Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10177–10182, September 1996
Biophysics

The fifth epidermal growth factor-like domain of thrombomodulin does not have an epidermal growth factor-like disulfide bonding pattern

[protein C/thrombin/*Pichia pastoris*/yeast expression/tris(2-carboxyethyl)phosphine]

CHRISTOPHER E. WHITE, MICHAEL J. HUNTER, DAVID P. MEININGER, SIV GARROD, AND ELIZABETH A. KOMIVES*

Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0601

Communicated by Bruno H. Zimm, University of California at San Diego, La Jolla, CA, June 28, 1996 (received for review March 28, 1996)

ABSTRACT **The disulfide bonding pattern of the fourth and fifth epidermal growth factor (EGF)-like domains within the smallest active fragment of thrombomodulin have been determined. In previous work, this fragment was expressed and purified to homogeneity, and its cofactor activity, as measured by k_{cat} for thrombin activation of protein C, was the same as that for full-length thrombomodulin. CNBr cleavage at the single methionine in the connecting region between the domains and subsequent deglycosylation yielded the individual EGF-like domains. The disulfide bonds were mapped by partial reduction with tris(2-carboxyethyl)phosphine according to the method of Gray [Gray, W. R. (1993) *Protein Sci.* 2, 1732–1748], which provides unambiguous results. The disulfide bonding pattern of the fourth EGF-like domain was (1–3, 2–4, 5–6), which is the same as that found previously in EGF and in a synthetic version of the fourth EGF-like domain. Surprisingly, the disulfide bonding pattern of the fifth domain was (1–2, 3–4, 5–6), which is unlike that found in EGF or in any other EGF-like domain analyzed so far. This result is in line with an earlier observation that the (1–2, 3–4, 5–6) isomer bound to thrombin more tightly than the EGF-like (1–3, 2–4, 5–6) isomer. The observation that not all EGF-like domains have an EGF-like disulfide bonding pattern reveals an additional element of diversity in the structure of EGF-like domains.**

Thrombomodulin (TM) has a cysteine-rich extracellular domain with six regions of sequence that resemble epidermal growth factor (EGF-like domains) (1–3). To date, more than 300 sequences have been identified and classified as EGF-like domains (4, 5). The classification is based mainly on the spacing and presence of six cysteine residues within approximately 40 amino acid residues. It is commonly held that all EGF-like domains have the same disulfide bonding pattern as EGF, but we report here that the fifth EGF-like domain of TM appears to have an anomalous disulfide bonding pattern.

The structure of EGF itself is well-determined by NMR (6, 7). The molecule has a central two-stranded β-sheet and a second shorter sheet, or double hairpin, at the C terminus. EGF contains six cysteines that form three disulfide bonds connecting the first cysteine to the third, the second to the fourth, and the fifth to the sixth (8). The disulfide bonds that connect the first cysteine to the third and the second cysteine to the fourth cross each other near the middle of the central two-stranded β-sheet. Thus, the central two-stranded β-sheet is cemented by the crossing first-to-third and second-to-fourth disulfide bonds, and the double hairpin is cemented by the fifth-to-sixth disulfide bond (Fig. 1). The fourth and fifth cysteines are always only one amino acid apart, so that a close connection exists between the two β-strand substructures.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

FIG. 1. Stereoview of a ribbon diagram of EGF [3EGF from the Protein Data Bank (6)]. The disulfide bonds are shown by black lines, and the cysteines are numbered.

Structures of nine EGF-like domains have been determined, and all of the structures are consistent with an EGF-like disulfide bonding pattern. Seven of these EGF-like domains, the first EGF from coagulation factor X (9), the first EGF from coagulation factor IX (10), transforming growth factor type α (11), the urokinase EGF-like domain (12), the EGF-like domain from E-selectin (13), the heregulin EGF-like domain (14), and the EGF-like domain from tissue plasminogen activator (15), are very similar to EGF. Two others, the second EGF-like domain from factor X (16) and the fourth EGF-like domain from TM (17), resemble EGF in the N-terminal half of the molecule, but the C-terminal loops have a broadened shape that does not superimpose well on the C-terminal loop of EGF.

The disulfide bonding pattern has been unambiguously determined by either chemical methods or by proteolytic mapping for murine EGF (8), the first EGF-like domains from coagulation factors IX (18) and X (19), transforming growth factor type α (20), and a synthetic version of the fourth EGF from TM (17). These domains all have the same disulfide bonding pattern as EGF despite their relatively low sequence similarity and the large degree of variability in the numbers of amino acids between the cysteines (Fig. 2). Although more than 300 EGF-like domains have been identified by sequence similarity, there are relatively few for which the disulfide bonding pattern is unambiguous.

Our previous work on a synthetic peptide corresponding to the fifth EGF-like domain of TM showed that the domain formed several disulfide bonded isomers when refolded in a redox buffer, and that all of the isomers bound to thrombin. Surprisingly, thrombin binding affinity correlated with lack of crossing in the disulfide bonds, so that the uncrossed (1–2, 3–4, 5–6) isomer bound to thrombin nearly an order of magnitude more tightly than the EGF-like (1–3, 2–4, 5–6) isomer (23). At

Abbreviations: EGF, epidermal growth factor; TM, thrombomodulin; TMEGF(4–5), a fragment that is composed of the fourth and fifth EGF-like domains of TM; TCEP, tris(2-carboxyethyl)phosphine; TMEGF4, the fourth EGF-like domain of TM; TMEGF5, the fifth EGF-like domain of TM.

*To whom reprint requests should be addressed. e-mail: ekomives@ucsd.edu.

Source	Sequence	% Similarity
consensus	XXXXC- X_{1-7}-CXXXGXC--- X_{1-13} ---CXCXXG(F/Y)X_{1-6}GXXCX_{1-11}	
mEGF	NSD-SECPLSHDGYCLHDGVCMYIEALDK---YACNCVVGYI-----GERCQYRDLKWWELR	100%
hTGFα	VSHFNDC-DSHTQFCFH-GTCRFLVQEDK----ACVCHSGYV-----GARCEHADLLA	41%
F.IXEGFI	YVDGDQC-ESNP--CLNGGSCK--DDINS---YECWCPFGFE-----GKNCELDVT	33%
F.XEGFI	-KDGDQC-EGHP--CLNQGHCK--DGIGD---YTCTCAEGFE-----GKNCEFSTR	35%
TMEGF4	--PVDPCFRAN---CEYQ--CQPLNQ-TS---YLCVCAEGFAPIPHEPHRCQMF	27%
F.XEGFII	---RKLCSLDNGD-C-DQ-FCH--EEQNS---VVCSCARGY-TLADNGKACIP	29%
tPAEGF	---VKSCSEPR---CFNGGTCQQALYFSD---FVCQCPEGFA-----GKSCEID	26%
uPAEGF	QVP-SNCD------CLNGGTCVSNKYFSNI--HWCNCPKKFG-----GQHCEIDK	28%
heregEGF	TSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFT-----GARCTENV	27%
Esel-EGF	-----ACTN---TSCSGHGECVETINN-----YTCKCDPGFS-----GLKCEQ	29%
TMEGF5	----MFC--NQT-AC--PADC---DPNTQAS---CECPEGY--ILDDGFICTDIDE	28%

FIG. 2. Comparison of the sequences of various EGF-like domains. For the first group, the disulfide bonding pattern has been chemically determined. mEGF, murine EGF; hTGFα, human transforming growth factor type α; F.IXEGFI, the EGF-like domain from coagulation factor IX closest to the N terminus; F.XEGFI, the EGF-like domain from coagulation factor X closest to the N terminus; TMEGF4, the fourth EGF-like domain from TM. For the second group, structures that are consistent with an EGF-like disulfide bonding pattern have been determined. F.XEGFII, the EGF-like domain from coagulation factor X closest to the C terminus; tPAEGF, the EGF-like domain from tissue plasminogen activator; uPAEGF, the EGF-like domain from urokinase; heregEGF, heregulin; Esel-EGF, the EGF-like domain from E-selectin. The sequence of TMEGF5 (the fifth EGF-like domain from TM) is given last. The degree of similarity among the sequences was determined by the method of Feng and Doolittle (21) using the BLOSUM algorithm of Henzikoff and Henzikoff (22).

first, this anomalous result was thought to be an artifact of removing the fifth domain from the context of the rest of the TM molecule. Indeed, the first cysteine in the domain is only the fourth residue from the end of the polypeptide, whereas in the full-length TM molecule, it follows the fourth domain. Misfolding of the domain could easily result from end effects. Furthermore, the connecting region between the fourth and fifth EGF-like domains in TM is critical for activity, so disconnection at this point could easily perturb results from biological assays. We have, therefore, determined the disulfide bonding pattern of the fifth domain within the context of a fragment of TM with full cofactor activity. Although sufficient quantities of the native protein were not available, expression and kinetic analysis of a fragment of TM in *Pichia pastoris* have now provided sufficient quantities of a TM fragment with convincing cofactor activity. This fragment [TMEGF(4–5)] is composed of the fourth and fifth EGF-like domains of TM and extends from residue E346 to residue E426 of the human TM sequence (3). We present results that show that the fifth EGF-like domain, isolated from the active TMEGF(4–5) fragment, has an anomalous disulfide bonding pattern.

EXPERIMENTAL PROCEDURES

Reagents. Trifluoroacetic acid was from Aldrich, PNGase F was from Glyko (Novato, CA), and endoglycosidase H was from Boehringer Mannheim. All other reagents and chemicals were reagent grade from Fisher Scientific or Sigma.

TM Fragment Characterization. TMEGF(4–5) was produced by high-density fermentation of the SMD1168 strain of *P. pastoris* transformed with the gene for TMEGF(4–5) as described (3). The expressed, folded protein was secreted into the culture supernatant from which it was purified by anion exchange chromatography, ultrafiltration (PM30 membrane, Amicon) to remove high molecular weight contaminants, HiLoad Q chromatography, and finally reverse-phase HPLC (3). The purified TMEGF(4–5) was greater than 95% pure by N-terminal sequencing, amino acid analysis, and reverse-phase HPLC. The kinetic parameters of this fragment were determined previously; the K_m for TM is 120 nM, which is 10-fold higher than that for native TM and is due to the absence of the sixth domain (3). The k_{cat} for thrombin activation of protein C by the expressed TMEGF(4–5) fragment was 2 sec^{-1}, which is the same as that for full-length human TM (24).

Proteins produced in *P. pastoris* have been shown to contain N-linked high mannose sugars, and N-terminal analysis indicated the probable presence of glycosylation on two NQT sequences. Deglycosylation of TMEGF(4–5) with PNGase F (Glyko) appeared to remove all of the sugars from only the NQT site in the fifth domain, so the sugars on the fourth domain were partially removed by treatment with endoglycosidase H (Boehringer Mannheim), which removes high mannose sugars but not the core sugar residues.

CNBr Cleavage of TMEGF(4–5). Glycosylated, HPLC purified TMEGF(4–5) (30 mg) was resuspended in 50 ml of 6 M guanidine hydrochloride in 70% formic acid (Sigma). After addition of 3.5 g of CNBr, the reaction proceeded for 20 hr in the dark. The products were diluted with 500 ml of purified H_2O (Milli-Q water purification system, Millipore) and freeze-dried. The lyophilized protein was resuspended in Milli-Q-purified H_2O and chromatographed in three portions on a C_{18} HPLC column (10 × 250 mm; Vydac, Hesperia, CA) at a flow rate of 3 ml/min. The gradient was 0.1% trifluoroacetic acid for 10 min, then 0–10% acetonitrile over 10 min, and finally 10–40% acetonitrile over 90 min at a flow rate of 3 ml/min. Detection was at 280 nm. The glycosylated fourth and fifth EGF-like domains of TM do not separate under these conditions.

Deglycosylation of the Mixture of the Fourth and Fifth EGF-Like Domains. A dry 5-mg portion of the purified CNBr products containing both the glycosylated fourth and fifth EGF-like domains was resuspended in 500 μl of Milli-Q-purified H_2O and apportioned into 20 1.5-ml Eppendorf tubes containing 25 μl each. After addition of 25 μl of 2× reaction buffer (100 mM sodium phosphate, pH 7.5) and 2 μl of PNGase F (2.5 units/ml), incubation proceeded for 20 hr at 37°C, each reaction was diluted to 100 μl with Milli-Q-purified H_2O, the pH was adjusted to 5.5 with 0.5 N HCl, and 10 μl of endoglycosidase H was added (Boehringer Mannheim; 1 milliunit/μl). The tubes were incubated for an additional 16 hr, and the samples were pooled and purified on a Vydac C_{18} HPLC column (10 × 250 mm) using the same gradient described above. This chromatography step afforded the separation of partially deglycosylated forms of the fourth domain; however, the major partially deglycosylated fourth domain fraction still coeluted with the fifth domain fraction as assessed by N-terminal sequencing analysis of the single major peak.

Separation of the Deglycosylated Fourth and Fifth EGF-Like Domains. The single major peak was chromatographed on a Vydac C_{18} HPLC column (4.6 × 250 mm) at a flow rate of 1 ml/min using the same extended gradient and detection described above. This analytical scale separation step afforded separation of the fourth and fifth domains. N-terminal sequencing confirmed that the leading, broader peak was the fourth domain (referred to as TMEGF4) and the sharp peak that eluted later was the completely deglycosylated fifth domain (referred to as TMEGF5).

Determination of the Disulfide Bonding Pattern of the Fourth and Fifth EGF-Like Domains. Approximately 100 μg of TMEGF4 or TMEGF5 was partially reduced by tris(2-carboxyethyl)phosphine (TCEP) to map the disulfides (25, 26). The TCEP solution was prepared by dissolving 32.4 mg of TCEP and 250 mg of citric acid in 5 ml of Milli-Q-purified H_2O and adjusting the pH to 3.0 by the dropwise addition of 1 M NaOH. The TMEGF4 or TMEGF5 was resuspended in 700 μl of 0.1% trifluoroacetic acid, 700 μl of TCEP solution was added, and the mixture was vortexed and incubated 60 min for TMEGF4 or 30 min for TMEGF5. The different reaction times were determined from previous work on the synthetic fourth and fifth domains (17, 23). Partial reduction products were separated on a Vydac C_{18} (4.6 × 250 mm) column with a gradient of 0.1% trifluoroacetic acid for 10 min, then 0–10% acetonitrile over 10 min, and finally 10–40% acetonitrile over 90 min at a flow rate of 1 ml/min and detection at 280 nm. Portions (500 μl) of each HPLC peak were collected and immediately injected into a Falcon tube containing 400 μl of 0.5 M Tris-acetate buffer (pH 8.0), 2 mM EDTA, and either 2.2 M iodoacetamide for the fourth domain or 2.2 M *N*-methyliodoacetamide for the fifth domain (12, 23). Iodoacetamide could not be used as the alkylating agent for the fifth domain because the phenylthiohydantoin derivative of *S*-carboxamidomethylcysteine elutes at the same time as that of glutamic acid, which is the single residue between the fourth and fifth cysteines (23). The reactions were quenched after 1 min by acidification with 800 μl of 0.5 M citric acid. The alkylated products were purified by analytical reverse phase HPLC using the same conditions as above, and characterized by N-terminal sequencing.

RESULTS

Kinetic Characterization of TMEGF(4–5). The TMEGF(4–5) fragment extends from amino acid E346 to E426 in the human TM sequence (Fig. 3). Kinetic characterization of this fragment revealed that it had full TM cofactor activity as assessed by the k_{cat} for protein C activation by the complex formed between the TM fragment and thrombin, which was 2 sec^{-1}, the same as that found for full-length human TM (3, 24). Thus, although this fragment was derived from an expression system, the fact that it has full cofactor activity strongly suggests that its disulfide bonding pattern will be the same as that found in the full-length protein.

Separation of the Fourth and Fifth Domains. The single methionine at position 388 in the connecting region between the fourth and fifth EGF-like domains of TM offered a convenient mechanism for separating the two domains (1). Addition of guanidine to the CNBr cleavage reaction greatly facilitated the cleavage, and the reaction could be carried to more than 50% of completion before additional side reactions occurred. After CNBr treatment of TMEGF(4–5), the fourth and fifth domains could not be separated physically (Fig. 4*A*). The most likely reason for the lack of separability of the two domains was the broad elution profile of glycosylated proteins under standard reverse phase HPLC conditions. Even after deglycosylation, a single major peak was isolated; this peak was shown by N-terminal sequencing to contain both the fourth and fifth domains (Fig. 4*B*). The other peaks in the chromatogram were also analyzed by N-terminal sequencing and were all shown to have the N-terminal sequence of the fourth domain. These other products most likely result from incomplete deglycosylation of the fourth domain or incomplete CNBr cleavage. Analytical scale chromatography of the single

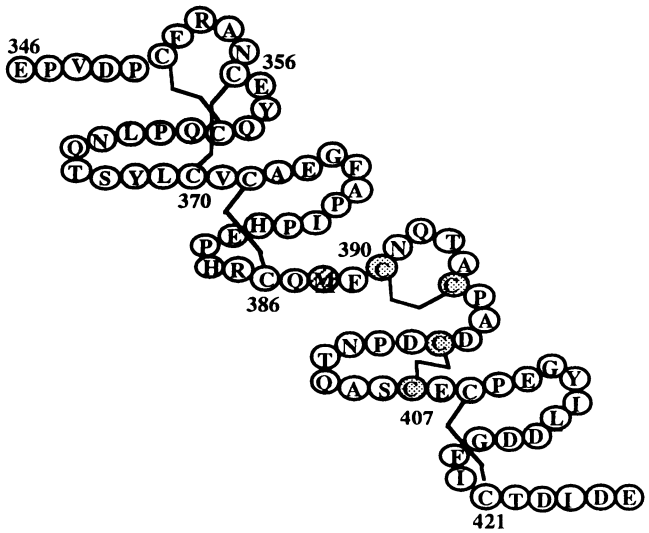

FIG. 3. Schematic diagram of TMEGF(4–5). M388, at which CNBr cleavage takes place, is striped. The cysteines (C390, C395, C399, and C407) for which the disulfide bonding pattern is questionable are shaded. The disulfide bonding pattern shown for each domain is that determined in the results presented here.

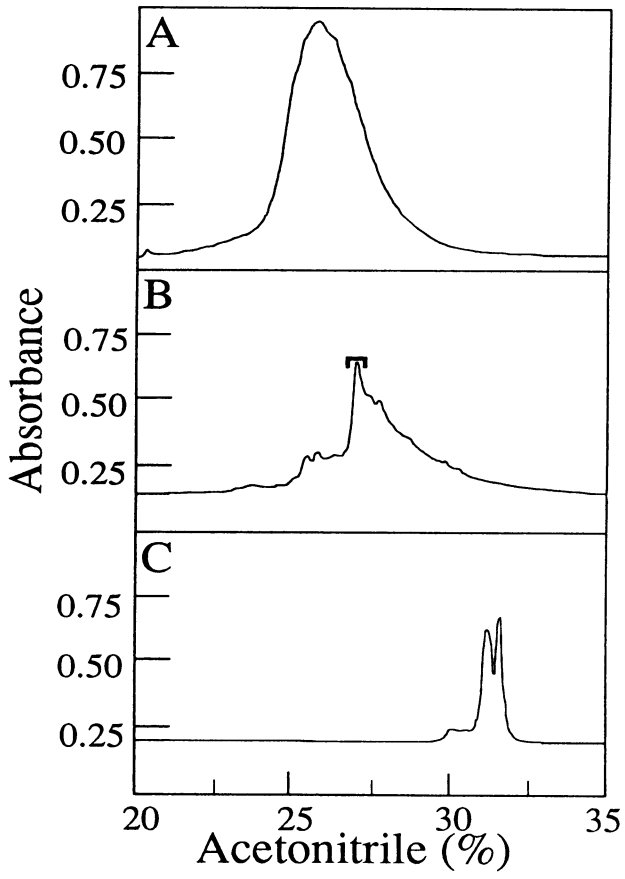

FIG. 4. HPLC purification of the fourth and fifth EGF-like domains after CNBr cleavage. (*A*) HPLC trace from purification of a portion of the CNBr reaction products. (*B*) HPLC trace from purification of the deglycosylated domains. (*C*) HPLC trace of the separation of the deglycosylated fourth and fifth domains by repurification of the peak marked with the bracket in *B*.

major peak obtained from reverse phase HPLC yielded the separated domains (Fig. 4*C*). The earlier eluting peak, which had the N-terminal sequence of the fourth domain, was somewhat broad because it still contained core sugar residues. The later eluting peak had the N-terminal sequence of the fifth domain and was sharp because the sugars had been completely removed. Proof that the removal of the sugars had occurred was obtained from the N-terminal sequencing data, which showed N391 had been converted to aspartic acid, which is the expected result of PNGase F deglycosylation.

Standardization of the Partial Reduction Method for Mapping Disulfide Bonds. In 1993, Gray published two extensive reports of a novel method for determining the disulfide bond connectivities in highly disulfide bonded proteins (25, 26). The method relies on partial reduction by TCEP at low pH, reverse phase HPLC purification of the partially reduced products (also at low pH), alkylation of the purified partially reduced products by rapid addition to a supersaturated solution of iodoacetamide, repurification of the alkylated products, and finally Edman sequencing to determine the location of the alkylated cysteines. Gray showed definitively that at pH values of 3 or below, disulfide bond exchange did not occur in any of 13 various disulfide bonded proteins he mapped, some of which had more closely spaced cysteines than those of EGF-like domains (25). Gray determined that disulfide bond exchange rarely occurred during the alkylation step, resulting in multiple products upon repurification of the alkylated, partially reduced peptide.

For a protein with three disulfide bonds, six partially reduced products may be formed upon treatment with TCEP, three with a single broken disulfide bond and three with two broken disulfide bonds. After HPLC separation, the reduced cysteines are alkylated, resulting in products that have alkylated cysteines in pairs corresponding to disulfide bonds originally present in the protein. N-terminal sequencing allows unambiguous determination of which cysteines were alkylated, and allows pairwise assignment of the disulfide bonds. For a protein with three disulfide bonds, a minimum of two sequences is required to unambiguously determine the disulfide bonds. One of the sequences must contain a single pair of alkylated cysteines, and the other sequence can contain either another single pair of alkylated cysteines or two pairs wherein one of the two pairs was that assigned by the first sequence.

We previously used the Gray method to map the disulfide bonds in synthetic peptides corresponding to the fourth and fifth EGF-like domain of TM (17, 23). The synthetic TMEGF4 (residues E346 to F389) folded into a single major product that had an EGF-like (1–3, 2–4, 5–6) disulfide bonding pattern (17). The results from the synthetic TMEGF4 are compared here with those obtained from TMEGF4 isolated from the expressed TMEGF(4–5) protein. After CNBr cleavage, the TMEGF4 was shorter by two amino acids at the C terminus compared with the synthetic version, so the HPLC traces of the separation of the partially reduced products were similar but not identical (data not shown). For both the synthetic and the isolated TMEGF4, the major partial reduction product had the fifth and sixth cysteines alkylated, indicating a C372–C386 (5–6) disulfide bond (Fig. 5*A*). The other products were less abundant, but two partial sequences showed no alkylation of the first and third cysteines and alkylation of the second cysteine. This information was sufficient to unambiguously identify the remaining disulfides as being in a (1–3, 2–4) pattern.

In previous work from our laboratory, synthetic peptides corresponding to three different disulfide bonded isomers of TMEGF5 (residues Q387 to E426) had been prepared and the disulfide bonding patterns determined by partial reduction (23). Despite their identical sequences, partial reduction of each disulfide bonded isomer produced a characteristic distribution of partially reduced products after HPLC separation (Fig. 6). Due to variations in the extent of reaction, direct proof for a particular disulfide bonding pattern was always obtained from alkylation, repurification, and Edman sequencing of the partially reduced products. Repurification of the alkylated products always showed a single major product indicating that exchange during alkylation had not occurred. The N-terminal sequencing data used to assign the different disulfide bonded synthetic standards are shown in Fig. 5*B*. The sequences of peaks 2 and 4 allowed the unambiguous assignment of the disulfides as being in a (1–2, 3–4, 5–6) pattern.

Source	Sequence	Result
A. TMEGF4 Isolated from TMEGF(4-5)		
SEQUENCE	E P V D P **C** F R A N **C** E Y Q **C** Q P L N Q T S Y L **C** V **C** A E G F A P I P H E P H R **C** Q M F	
peak 2	E P V D P **C** F R A N **C** E Y Q **C** Q P L N Q T S Y L **C** V **X** A E G F A P I P H E P H R **X** Q M F	(5-6)
peak 4	E P V D P **C** F R A N **X** E Y Q **C** Q P L	(1-3)
B. Synthetic fifth domain standards		
SEQUENCE	Q M F **C** N Q T A **C** P A D **C** D P N T Q A S **C** E **C** P E G Y I L D D G F I **C** T D I D E	
(1-2,3-4,5-6)		
peak 2	Q M F **X** N Q T A **X** P A D **C** D P N T Q A S **C** E **C** P E G Y I L D D G F I **C** T	(1-2)
peak 4	Q M F **X** N Q T A **X** P A D **X** D P N T Q A S **X** E **C** P E G Y I L D D G F I **C** T	(3-4)
(1-3,2-4,5-6)*		
peak 4	Q M F **X** N Q T A **X** P A D **X** D P N T Q A S **X** E **C** P E G Y I L D D G F I **C** T	(1-3)
(1-3,2-5,4-6)		
peak 1	Q M F **C** N Q T A **X** P A D **C** D P N T Q A S **C** E **X** P E G Y I L D D G F I **C** T	(2-5)
peak 3	Q M F **X** N Q T A **X** P A D **X** D P N T Q A S **C** E **X** P E G Y I L D D G F I **C** T	(1-3)
C. TMEGF5 Isolated from TMEGF(4-5)		
SEQUENCE	F **C** N Q T A **C** P A D **C** D P N T Q A S **C** E **C** P E G Y I L D D G F I **C** T D I D E	
peak 2	F **X** D Q T A **X** P A D **C** D P N T Q A S **C** E **C** P E G Y I L D D G F I **C** T	(1-2)
peak 4	F **X** D Q T A **X** P A D **X** D P N T Q A S **X** E **C** P E G Y I L D D G F I **C** T	(3-4)

FIG. 5. Results from N-terminal sequencing of the repurified, alkylated products from partial reduction of TMEGFs. C denotes a cysteine that remained in a disulfide bond and was registered as a blank in the N-terminal sequencer. X denotes an alkylated cysteine. (*A*) Sequences used to determine the disulfide bonding pattern of TMEGF4 isolated from TMEGF(4–5). (*B*) Sequences used to determine the disulfide bonding patterns of each of the synthetic standards for the fifth EGF-like domain of TM (23). *Only one analysis was carried out for the (1–3, 2–4, 5–6) disulfide bonded isomer because the 2–4 disulfide bond was chemically directed. (*C*) N-terminal sequences used to determine the disulfide bonding pattern of TMEGF5 isolated from TMEGF(4–5). Sequences begin at F389, immediately following CNBr-sensitive M388, and end at T422 after the last cysteine. Deglycosylation of N391 by PNGase F results in an aspartic acid at this position.

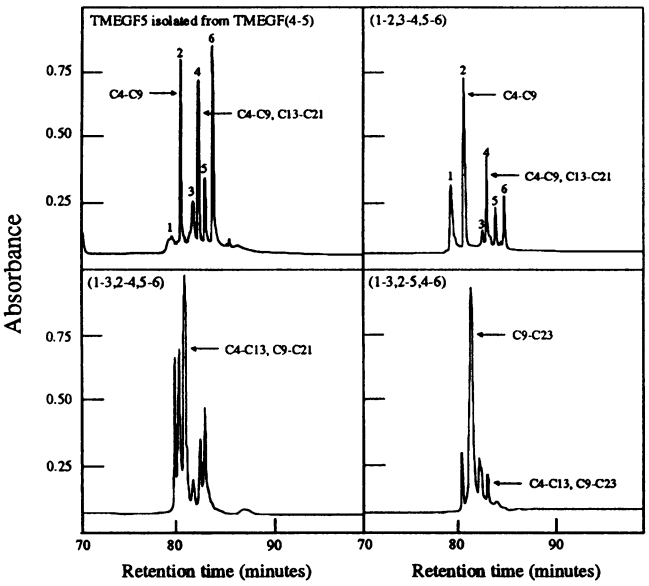

FIG. 6. HPLC separation of the partial reduction products of isolated TMEGF5 and the synthetic disulfide bond isomer standards. TMEGF5 isolated from TMEGF(4–5) is shown in the upper left panel. For the TMEGF5 trace and for the synthetic (1–2, 3–4, 5–6) isomer trace, the peaks are numbered in order of elution. Each synthetic standard panel is marked with the disulfide bonding pattern that was determined from N-terminal sequencing analysis of alkylated products.

Disulfide Bonding Pattern of the Fifth EGF-Like Domain of TM. The TMEGF5 isolated from CNBr digestion of TMEGF(4–5) was subjected to partial reduction under identical conditions used to analyze the synthetic fifth domain standards. The profile of the partially reduced products obtained from the TMEGF5 isolated from the fully active TMEGF(4–5) fragment most closely resembles the profile of the (1–2, 3–4, 5–6) synthetic standard (Fig. 6). The traces are not identical because the TMEGF5 isolated by CNBr digestion of TMEGF(4–5) does not contain Q387 or M388, and the partial reduction of TMEGF5 proceeded somewhat further than that of the synthetic (1–2, 3–4, 5–6) standard, resulting in less of peak 1 (fully oxidized TMEGF5) and more of peak 6 (fully reduced TMEGF5).

Definitive results were obtained from alkylation, repurification, and Edman sequencing of the partially reduced products. Repurification of the alkylated products again showed a single major product, strongly suggesting that disulfide bond exchange had not occurred (Fig. 7). Peaks 2 and 4 from the separation of partially reduced products were analyzed by N-terminal sequencing. The peptide isolated from peak 2 was alkylated only at the first and second cysteines, and the peptide isolated from peak 4 was alkylated on the first and second and third and fourth cysteines (Fig. 5*C*). The same two peaks (albeit with two more amino acids on the N terminus) obtained for the synthetic fifth domain with the (1–2, 3–4, 5–6) disulfide bonding pattern gave the same alkylation patterns upon N-terminal sequence analysis. Thus, the TMEGF5 isolated from TMEGF(4–5) can be unambiguously assigned the non-EGF-like (1–2, 3–4, 5–6) disulfide bonding pattern.

DISCUSSION

During previous studies of various disulfide bonded isomers of a synthetic peptide corresponding to the fifth EGF-like domain of TM, we discovered that a non-EGF-like disulfide bonded isomer bound to thrombin more tightly than the EGF-like disulfide bonded isomer (23). The synthetic single EGF-like domains, however, only bound to thrombin; they did

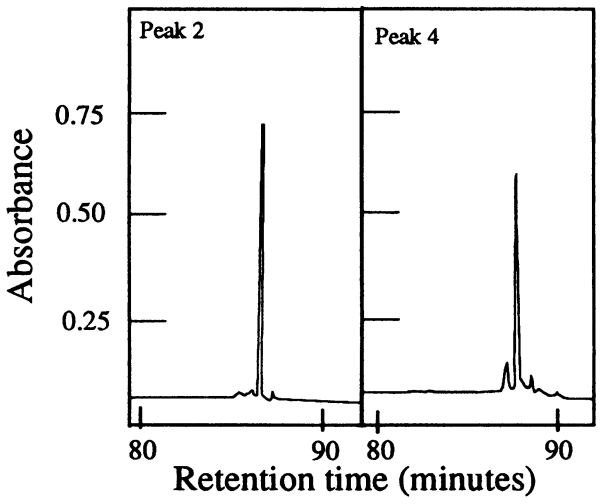

FIG. 7. HPLC traces from repurification of the alkylated, partially reduced products from TMEGF5. Peaks 2 and 4 gave sufficient Edman sequencing information to unambiguously identify the disulfide bonding pattern.

not possess TM cofactor activity. Also, even the isomer that bound to thrombin most tightly still bound 100-fold less tightly than full-length TM. It, therefore, seemed important to determine the disulfide bonding pattern of TMEGF5 isolated from a fully active TM fragment. The work previously carried out on the synthetic peptides provided a useful set of standards to calibrate the partial reduction method and with which to compare the results obtained on the native TMEGF5.

TM can only be obtained in low yields from natural sources, so we turned to an expression system to produce large amounts of active TM fragments. The TMEGF(4–5) fragment expressed in *P. pastoris* had the same k_{cat} for protein C activation as full length TM (3). Small changes in the region of the TMEGF(4–5) fragment from residues 372 to 395 alter k_{cat} but not K_m; for example, an H381G or M388L alteration results in a doubling of the specific activity, and oxidation of M388 results in a 90% drop (27, 28). Mutation of M388 affects k_{cat} but not K_m (3). Thus, the fact that the TMEGF(4–5) fragment has the same k_{cat} as full-length TM strongly suggests that the disulfide bonds at C390 and C395 are the same as that in full-length TM. These disulfide bonds are in the middle of the TMEGF(4–5) fragment and are not subject to end effects. The results presented here with TMEGF5, isolated from fully active TMEGF(4–5), can be taken as much stronger evidence that TMEGF5 has an anomalous disulfide bonding pattern.

This result calls into question the dependability of sequence similarity algorithm results of 25–30% for indicating relatedness among sets of small, cysteine-rich polypeptides such as EGF-like domains. Indeed, sequence alignment and classification of more than 300 EGF-like domains showed that similarity scores within the set were typically in the 25–35% range (5). The sequence differences are not only from amino acid substitutions but also from a large variability in the number of amino acids between the six cysteines (Fig. 2). The results presented here suggest that caution should be taken when interpreting similarity scores of 25–30% as indicating a similar disulfide bonding pattern for EGF-like domains. These results also call into question the commonly held belief that all EGF-like domains have the same disulfide bonding pattern.

In general, if two proteins have similar sequences, this can be interpreted as indicating that the proteins will have similar three-dimensional folds. In the case of EGF and EGF-like domains, it is not clear how or whether alteration of the disulfide bonding pattern will alter the overall fold. To date, at least 10 EGF-like domains have been studied, and the number

is increasing rapidly. All of these have folds very similar to that of EGF (Fig. 1). TMEGF5 has the same sequence similarity score as most other EGF-like domains but a different disulfide bonding pattern. If sequence determinants other than the disulfide bonds are important, TMEGF5 will have a fold similar to EGF despite its different disulfide bonding pattern. On the other hand, the disulfide bonds could provide a major determinant of the overall fold, in which case the structure of TMEGF5 will be very different from other EGF-like domains.

TMEGF5 (1–2, 3–4, 5–6) has three simple disulfide bonded loops and no crossing disulfides. A likely result of the uncrossing of the disulfides is to increase the flexibility of the domain. The C-terminal loop of the fifth domain binds to thrombin and has been shown to be unstructured in solution and to become a tri-stranded β-sheet upon binding to thrombin (29–32). Thus, evidence points toward a conformational change upon binding, and perhaps an induced-fit binding mechanism.

Is this a single anomaly, or will other examples of EGF-like domains with anomalous disulfide bonding patterns be found in the future? The functional relatedness of TM and the lipoprotein receptors has recently been pointed out by Davis (33). If the disulfide bonding pattern really does have something to do with the function, as we believe to be the case for TMEGF5, then the disulfide bonding patterns of the EGF-like domains in the lipoprotein receptors would be a good place to begin the search for other anomalies.

CONCLUSIONS

The disulfide bonding pattern of the fifth EGF-like domain of TM has been shown to be (1–2, 3–4, 5–6), which is different from the disulfide bonding pattern of EGF and of all other known EGF-like domains. This result calls into question the relatedness among domains defined by sequence similarity algorithms as EGF-like. The fact that TMEGF5 has the same sequence similarity score as most other EGF-like domains but a different disulfide bonding pattern also raises the question of how important the disulfide bonding pattern is in determining the overall fold of EGF-like domains.

We thank Dr. Da Fae Feng for carrying out the sequence alignments. Funding for this project was provided by National Institutes of Health Grant HL47463, by the Searle Scholars Fund, and by the Rita Allen Foundation. C.E.W. acknowledges support from Growth Regulation Training Grant T32 CA09523 from the National Cancer Institute.

1. Stearns, D. J., Kurosawa, S. & Esmon, C. T. (1989) *J. Biol. Chem.* **264,** 3352–3356.
2. Esmon, C, T. (1995) *FASEB J.* **9,** 946–955.
3. White, C. W., Hunter, M. J., Meininger, D. P., White, L. R. & Komives, E. A. (1995) *Protein Eng.* **8,** 1177–1187.
4. Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) *Nature (London)* **307,** 558–560.
5. Campbell, I. D. & Bork, P. (1993) *Curr. Opin. Struct. Biol.* **3,** 385–392.
6. Montelione, G. T., Wuthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D. & Scheraga, H. A. (1992) *Biochemistry* **31,** 236–249.
7. Hommel, U., Harvey, T. S., Driscoll, P. C. & Campbell, I. D. (1992) *J. Mol. Biol.* **227,** 271–282.
8. Savage, C. R., Hash, J. H. & Cohen, S. (1973) *J. Biol. Chem.* **248,** 7669–7672.
9. Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J. & Drakenberg, T. (1992) *J. Biol. Chem.* **267,** 19642–19649.
10. Baron, M., Norman, D. G., Harvey, T. S., Handford, P. A., Mayhew, M., Tse, A. G. D., Brownlee, G. G. & Campbell, I. D. (1992) *Protein Sci.* **1,** 81–90.
11. Moy, F. J., Li, Y.-C., Rauenbuehler, P., Winkler, M. E., Scheraga, H. A. & Montelione, G. T. (1993) *Biochemistry* **32,** 7334–7353.
12. Hansen, A. P., Petros, A. M., Meadows, R. P., Nettsheim, D. G., Mazar, A. P., Olejniczak, E. T., Xu, R. X., Pederson, T. M., Henkin, J. & Fesik, S. W. (1994) *Biochemistry* **33,** 4847–4864.
13. Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P. C., Wolitsky, B. A. & Burns, D. K. (1994) *Nature (London)* **367,** 532–538.
14. Nagata, K., Kohda, D., Hatanaka, H., Ichikawa, S., Matsuda, S., Yamamoto, T., Suzuki, A. & Inagaki, F. (1994) *EMBO J.* **14,** 3517–3523.
15. Smith, B. O., Downing, A. K., Driscoll, P. C., Dudgeon, T. J. & Campbell, I. D. (1995) *Structure* **3,** 823–833.
16. Padmanabhan, K., Padmanabhan, K. P., Tulinsky, A., Park, C. H., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D. & Kisiel, W. (1993) *J. Mol. Biol.* **232,** 947–966.
17. Meininger, D. P., Hunter, M. J. & Komives, E. A. (1995) *Protein Sci.* **4,** 1683–1695.
18. Huang, L. H., Ke, X.-H., Sweeney, W. & Tam, J. P. (1989) *Biochem. Biophys. Res. Commun.* **160,** 133–139.
19. Hojrup, P. & Magnusson, S. (1987) *Biochem. J.* **245,** 887–892.
20. Winkler, M. E., Bringman, T. & Marks, B. J. (1986) *J. Biol. Chem.* **261,** 13838–13843.
21. Feng, D. F. & Doolittle, R. F. (1987) *Methods Enzymol.* **189,** 375–390.
22. Henikoff, S. & Henikoff, J. G. (1992) *Proteins Struct. Funct. Genet.* **89,** 10915–10925.
23. Hunter, M. J. & Komives, E. A. (1995) *Protein Sci.* **4,** 2129–2134.
24. Parkinson, J. F., Grinnell, B. W., Moore, R. E., Hoskins, J., Vlahos, C. J. & Bang, N. U. (1990) *J. Biol. Chem.* **265,** 12602–12610.
25. Gray, W. (1993) *Protein Sci.* **2,** 1732–1748.
26. Gray, W. (1993) *Protein Sci.* **2,** 1749–1755.
27. Adler, M., Seto, M. H., Nitecki, D. E., Lin, J. H., Light, D. R. & Morser, J. (1995) *J. Biol. Chem.* **270,** 23366–23372.
28. Glaser, C. B., Morser, J., Clarke, J. H., Blasko, E., McLean, K., Kuhn, I., Chang, R.-J., Lin, J.-H., Vilander, L., Andrews, W. H. & Light, D. R. (1992) *J. Clin. Invest.* **90,** 2565–2573.
29. Srinivasan, J., Hu, S., Hrabal, R., Zhu, Y., Komives, E. A. & Ni, F. (1994) *Biochemistry* **33,** 13553–13560.
30. Hrabal, R., Komives, E. A. & Ni, F. (1996) *Protein Sci.* **5,** 195–203.
31. Lougheed, J. L., Bowman, C. A., Meininger, D. P. & Komives, E. A. (1995) *Protein Sci.* **4,** 773–780.
32. Blackmar, C., Healy, V. L., Hrabal, R., Ni, F. & Komives, E. A. (1995) *Bioorg. Chem.* **23,** 519–527.
33. Davis, C. G. (1990) *New Biol.* **2,** 410–419.