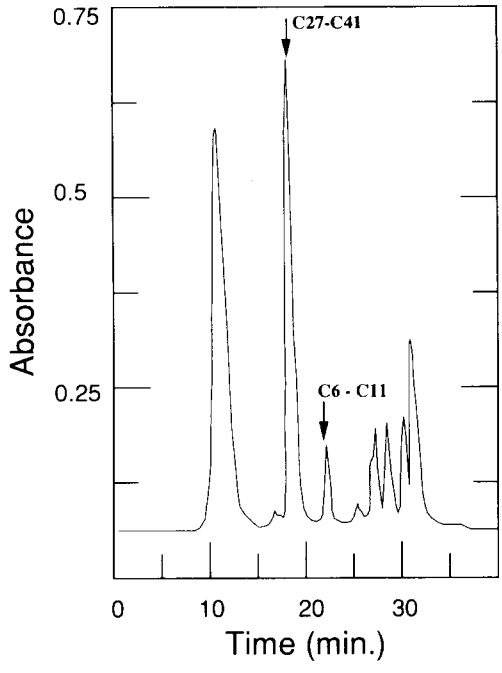

Fig. 2. Reverse-phase HPLC separation of the products of partial reduction of TM4 by TCEP. The first peak observed represented peptide that remained fully oxidized and eluted approximately 10 min after the start of the extended gradient from 10% to 40% acetonitrile. All eight possible partial reduction products were isolated and the sequences of two peaks were determined by N-terminal sequence. In the sequence data, acetamidylated cysteines elute as glutamic acid, and cystines are absent from the analysis. Peaks 2 and 3 gave sufficient data to establish the disulfide bonding pattern as (1-3, 2-4, 5-6). The N-terminal sequence of peak #2 was EPVDP-FRAN-EYQ-QPLNQTSYL V"E"AEGFAPIPHEPHR"E"end, and the N-terminal sequence of peak #3 was: EPVDP"E"FRAN-EYQ"E"QPLNQT-YL-end.

at concentrations up to 350 μM. Finally, assays in which TM4 replaced TM were performed to determine whether the isolated TM4 domain might have cofactor activity. No cofactor activity was detectable up to concentrations of 350 μM.

Structure calculations

Nearly complete spin system identifications for the nonlabile hydrogen atoms in TM4 were obtained and are listed in Table 1. In addition, labile side-chain proton resonance assignments were obtained for R8, N10, Q16, N19, R40, and Q42 and are also listed in Table 1. An example of the sequential connectivities from C6 to Q16 and from L18 to A32 is shown on a plot of the fingerprint region of the 300-ms jump-return NOESY spectrum in Figure 3. An initial set of structures was calculated from 273 distance constraints derived from NOEs observed in a jump-return NOESY data set collected in H_2O, which could be assigned solely and unambiguously on the basis of resonance frequency. Of the 50 calculated structures, the structure with the smallest RMS deviation (RMSD) to the average structure was then used as a model to resolve ambiguities in the assignments of the NOEs. Additional constraints from the jump-return NOESY data sets and constraints created from 100-, 200-, and 300-ms NOESY data sets collected in D_2O were assigned iteratively over the course of several rounds of structure calculations. A total of 519 NOE-based distance constraints was used for the final calculation. These consisted of 192 intraresidue, 117 sequential, 21 medium-range (less than five residues apart), and 190 long-range (five or more residues apart) restraints (Wüthrich, 1986). A final set of 50 embedded structures was calculated starting from the TM4 structure modeled on the EGF protein using these 519 distance constraints.

A residue versus residue plot of the final set of NOE-based constraints is shown in Figure 4. Two regions of apparent β-sheet can be discerned, one in the N-terminal half of the molecule from residues 14 to 26 with a turn at residues 19–21, and the other in the C-terminal half of the molecule from residues 26 to 40 with turns at residues 29–30 and 38–39. In particular, d_{NN}, $d_{N\alpha}$, and/or $d_{\alpha\alpha}$ crosspeaks were observed in the N-terminal half of the molecule between residues 14 and 26, residues 15 and 25, residues 15 and 26, residues 16 and 26, residues 17 and 24, residues 18 and 23, residues 18 and 24, and residues 19 and 23 (Fig. 5). Slowly exchanging amide protons were observed for residues 16, 18, 24, and 26, and moderately slowly exchanging amide protons were observed for residues 14, 15, 23, and 25. These data were consistent with the presence of the central two-stranded antiparallel β-sheet as has been found in other known EGF-like domains. In the C-terminal half of the molecule, d_{NN}, $d_{N\alpha}$, and/or $d_{\alpha\alpha}$ crosspeaks between residues 26 and 39, and residues 27 and 40 were observed. The amide proton for residue number 40 was observed to be slowly exchanging, and moderately slowly exchanging amide protons were observed for residues 27 and 39. These data suggest a short region in which the two β-strands run parallel. In addition, d_{NN}, $d_{N\alpha}$, and/or $d_{\alpha\alpha}$ crosspeaks were observed between residues 31 and 43, residues 31 and 44, residues 32 and 42, residues 32 and 43, residues 32 and 44, residues 33 and 42, residues 34 and 41, and residues 34 and 42, indicating a region of antiparallel β-sheet between residues 31 and 34, and residues 41 and 44. Moderately slowly exchanging amide protons were observed for residues 34 and 44, supporting the presence of this region of antiparallel β-sheet.

The overall fold of the structures produced during later rounds of calculations did not differ substantially from the fold of the structures generated by the first round of calculations, but the additional constraints improved the convergence of the calculated structures, especially the side-chain positions. Of the 50 final calculated structures, 41 exhibited a DGII error function value less than 0.5 after the simulated annealing phase of optimization and were selected for further analysis. The DGII structures were then energy minimized in three steps in Discover, with 2 of the 41 structures discarded prior to minimization due to excessively high starting energies. The 39 minimized structures were then subjected to a final phase of X-PLOR-based refinement consisting of one round of simulated annealing followed by energy minimization. The structures generated by the two refinement procedures had identical folds, and the majority of the structures had no NOE violations above 0.35 Å. The structures resulting from both refinement procedures superimposed with similar RMSDs to the average. The Discover refinement produced structures in which the nonbonded interactions were minimized, but bond angles and lengths differed significantly from ideality as assessed by the initial energy evaluation in X-PLOR. The X-PLOR refinement produced structures that had slightly higher energies associated with nonbonded interactions, but that also exhibited bond angles and lengths much

Table 1. *^{1}H NMR chemical shifts for TM4 at pH 5.2 and 25 °C*

Residue[a]							
In TM4	In TM		Chemical shift index[b]	HN	Hα	Hβ	Others
1	346	Glu	0	–[c]	4.26	1.99, 2.08	γCH2 2.37[d]
2	347	Pro	−1	–	4.27	1.81, 2.23	γCH2 1.91[d]; δCH2 3.47, 3.62
3	348	Val	0	8.27	3.99	1.99	γCH3 0.90[d]
4	349	Asp	1	8.20	4.87	2.60, 2.69	
5	350	Pro	0	–	4.40	2.37[d]	γCH2 1.90[d]; δCH2 3.88, 4.01
6	351	Cys	−1	8.40	4.02	2.90, 3.15	
7	352	Phe	−1	7.77	4.28	3.02, 3.29	2,6H 7.24; 3,5H 7.31; 4H –
8	353	Arg	−1	8.04	3.98	1.51, 1.83	γCH2 1.03, 1.14; δCH2 2.98[d]; εNH 7.09
9	354	Ala	−1	6.93	3.79	0.74	
10	355	Asn	−1	8.02	4.53	2.64, 2.74	NH2 6.46, 7.26
11	356	Cys	0	7.64	4.69	2.21, 3.58	
12	357	Glu	−1	8.82	3.94	1.90, 2.33	γCH2 2.15[d]
13	358	Tyr	1	9.02	4.83	2.95, 3.67	2,6H 6.99; 3,5H 6.61
14	359	Gln	0	7.30	4.41	1.51, 1.83	γCH2 1.75[d]; NH2 –
15	360	Cys	1	8.65	5.16	2.81[d]	
16	361	Gln	1	9.51	5.06	1.96, 2.17	γCH2 2.26[d]; NH2 6.80, 7.32
17	362	Pro	0	–	4.50	2.21[d]	γCH2 2.04[d]; δCH2 3.87[d]
18	363	Leu	1	8.62	4.49	1.36[d]	γCH 1.41; δCH3 0.78[d]
19	364	Asn	−1	8.20	4.47	3.17, 3.27	NH2 6.89, 7.66
20	365	Gln	−1	8.59	4.14	2.48[d]	γCH2 2.22[d]; NH2 –
21	366	Thr	1	8.25	4.60	4.43	γCH3 1.43
22	367	Ser	1	7.96	4.91	3.66, 3.79	
23	368	Tyr	1	8.10	5.19	2.44, 3.02	2,6H 6.73; 3,5H 6.68
24	369	Leu	1	9.15	4.64	1.42, 1.72	γCH 1.49; δCH3 0.85, 0.95
25	370	Cys	1	8.95	5.64	2.82, 3.69	
26	371	Val	1	8.81	4.60	2.15	γCH3 0.84, 0.96
27	372	Cys	1	8.93	4.78	2.47, 3.02	
28	373	Ala	−1	7.69	3.67	0.61	
29	374	Glu	−1	8.18	4.12	1.92, 2.30	γCH2 2.00[d]
30	375	Gly	1	8.88	3.39, 4.11		
31	376	Phe	1	8.43	4.98	2.48, 3.26	2,6H 6.75; 3,5H 7.23; 4H 7.10
32	377	Ala	1	9.19	4.85	1.34	
33	378	Pro	1	–	5.07	2.46[d]	γCH2 1.73[d]; δCH2 3.59, 3.81
34	379	Ile	0	7.63	4.01	1.85	γCH2 and γCH3 0.75; δCH3 0.82
35	380	Pro	1	–	4.75	2.10[d]	γCH2 1.78[d]; δCH2 3.64, 4.07
36	381	His	0	8.16	4.58	3.30, 3.41	2H 7.15; 4H 8.35
37	382	Glu	1	7.70	4.79	1.67, 1.85	
38	383	Pro	0	–	4.50	1.89, 2.28	γCH2 2.01[d]
39	384	His	1	8.70	4.80	3.33[d]	2H 7.06; 4H 8.26
40	385	Arg	1	8.41	4.79	1.66, 1.92	γCH2 1.14, 1.29 δCH2 3.07, 3.15; εH 7.25
41	386	Cys	1	8.42	5.24	2.74, 2.89	
42	387	Gln	1	9.39	4.92	1.93, 2.08	γCH2 2.25[d]; NH2 6.81.7.45
43	388	Met	−1	8.70	3.72	1.50, 1.68	γCH2 1.93[d]; εCH3 –
44	389	Phe	−1	7.62	4.36	2.87[d]	2,6H 7.09; 3,4,5H –

[a] Numbers of the residues are given in the left column for the individual TM4 domain. Numbers of the residues within the whole thrombomodulin molecule are given in the right column for reference.
[b] The chemical shift index was determined by the method of Wishart et al. (1992).
[c] – Indicates that the resonance could not be assigned or was not observed.
[d] Protons with degenerate chemical shifts.

closer to ideality (Table 2). The structures refined in Discover had large energies associated with improper torsion angles and these were minimized to much lower values by X-PLOR. Twenty of the 39 structures refined in X-PLOR had no NOE violations greater than 0.35 Å. Twelve of these structures had energies less than 120 kcal·mol^{-1} and these were collected as the 12 best structures. The backbone RMSD for these 12 best structures after the X-PLOR refinement was 1.93 Å and the RMSD for all heavy atoms was 2.46 Å. The five N-terminal residues were not well constrained by the NMR data. Removal of these residues from the RMSD calculations resulted in a backbone atom RMSD of 1.21 Å and an RMSD for all heavy atoms of 1.81 Å. A ste-

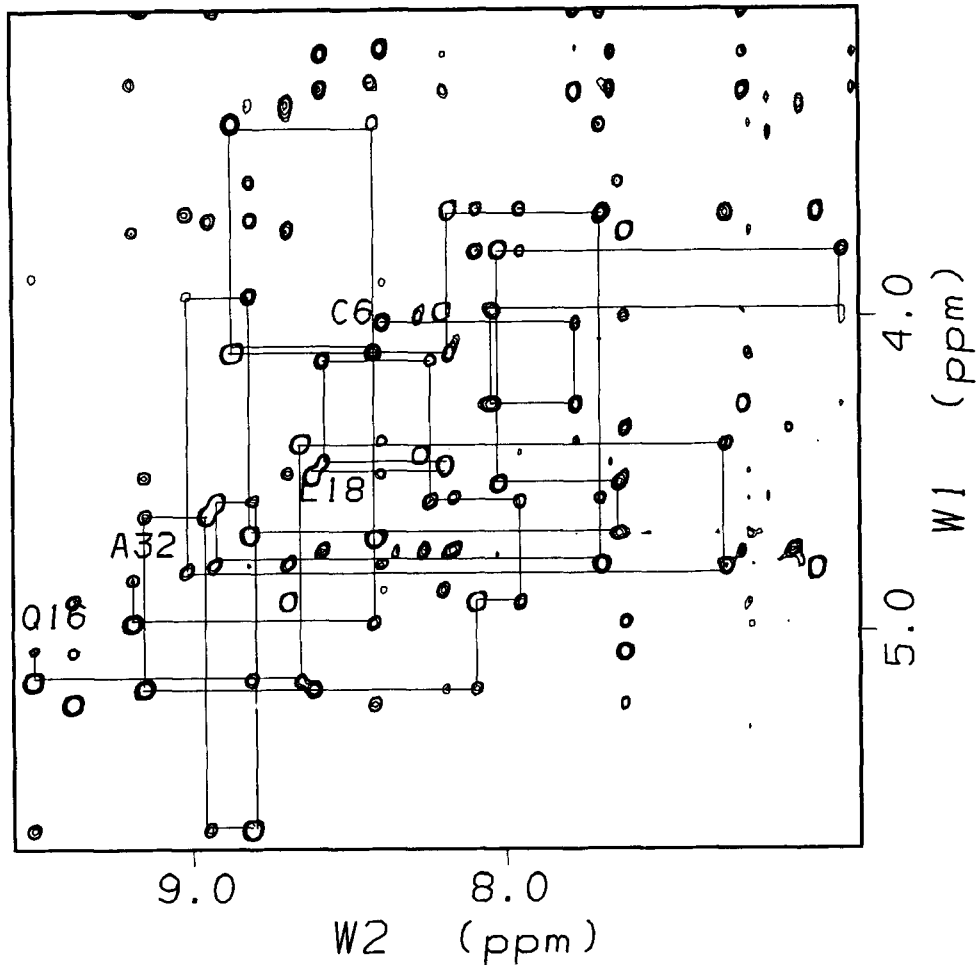

Fig. 3. Fingerprint region of the 300-ms jump-return NOESY spectrum of TM4. The sequential connectivities are shown for residues C6–Q16 and for residues L18–A32. The sequence of the peptide throughout this stretch is CFRANCEYQCQPLN QTSYLCVCAEGFA.

reo view of these 12 best structures superimposed on the backbone atoms using the program developed by Diamond (1992) is shown in Figure 6 and Kinemage 1.

Discussion

Activity of TM4

TM4 showed no activity in either assay system in which it was tested. This was a surprising result because the peptide corresponding to C27–F44 (residues C372–F389 of the TM sequence) was an inhibitor of thrombin clotting of fibrinogen that doubled the clotting time at a concentration of 550 μM (Lougheed et al., 1995). Although the C-terminal loop peptide was a fairly weak inhibitor, it was surprising that the presence of the rest of the domain diminished the observed inhibition. We do not know whether the conformation of the loop alone differs from its conformation within the whole domain. It is possible that the observed inhibition by the peptide resulted from nonspecific hydrophobic interactions. Because the fourth domain is required for TM cofactor activity, it was also surprising that the isolated domain did not show any cofactor activity, even at very high concentrations. TM is active at a concentration of 0.1 pM, whereas the domain showed no activity at concentrations up to 350 μM. In other experiments, TM4 was mixed with a peptide corresponding to the fifth domain of TM and no cofactor activity was observed for the mixture. The two-domain fragment comprised of the fourth and fifth domains binds to thrombin 200-fold more tightly than the fifth domain alone, suggesting that the fourth domain is capable of binding to thrombin in the context of this two-domain fragment (White et al., 1995).

Comparison of the fourth and fifth domains of TM

The fingerprint region of the NOESY spectrum of TM4 showed good dispersion of the $d_{\alpha N}$ crosspeaks with narrow line widths, indicating that TM4 is a well-structured molecule. These results were in sharp contrast to those observed for the fifth domain of TM, which exhibits both a significant amount of spectral overlap and rather broad line widths in the fingerprint region (M. Hunter, unpubl.). The fifth domain also appears to tolerate different disulfide-bonding patterns and the isomer with the highest thrombin-binding affinity is non-EGF-like (Hunter & Komives, 1995). Thus, it is interesting to speculate on the roles

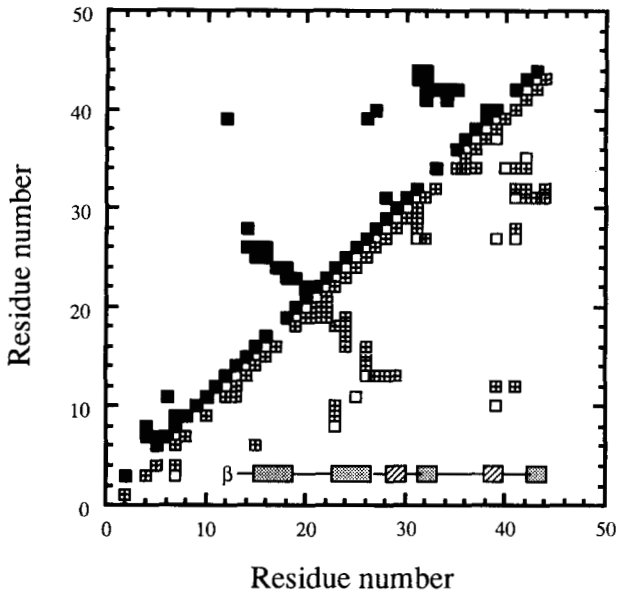

Fig. 4. Survey of the NOE-derived distance constraints for TM4 in H_2O at pH 5.2 and 25 °C. Each axis is labeled with the amino acid sequence numbers. Solid, crossed, and open squares at coordinate *i*, *j* indicate that one or more backbone-to-backbone, backbone-to-side chain, or side-chain to side-chain NOEs, respectively, have been identified between the atoms of residues *i* and *j*. The NOEs involving only backbone resonances are shown on the top of the diagonal (■), and the NOEs involving main chain-to-side chain resonances (⊞) and side chain-to-side chain resonances (□) are shown on the bottom of the diagonal. Regions of antiparallel β-sheet are indicated by shaded bars at the bottom of the graph and regions of parallel β-sheet are indicated by striped bars.

of each of these essential EGF-like domains in the thrombin–TM interaction. The fifth domain is not well structured but binds tightly to thrombin. Transferred NOE studies of the C-terminal loop of the fifth domain show that this part of the domain becomes structured upon binding to thrombin, suggesting that the fifth domain may bind to thrombin by an induced fit mechanism (Srinivasan et al., 1994). The fourth domain, which is required for cofactor activity, appears to adopt a well-defined structure but does not bind to thrombin. Based on the apparent rigidity of the fourth domain, it seems unlikely that its structure would change when connected to the fifth domain. The fifth domain may open up a binding site for the fourth domain on thrombin. The observation that cofactor activity correlates with a conformational change far from the active site of thrombin supports this hypothesis (Ye et al., 1991), but the fact that no cofactor activity is observed from mixtures of the two domains suggests that the fourth domain cannot bind if it is not connected to the fifth domain. These speculations await clarification that could be offered by the structure of the TM fragment containing both the fourth and fifth EGF-like domains, which is in progress.

Preliminary structure of TM4

TM4 formed a single major product upon oxidative refolding in the presence of glutathione. This product, presumably the thermodynamically most stable, had the disulfide-bonding pattern that is expected for EGF-like domains (1–3, 2–4, 5–6). The fingerprint region of the TOCSY spectrum shows that all 37 of the expected $d_{\alpha N}$ crosspeaks were present and well dispersed, indicating a high degree of structure within the domain. In addition, the total number of NOE cross peaks was large, resulting in preliminary structure calculations utilizing 11.8 NOE-based distance constraints per residue, or 12.6 NOE-based distance constraints per residue if the N-terminal five residues and NOEs involving these residues were ignored. The structure calculations gave a well-converged family of structures despite the absence of dihedral angle constraints, stereospecific assignments, and hydrogen bond constraints.

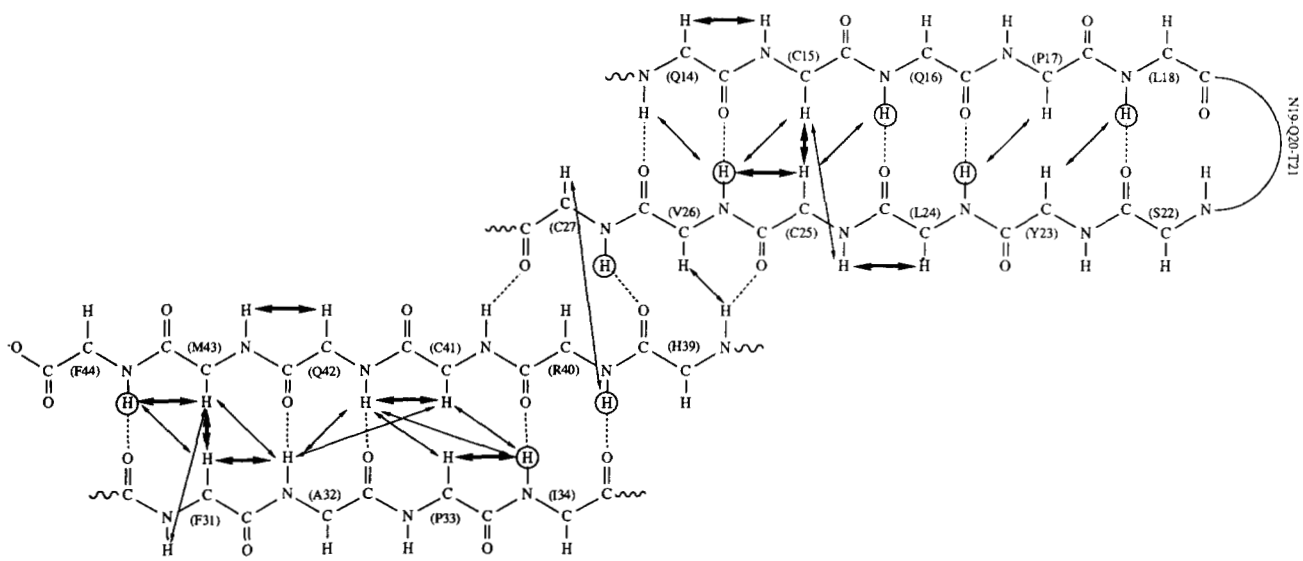

Fig. 5. Schematic diagram of the β-sheets found in TM4. NOE connectivities are depicted by double-ended arrows with darker arrows indicating strong NOEs. Slowly exchanging protons are circled. The amide-to-carbonyl distances that were measured to be close enough to be hydrogen bonded are shown with dashed lines. Very few constraints were observed to the α-proton of R40 due to the fact that this proton was nearly on resonance with the water signal.

Table 2. *Statistical information for TM4 structures at various stages of refinement*

Statistic	After DGII	After Discover	After X-PLOR	12 best structures
Number of structures	41	39	35	12
Avg. error function[a]	0.26	N/A	N/A	N/A
Highest	0.45	N/A	N/A	N/A
Lowest	0.15	N/A	N/A	N/A
Avg. energy (kcal·mol^{-1})	5,560	284	123	112
Avg. number of NOE violations per structure >0.3 Å	0.46	1.9	1.2	0.75
Highest	1	4	4	3
Lowest	0	0	0	0
Avg. RMSD[b] (Å) backbone atoms	1.33	1.70	1.40	1.21
Avg. RMSD (Å) heavy atoms	1.94	2.38	1.99	1.81
Avg. RMSD (Å) N-terminal subdomain backbone atoms	0.91	1.17	1.40	0.97
Avg. RMSD (Å) N-terminal subdomain heavy atoms	1.54	1.88	1.76	1.66
Avg. RMSD (Å) C-terminal subdomain backbone atoms	0.96	1.10	0.91	0.87
Avg. RMSD (Å) C-terminal subdomain heavy atoms	1.68	1.87	1.53	1.47
Avg. RMSD (°) from ideal bond angles	—	—	0.691 ± 0.034	0.679 ± 0.009
Avg. RMSD from (Å) ideal bond lengths	—	—	0.0027 ± 0.0003	0.0026 ± 0.0001
Avg. RMSD (°) from ideal impropers	—	—	0.264 ± 0.049	0.244 ± 0.017
Avg. RMSD (Å) for NOE constraints	—	—	0.031 ± 0.005	0.026 ± 0.002

[a] The error function is a parameter evaluated during the DGII calculations and was therefore not determined for the minimized structures. Energies were evaluated using the Discover software.

[b] All RMSDs were calculated relative to the mean coordinates using the best superposition program developed by Diamond (1992) and did not include the five disordered N-terminal residues.

The overall architecture of TM4 is similar to that observed for other EGF-like domains. The central portion of the molecule, from residues 14 to 26, forms a two-stranded β-sheet, and the stretch of amino acids between 19 and 21 forms a β-hairpin of similar structure to that observed in other EGF-like domains (Fig. 7; Kinemage 2). Confirmation of the presence of β structure was obtained from the chemical shift index, which indicated a region of β-sheet in the N-terminus from amino acids C13 to L18 and from T21 to C27 (Wishart et al., 1992). The N-terminus of TM4 contains a loop between residues 4 and 11, which is different from the helical segment observed in EGF protein (Hommel et al., 1992). This loop appears to form because of a close interaction between residues D4 and R8, indicated by several side-chain to side-chain NOEs between these residues, which

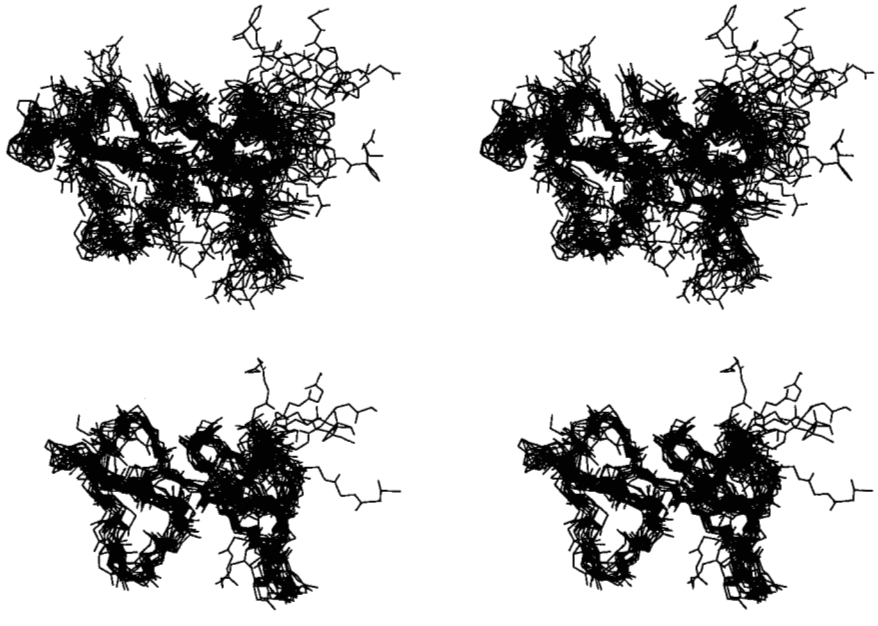

Fig. 6. Stereo view of the 12 best structures obtained for TM4 after the final X-PLOR-based refinement superposed on the backbone atoms by the best superposition program of Diamond (1992). The structures were those that had total energies less than 120 kcal·mol^{-1} and no NOE violations greater than 0.35 Å. The top view is all heavy atoms; the bottom view is backbone atoms.

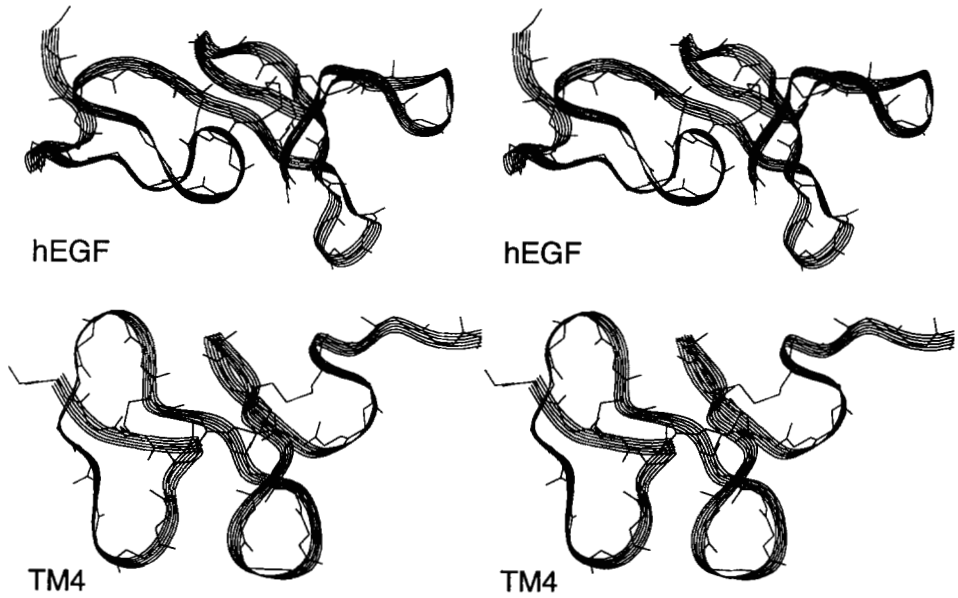

Fig. 7. Stereo views of ribbon diagrams of the structure of human EGF protein (3EGF from the most current Brookhaven Protein Data Bank release) and the lowest energy structure of TM4. The ribbon is superimposed on the backbone of each molecule and the cysteine disulfide bonds are shown as thin black lines.

may be due to the presence of a salt bridge between the charged side chains. The effect of the loop is to bring the N-terminus into close proximity to the residues E12–Q14 (E357–Q359) in TM4.

Structure of the C-terminal loop

Superposition of TM4 and EGF protein on the cysteines shows that the N-terminal subdomains superimpose well despite differences in loop sizes (Kinemage 2). The central two-stranded β-sheet also superimposes well, but large differences are seen between the C-terminal subdomains of the two proteins. The architecture of the C-terminal loop of TM4 is different from that found in EGF protein and in other previously studied EGF-like domains. A model of TM4 based on the structure of EGF protein (Montelione et al., 1987; Hommel et al., 1992) that was generated by homology modeling had a tri-stranded β-sheet structure very similar to EGF, but many unambiguously assigned NOEs were inconsistent with this model, especially for the C-terminal subdomain.

In TM4, the overall structure of the C-terminal loop is still a tri-stranded β-sheet. In fact, the chemical shift index indicates that all of the residues from G30 to Q42 are in β-strands. Backbone-to-backbone NOEs between residues V26 and H39 and between C27 and R40 support the fact that the strand extending from C25 to C27 and the strand extending from H39 to C41 (the C-terminal strand) run parallel to each other. The strand extending from H39 to F44 runs antiparallel to the strand extending from I34 to F31. The C-terminal strand in TM4 lies directly between the other two strands in the C-terminal loop, whereas the C-terminal strand in EGF protein lies on the outside of the other two strands in the C-terminal loop for most of its length (Fig. 7; Kinemage 2). Thus, the architecture of the C-terminal portion of TM4 may be described as a broken, irregular parallel-antiparallel sheet with one crossover rather than as a simple antiparallel sheet, as is observed in EGF protein and other EGF-like domains. The C-terminal sheet in TM4 is somewhat irregular with β-turns between residues A28 and G30 and P35 and E37, which partially distort the structure.

The middle strand of the C-terminal tri-stranded β-sheet in TM4 contains the amino acids C-terminal to C41, which are the same amino acids that connect the fourth domain with the fifth domain in TM. Thus, the structure predicts an intimate and rigid contact between the fourth and fifth domains in TM. The hypothesis of a rigid contact region between the two domains is supported by the fact that both domains are required for TM cofactor activity, whereas mixtures of the two have no cofactor activity. As more studies of tandem small domains are reported, it is becoming clear that the original beads-on-a-string hypothesis seldom holds (Smith et al., 1994). It is interesting to speculate that the novel architecture of TM4 has the function of fixing the residues C-terminal to the last cysteine in TM4 in order to spatially stabilize the interaction between the fourth and fifth EGF-like domains of TM that is probably required for cofactor activity. Perhaps this motif will be found in other EGF-like domains that function in tandem.

Location of the essential residues identified by alanine scanning

Several residues in the fourth EGF-like domain of TM have been shown to be essential by alanine scanning mutagenesis (Nagashima et al., 1993). In particular, mutation of D4, E12, Y13, E29, and F31 (D349, E357, Y358, E374, and F376) to alanine resulted in at least a 75% loss of activity. Also, oxidation of M43 (M388) to the sulfoxide results in a 90% loss of activity (Glaser et al., 1992). Figure 8 (and Kinemage 3) shows that these residues all lie in a "patch" in the domain formed by the N-terminus of the domain, the short loop between C11 and C15,

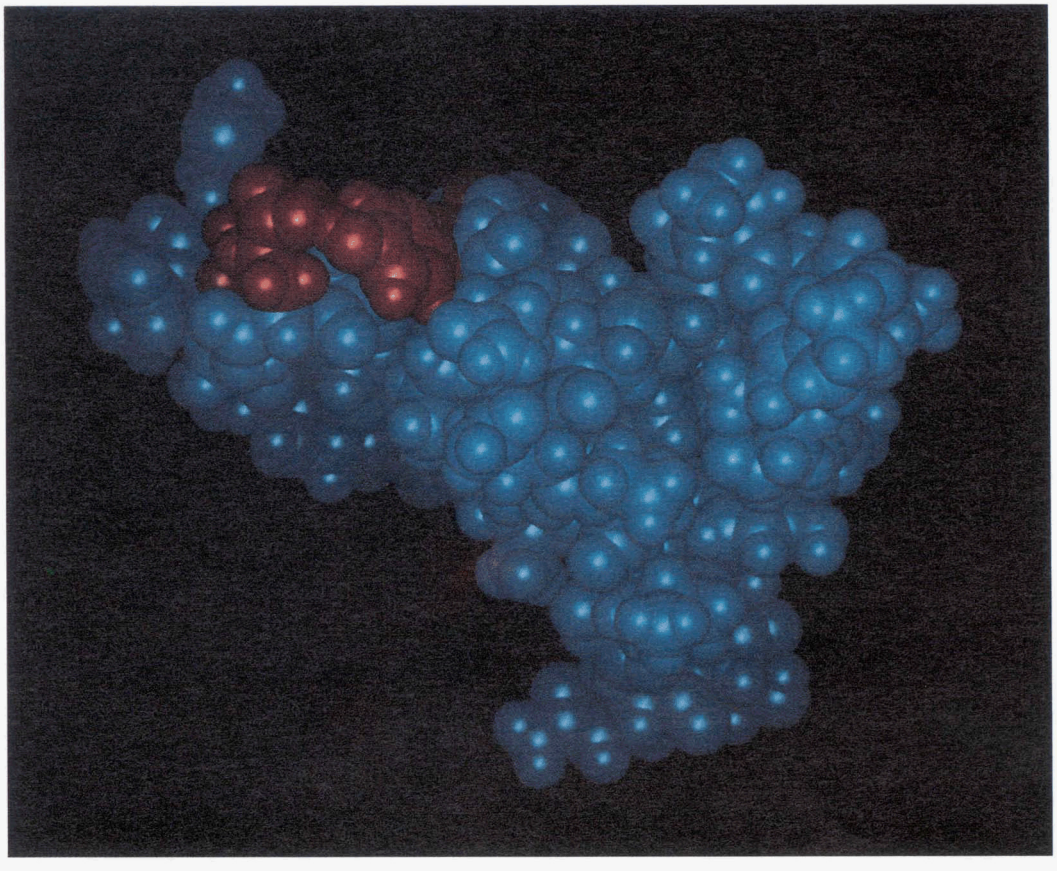

Fig. 8. CPK structure of TM4 showing the residues known to be essential for TM cofactor activity as shown by alanine scanning mutagenesis. The TM4 domain is shown in blue, and the essential residues, D349, E357, Y358, E374, and F376, are shown in red.

and the first strand of the C-terminal loop, C27–F31. It is striking that the residues that form the patch are all either hydrophobic or acidic, both characteristics of molecules that bind in the anion-binding exosite of thrombin. The essential residues, D4, E12, and E29 all contact very few other residues, are solvent accessible, and are close to one another within the patch. The whole patch is acidic except for one edge that is bound by M43 (M388) and is hydrophobic.

At least two of the essential residues, Y13 (Y358) and F31 (F376), most likely play structural roles because these residues appear buried in the structure and make contacts to several other residues. In particular, Y13 exhibits NOEs to V26, C27, A28, and E29 and appears to cement the position of the N-terminal subdomain with respect to the C-terminal subdomain. F31 exhibits NOEs to C27, A28, E29, C41, Q42, M43, and F44 and appears to cement the core of the C-terminal tri-stranded β-sheet. M43 and F44 appear to form a hydrophobic protrusion at the edge of the patch containing the acidic residues D4, E12, and E29. This patch of essential residues will most likely be in close proximity to the fifth domain, the first cysteine of which follows directly F44. This structure predicts an intimate contact between the C-terminus of the fourth domain and the N-terminus of the fifth domain and suggests that a much larger surface may form in the TM fragment containing both the fourth and fifth domains that could interact with thrombin.

Materials and methods

Peptide synthesis

TM4, the 44-amino acid peptide corresponding to the sequence in Figure 1, was prepared by standard solid-phase peptide synthesis on a Milligen 9050 peptide synthesizer. The amino acids used were obtained from Milligen (Burlington, Massachusetts), were protected on the amine with fluorenylmethoxycarbonyl, and were preactivated as the pentafluorophenyl esters of the carboxylic acids. TM4 proved to be particularly difficult to produce. Several syntheses were attempted using hydroxybenzotriazole obtained from Milligen (Burlington, Massachusetts) as the coupling reagent, but the full-length peptide could not be isolated from these syntheses. Eventually, two separate syntheses were successfully carried out using slightly different methodologies and using hydroxyazabenzotriazole, again from Milligen, as the coupling reagent. For the first synthesis, KA solid support (Novabiochem) was used. Phenylalanine was manually coupled to the resin as described in the Milligen Peptide Synthesizer Manual to a substitution level of 0.087 meq/g. The amino acid side-chain protecting groups for the first synthesis were as follows: Asn, Cys, and Gln-Trt; Asp and Glu-OtBu; Ser, Thr, and Tyr-tBu: His-Boc; Arg-Mtr. For the second synthesis, PEG-PS-phenylalanine solid support with a substitution level of

0.160 meq/g obtained from Milligen was used and the amino acid side-chain protecting groups were the same as those in the first synthesis with the exception of Arg for which Arg(Pmc)-OPfp was used. Fmoc-L-Arg(Pmc)-OPfp was synthesized from Arg(Pmc)-OH and pentafluorophenol as described previously using ethylacetate as the solvent and with care taken to keep the reaction solution ice cold at all times during the synthesis (Kisfaludy et al., 1973). Dimethylformamide (glass distilled) from EM Sciences (distributed by VWR, Los Angeles, California) was used as the solvent for both syntheses.

The resin was washed extensively with dichloromethane upon completion of synthesis and dried overnight under vacuum. The peptide was cleaved from the resin and the side-chain protecting groups were removed by treatment of the resin with a mixture of 90% trifluoroacetic acid (TFA), 5% thioanisole, 3% ethanedithiol, and 2% anisole for 12 h for the first synthesis (with Arg-Mtr) and for 4 h for the second synthesis (with Arg-Pmc). The resin was removed by filtration through glass wool and the peptide was isolated by ether precipitation at −20 °C overnight. Crude peptide was collected on a cintered glass funnel, washed with cold ether, dissolved in 5% ammonium acetate buffer, pH 7.5, and lyophilized. The yield of crude peptide was 640 mg for the first synthesis (72.5% of the theoretical yield) and 750 mg for the second synthesis (66.0% of the theoretical yield).

Fully reduced peptide was prepared by dissolving 25-mg portions of crude peptide in 20 mL of 100 mM Tris-HCl, pH 8.0 buffer containing 20 mM dithiothreitol under nitrogen for 8–12 h. Purification of 100-mg portions of crude peptide by reverse-phase HPLC on a Waters DeltaPak C18 column (19 × 300 mm) using a gradient of argon-sparged 0.1% TFA to 50% acetonitrile over 1 h produced a yield of 25–30 mg pure, fully reduced peptide. The major HPLC product eluted as a single, sharp peak that was collected and lyophilized and shown to be of the correct molecular weight by mass spectrometry and the correct amino acid composition by amino acid analysis. The minor, later eluting product appeared on further analysis to be a mixture of species and was not characterized further.

The purified, reduced peptide was then refolded by slow air oxidation in a redox buffer. The peptide (25 mg) was dissolved in 1 L of 0.5% ammonium acetate buffer, pH 5.0, containing 1 mM oxidized glutathione and 9 mM reduced glutathione and the pH of the resulting solution was then slowly brought to 8.3 by the dropwise addition of 5% NH_4OH with vigorous stirring. The pH of the oxidation solution was maintained at 8 to 8.3 by daily addition of 5% NH_4OH for 4–5 days to ensure complete oxidation of the peptide. The oxidized, folded product was isolated by loading the oxidation solution in two 500-mL portions onto an HPLC column directly through the "A" buffer pump. The column, gradient, and buffers were the same as those above except that the buffers were not sparged with argon and the gradient was held at 30% acetonitrile for better product separation. One major oxidation species was produced that eluted earlier than a mixture of unresolved minor species. After purification to >95% homogeneity, the overall yield for the major oxidation product was approximately 10% of the crude.

Determination of disulfide bonds

TM4 contains six cysteine residues that form three disulfide bonds. The disulfide bonding patterns of the peptides were determined by a partial reduction method (Gray, 1993a). The disulfide bonding pattern of the major oxidation species was mapped using Tris-carboxyethylphosphine (TCEP) as a partial reducing agent (Gray, 1993b). Approximately 500 μg of purified, oxidized peptide was repurified on a Vydak C18 HPLC column (4.6 × 250 mm) at a flow rate of 1 mL/min, using a gradient of 0.1% TFA to 10% acetonitrile over 10 min and then to 40% acetonitrile over 1 h. Detection was at 280 nm. The TCEP solution was prepared by dissolving 32.4 mg TCEP and 250 mg citric acid in 5 mL MilliQ H_2O and adjusting the pH to 2.5 by the dropwise addition of 1 N NaOH. The repurified peptide was then dissolved in 700 μL 0.1% TFA. The same volume of TCEP solution was added to the peptide solution and the mixture was vortexed. After exactly 1 h, the partially reduced peptides were separated by HPLC using the same column and gradient described for the repurification of the peptide. Eight distinct reaction products were detected and collected separately. The free thiols in the reaction products were acetamidylated by forcibly injecting 500 μL of each HPLC peak into a Falcon tube containing 400 μL of a 0.5 M Tris-acetate buffer, pH 8.0, 2 mM EDTA, and 2.2 M iodoacetamide with continuous vortexing during the injection. Each of the eight alkylation reactions was quenched after exactly 1 min by acidification with 800 μL of 0.5 M citric acid. The acetylated products were purified separately using the above HPLC conditions.

Finally, the purified, alkylated products were characterized by N-terminal sequencing. The PTH derivative of reduced, acetamidylated cysteines elutes at the same retention time as the PTH derivative of glutamic acid and was easily distinguished from cystine.

TM activity assays

Peptides were dissolved in H_2O at a concentration of 1–2 mM and the pH was adjusted if necessary. The peptide solutions were divided into small portions and stored at −20 °C. The concentration of each peptide solution was determined by quantitative amino acid analysis using norleucine as the standard.

The direct inhibition of thrombin-induced clot formation by TM has been previously demonstrated by measuring the time it takes for thrombin to induce clot formation at different concentrations of TM in a solution containing fibrinogen. The experimental details have been described previously (Lougheed et al., 1995). TM4 was used in place of TM in this assay. The assay was carried out by incubating several different concentrations of TM4 with 6.6 U of human thrombin (a generous gift of Dr. John Fenton). After 10 min, fibrinogen was added and the time for the clot to form was measured using a stop watch. Concentrations of TM4 up to 175 μM were assayed.

TM cofactor activity, which results in the thrombin-dependent production of activated protein C, was measured using an assay in which thrombin and TM were incubated with protein C and the resulting activated protein C was then assayed with a chromogenic substrate (Stearns et al., 1989). Cofactor activity of TM4 was measured by incubating various concentrations of each peptide with human thrombin (0.4 μg/mL, 0.875 U/mL, 11 nM) in TBS containing BSA (1 mg/mL) and $CaCl_2$ (5 mM). After 10 min, human protein C (7.25 μg/mL, 120 nM, Hematologic Technologies) was added and the mixture was incubated for 20 min. The thrombin activity was quenched with heparin-antithrombin III (80 ng/mL and 220 ng/mL, respectively) and activated protein C activity was assayed by monitoring the re-

lease of *p*-nitroaniline at 405 nm from a chromogenic substrate, ChromozymPCa (Boehringer-Mannheim) or S-2366 (Chromogenix), at a concentration of 0.74 mg/mL. TM4 showed no activity in this assay up to a concentration of 350 μM. TM4 was also tested for thrombin binding as indicated by inhibition of protein C activation in the presence of rabbit TM (20 ng/mL, 0.29 nM) using the same assay, but no inhibition was seen up to concentrations of TM4 of 350 μM.

NMR spectroscopy

The folded, purified TM4 peptide was not soluble at pH 4.0–4.5, so care was taken to keep the pH of the samples above 5.0. Both 90% H_2O/10% D_2O and 100% D_2O samples were prepared. For the former, samples of TM4 were dissolved in 450 μL of H_2O, the pH was adjusted to 5.2 by the addition of microliter amounts of 1 N NaOH, and 50 μL of D_2O was added (Cambridge Isotope Labs, Cambridge, Massachusetts) for the deuterium lock. For D_2O samples, TM4 was dissolved in 500 μL of D_2O and the pH was adjusted to 5.2 (uncorrected for isotope effects) by the addition of microliter amounts of 1 N NaOH. After lyophilization, D_2O samples were redissolved in 500 μL of "100%" D_2O and the pH was checked to verify that it was still approximately 5.2. For both samples types, the final peptide concentration was verified to be between 1.5 and 2 mM by amino acid analysis.

Spectra were acquired on a Bruker AMX spectrometer operating at 500.13 MHz. All 2D data sets were processed on a Silicon Graphics workstation using FELIX 2.30 software (Biosym Technologies, San Diego, California). All spectra were recorded in the phase-sensitive mode by using time-proportioned phase incrementation (Marion & Wuthrich, 1983). The carrier was set on the water resonance and the water resonance was suppressed by selective irradiation by using the decoupler during a relaxation delay period of 1.0 s. To further aid in water suppression, all H_2O time domain data were processed with a zero frequency subtraction routine in t_2 prior to Fourier transformation (Marion et al., 1989), which was provided by Dr. Ed Wang in the laboratory of Dr. Julie Feigon.

^{1}H NMR assignments of TM4

To obtain ^{1}H resonance assignments clean-TOCSY (Griesinger et al., 1988), DQF-COSY (Braunschweiler et al., 1983), jump return-NOESY (Plateau & Guéron, 1982), and NOESY (Kumar et al., 1980; Macura & Ernst, 1980) experiments were recorded. Clean-TOCSY spectra were recorded in H_2O with a mixing time of 75 ms at 15, 25, 35, and 45 °C, with mixing times of 35 and 50 ms at 25 °C, and in D_2O with a mixing time of 35 ms at 25 °C. DQF-COSY spectra were recorded in H_2O and in D_2O at 25 °C. Jump return-NOESY spectra were recorded in H_2O with mixing times of 100, 200, and 300 ms, and with a mixing time of 300 ms at 15 °C. NOESY spectra were recorded in D_2O with mixing times of 100, 200, and 300 ms at 25 °C. A short Hahn-echo period was added at the end of the clean-TOCSY and NOESY pulse sequences to improve the quality of the baseline (Davis, 1989). Clean-TOCSY experiments were carried out with sweep widths of 12,500 Hz in t_2 and 6,250 Hz in t_1, and with the DIPSI-2 relaxation compensated mixing sequence to optimize Hartman–Hahn transfer and suppress ROESY transfer. DQF-COSY, jump return-NOESY, and NOESY experiments were acquired with sweep widths of 6,250 in both t_2 and t_1. DQF-COSY experiments were recorded with 1,024 t_1 increments each with 112–144 scans of 4,096 complex data points and were zero filled to 8,192 points in t_2 and t_1 and multiplied by a skewed sinebell-squared window function prior to Fourier transformation. Other data sets were recorded with 450–512 t_1 increments each with 80–96 scans of 2,048 complex data points and were zero filled to 1,024 points in t_1 and multiplied by a skewed sinebell-squared window function prior to Fourier transformation.

Jump-return NOESY and NOESY experiments at 25 °C afforded at least one d_{NN}, $d_{\alpha N}$, or $d_{\beta N}$ sequential connectivity throughout the entire length of the peptide chain. Proline spin systems were inserted into their correct positions in the sequence based on observed $d_{\alpha\delta}$, $d_{\beta\delta}$, $d_{\gamma\delta}$, $d_{N\delta}$, $d_{\alpha N}$, $d_{\delta N}$, and $d_{\gamma N}$ crosspeaks. All six proline residues showed connectivities to both the preceding and the following residues in the sequence. The Felix 2.30 Assign module provided verification of the manually generated spin system identities and peptide spanning sequential connectivities.

Deuterium exchange rates were determined using a pretuned, preshimmed instrument by lyophilizing fully protonated TM4, for which the pH was previously adjusted to pH 5.2, and redissolving the peptide in 100% D_2O. A series of eight 3.5-h clean-TOCSY spectra were collected with the first experiment beginning 10 min after redissolving the sample in 100% D_2O. Each exchange experiment was acquired at 298 K with 256 t_2 increments each with 32 scans of 2,048 complex data points. Exchange spectra were processed in the same manner as other clean-TOCSY spectra. Slowly exchanging amide protons were identified as those continuing to elicit signal 24 h after the peptide was redissolved, whereas amide protons continuing to produce signal for anywhere from 10 min to less than 24 h were classified as moderately slowly exchanging.

Structure calculations

NOESY crosspeak intensities at 100, 200, and 300 ms at 25 °C in both H_2O and D_2O were measured and converted to upper-bound distance restraints through the use of NOE buildup curves calculated in the Felix assign module. NOE buildup curves were created from the peak volumes measured at the three mixing times and were calibrated to interproton distances based on a small set of clean intraresidue peaks chosen from the 100-ms jump-return NOESY and NOESY data sets, which corresponded to proton interactions of a fixed distance (Ni et al., 1992). Dirty or overlapped peaks were not included as NOE-based distance restraints but were manually converted to generic distance restraints on the basis of estimated peak volumes at 100 ms. Peaks suspected of being affected by spin diffusion as evidenced by nonlinear buildup curves were discarded. Intraresidue constraints that were redundant with covalent limits were also discarded. NOEs were classified as strong, medium, and weak corresponding to interproton distance restraints of 1.8–3.0, 1.8–4.0, and 1.8–5.0 Å. Pseudoatom corrections (Wüthrich, 1986) were used only in cases of suspected degenerate signals for the DGII and Discover-based calculations. In cases where prochiral constituents were spectroscopically distinguishable but not assigned stereospecifically, pseudoatom corrections were not used. In such cases, the resonances were assigned arbitrarily and were allowed to freely interchange their chirality during the structure calcu-

lation process. For distance restraints involving pseudoatoms, the upper bound was corrected.

Structures were calculated using distance geometry methods (Havel, 1991) through the NMR-Refine module of the INSIGHT II version 2.3.9 software package (Biosym Technologies, San Diego, California). Attempts to assign the NOE data in accordance with a model of TM4 based on the structure of EGF protein (Montelione et al., 1987; Hommel et al., 1992) were unsuccessful, especially for NOEs involving the C-terminal loop and intersubdomain NOEs. Therefore, an initial set of 273 unambiguous NOE-based distance restraints was used to calculate a set of 50 structures starting from the EGF-based model of TM4 using the DGII distance geometry software. The structure that had the smallest RMSD relative to the average of the 50 structures and that had no significant restraint violations was used as a model for further restraint generation. A homology-based structure of the periplasmic cyclophilin was also inconsistent with the NOE data, and this structure was calculated in an analogous manner (Clubb et al., 1994). Initial structure calculations were performed with the distance restraints enforced with a force constant of only 1 kcal/mol in order to emphasize proper molecular geometry over the distance restraints in order to facilitate the identification of misassigned NOEs. No significant differences were observed between structures calculated using the TM4 structure modeled on EGF protein as the starting structure or a TM4 structure from an intermediate round of calculations as the starting structure.

A final set of 50 embedded structures was calculated using 519 NOE-based and generic distance restraints enforced with a force constant of 30 kcal·mol^{-1}·Å^{-2}. The TM4 structure modeled on EGF protein was used as the starting structure. All other parameters were set to the default values in the DGII software (Biosym Technologies, 1993). The embedded structures were optimized through 10^6 iterations of DGII-based simulated annealing with the initial energy value set to 1,024 kcal/mol, the time step set to 0.2 ps, and all other parameters left at their default values (Biosym Technologies, 1993). Of the 50 final embedded structures, 41 achieved a DGII error function value of less than 0.5 after the simulated annealing phase of the optimization and were selected for further analysis. The 41 structures were then cooled and minimized against the distance restraints using the conjugate gradients algorithm until the derivative reached 1×10^{-5} (Biosym Technologies, 1993).

Discover refinement

The optimized structures were further refined in three steps using the Discover module (Biosym Technologies, San Diego, California) with 2 of the 41 structures discarded prior to further refinement due to excessively high starting energies. The three refinement steps used the cvff forcefield and a Leonard-Jones potential with a globally enforced dielectric constant of 1.0, charges, and the distance restraints again enforced with a force constant of 30 kcal·mol^{-1}·Å^{-2}, included in the simulation. All force constants were scaled to 1.0 during the three stages of further refinement (Clubb et al., 1994). The first stage of further refinement consisted of 100 iterations of steepest descents energy minimization. Second, the structures were subjected to a molecular dynamics protocol consisting of a 1-ps initialization at 300 K followed by 20 ps of further dynamics in two steps at 300 K. Finally, conjugate gradients minimization was carried out until the derivative reached 0.01. Further minimization did not substantially lower the final energies obtained. All three steps of refinement were essential for producing low energy structures, especially the molecular dynamics, which reduced the energies of the structures to less than 300 kcal·mol^{-1} compared to approximately 800 kcal·mol^{-1} when minimization alone was used. The lowest energy structures obtained were also those with the lowest residual distance constraint violations.

X-PLOR refinement

The program X-PLOR was used to optimize bond lengths and bond angles (Brünger, 1988). For these calculations, the restraint file from the DGII calculations was edited so that atom names referring to nonstereospecifically assigned resonances were designated by wild cards instead of the arbitrary numerical designations used in DGII and the upper bounds were increased to account for pseudoatom corrections (Wüthrich, 1986). Slow-cooling refinement was performed in which the system was cooled from its initial value of 1,000 K to its final value of 100 K in 50 K steps. All masses were set to uniform values (100 a.u.) to speed molecular dynamics (Nilges et al., 1988). An NOE force constant of 30 kcal·mol^{-1}·Å^{-2} was held constant throughout the protocol, and the van der Waals force constant was increased slowly over the course of the refinement from 0.003 kcal·mol^{-1}·Å^{-2} to 4.0 kcal·mol^{-1}·Å^{-2}. The final value of the van der Waals radii was set to 0.75 times the values in the X-PLOR parallhdg.pro parameter set, which is approximately the same as the hard sphere van der Waals radii employed by the programs DISMAN (Braun & Go, 1985) and DIANA (Güntert et al., 1991). A total of 2,000 cooling steps of 5-fs dynamics were employed followed by 500 steps of restrained Powell energy minimization. The lowest energy structures obtained after the X-PLOR-based refinement were again those that exhibited the smallest violations of the distance constraints.

Acknowledgments

We thank Dr. John Fenton for the generous gift of human α-thrombin. We thank Dr. Jack Kyte for use of his amino acid analyzer, Drs. Susan Taylor and Siv Garrod for N-terminal sequencing, Gary Siusdak at the Scripps Research Institute for electrospray mass spectrometry analysis, and Craig Dobbs for advice on peptide synthesis. This work was supported by NIH grant HL47463, by the Rita Allen Foundation, and by the Searle Scholars program. This work was done during the tenure of a graduate research fellowship from the American Heart Association, California Affiliate to M.J.H.

References

Baron M, Norman DG, Harvey TS, Handford PA, Mayhew M, Tse AGD, Brownlee GG, Campbell ID. 1992. The three-dimensional structure of the first EGF-like module of human factor IX: Comparison with EGF and TGFα. *Protein Sci 1*:81–90.

Biosym Technologies. 1993. *NMRchitect user guide, version 2.3*. San Diego, California: Biosym Technologies.

Braun W, Go N. 1985. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol 186*: 611–626.

Braunschweiler L, Bodenhausen G, Ernst RR. 1983. Analysis of networks of coupled spins by multi quantum NMR. *Mol Phys 48*:535–560.

Brünger AT. 1988. *X-PLOR version 3.1. A system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University Press.

Campbell ID, Bork P. 1993. Epidermal growth factor-like modules. *Curr Opin Struct Biol 3*:385–392.

Clubb RT, Ferguson SB, Walsh CT, Wagner G. 1994. Three-dimensional solution structure of *Escherichia coli* periplasmic cyclophilin. *Biochemistry 33*:2761–2772

Davis DG. 1989. Elimination of baseline distortions and minimization of artifacts from phased 2D NMR spectra. *J Magn Reson 81*:603–607.

Diamond R. 1992. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci 1*:1279–1287.

Esmon CT. 1989. The roles of protein C and thrombomodulin in the regulation of blood coagulation. *J Biol Chem 264*:4743–4746.

Esmon NL. 1989. Thrombomodulin. *Prog Hemostasis Thrombosis 9*:29–55.

Glaser CB, Morser J, Clarke JH, Blasko E, McLean K, Kuhn I, Chang RJ, Lin JH, Vilander L, Andrews WH, Light DR. 1992. Oxidation of a specific methionine in thrombomodulin by activated neutrophil products blocks cofactor activity. A potential rapid mechanism for modulation of coagulation. *J Clin Invest 90*:2565–2573.

Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, Presky DH, Familletti PC, Wolitsky BA, Burns DK. 1994. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature 367*:532–538.

Gray W. 1993a. Disulfide structures of highly bridged peptides: A new strategy for analysis. *Protein Sci 2*:1732–1748.

Gray W. 1993b. Echistatin disulfide bridges: Selective reduction and linkage assignment. *Protein Sci 2*:1749–1755.

Griesinger C, Otting G, Wüthrich K, Ernst RR. 1988. Clean TOCSY for ^{1}H spin system identification in macromolecules. *J Am Chem Soc 110*: 7870–7872.

Güntert P, Braun W, Wüthrich K. 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol 217*:517–530.

Havel TF. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog Mol Biol Biophys 56*:43–78.

Hayashi T, Zushi M, Yamamoto S, Suzuki K. 1990. Further localization of binding sites for thrombin and protein C in human thrombomodulin. *J Biol Chem 265*:20156–20159.

Hojrup P, Magnusson S. 1987. Disulfide bridges of bovine factor X. *Biochem J 245*:887–892.

Hommel U, Harvey TS, Driscoll PC, Campbell ID. 1992. Human epidermal growth factor high resolution solution structure and comparison with human transforming growth factor α. *J Mol Biol 227*:271–282.

Huang LH, Chen H, Pardi A, Tam JP, Sweeney W. 1991. Sequence-specific ^{1}H NMR assignments, secondary structure, and location of the calcium binding site in the first epidermal growth factor like domain of blood coagulation factor IX. *Biochemistry 30*:7402–7409.

Hunter MJ, Komives EA. 1995. Different disulfide bonding isomers of the fifth EGF-like domain of thrombomodulin have similar thrombin-binding affinities. *Protein Sci*. Forthcoming.

Kisfaludy L, Low M, Nyeki O, Szirtes T, Schon I. 1973. Die verwendung von pentafluorphenylestern bei peptid-synthesen. *Liebigs Ann Chem 9*: 1421–1429.

Kumar A, Ernst RR, Wüthrich K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross relaxation networks in biological macromolecules. *Biochem Biophys Res Commun 95*:1–6.

Lentz SR, Chen Y, Sadler JE. 1993. Sequences required for thrombomodulin cofactor activity within the fourth epidermal growth factor-like domain of human thrombomodulin. *J Biol Chem 268*:15312–15317.

Lougheed JL, Bowman CA, Meininger DP, Komives EA. 1995. Inhibition of thrombin by cyclic peptides from thrombomodulin. *Protein Sci 4*: 773–780.

Macura S, Ernst RR. 1980. Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol Phys 41*:95–117.

Marion D, Ikura M, Bax A. 1989. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J Magn Reson 84*:425–430.

Marion D, Wüthrich K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun 113*(3): 967–974.

Montelione GT, Wüthrich K, Nice EC, Burgess AW, Scheraga HA. 1987. Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. *Proc Natl Acad Sci USA 84*:5226–5230.

Moy FJ, Li YC, Rauenbuehler P, Winkler ME, Scheraga HA, Montelione GT. 1993. Solution structure of human type-α transforming growth factor determined by heteronuclear NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry 32*:7334–7353.

Nagashima M, Lundh E, Leonard JC, Morser J, Parkinson JF. 1993. Alanine scanning mutagenesis of the epidermal growth factor domains of human thrombomodulin identifies critical residues for its cofactor activity. *J Biol Chem 268*:2888–2892.

Ni F, Ripoll DR, Purisima EO. 1992. Conformational stability of a thrombin-binding peptide derived from the hirudin C-terminus. *Biochemistry 31*:2545–2554.

Nilges M, Clore M, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett 239*:129–136.

Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W, Huber R, Blankenship DT, Cardin AD, Kisiel W. 1993. Structure of Human Des (1-45) factor Xa at 2.2 Å resolution. *J Mol Biol 232*:947–966.

Plateau P, Guéron M. 1982. Exchangeable proton NMR without baseline distortion using new strong pulse sequences. *J Am Chem Soc 104*:7310–7311.

Savage CR, Hash JH, Cohen S. 1973. Epidermal growth factor location of disulfide bonds. *J Biol Chem 248*:7669–7672.

Selander-Sunnerhagen M, Ullner M, Persson E, Teleman O, Stenflo J, Drakenberg T. 1992. How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH_2-terminal EGF-like domain in coagulation factor X. *J Biol Chem 267*:19642–19649.

Smith BO, Downing AK, Dudgeon TJ, Cunningham M, Driscoll PC, Campbell ID. 1994. Secondary structure of fibronectin type 1 and epidermal growth factor from tissue-type plasminogen activator by nuclear magnetic resonance. *Biochemistry 33*:2422–2429.

Srinivasan J, Hu S, Hrabal R, Zhu Y, Komives EA, Ni F. 1994. Thrombin-bound structure of an EGF subdomain from human thrombomodulin determined by transferred nuclear Overhauser effects. *Biochemistry 33*:13553–13561.

Stearns DJ, Kurosawa S, Esmon CT. 1989. Microthrombomodulin. *J Biol Chem 264*:3352–3356.

Wen D, Dittman WA, Ye RD, Deaven LL, Majerus PW, Sadler JE. 1987. Human thrombomodulin: Complete cDNA sequence and chromosome localization of the gene. *Biochemistry 26*:4350–4357.

White CE, Hunter MJ, Meininger DP, Komives EA. 1995. Large scale expression, purification and characterization of small fragments of thrombomodulin: The roles of the sixth domain and of methionine 388. *Protein Engineering* (in press).

Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*:1647–1651.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: John Wiley and Sons.

Ye J, Esmon NL, Esmon CT, Johnson AE. 1991. The active site of thrombin is altered upon binding to thrombomodulin. *J Biol Chem 266*:23016–23021.