Localización de enlaces de disulfuro en el domínio de la cintina del factor de von Willebrand de humanos

Akira Katsumi, Elodee A. Tuley, Imre Bodó, and J. Evan Sadler†

From the Howard Hughes Medical Institute, Department of Medicine and Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

von Willebrand factor (VWF) es un polimérico de proteína de glucoproteína que es requerido para la adherencia platelet y para el subendotelio, y que se une a y estabiliza el factor de coagulación VIII en la circulación. El VWF se sintetiza en células endoteliales (2) y megacariocitos (3). La secuencia de VWF es notable por el agrupamiento de residuos de cisteína en regiones N-terminal y C-terminal. Cisteína es el aminoácido más abundante en la proteína y cuenta con 8,3% del total. Todas las cisteína residuos en el proteína en la forma de disulfuro (5), de las cuales algunas se encuentran entre subunidades. Después de la translocación al retículo endoplasmático, las subunidades de VWF se unen a través de enlaces de disulfuro entre las subunidades, formando multimeros que pueden superar 20 millones de Da de tamaño. La formación de ambos C-terminal y N-terminal intermolecular disulfuros es crítica porque sólo los largos VWF multimeros pueden efectivamente mediar la adherencia platelet (1).

El entendimiento del método de dimerización de VWF y las localizaciones de enlaces disulfuro en las subunidades dimerizadas no han sido ampliamente caracterizados. A C-terminal proteolítico fragment de VWF es dimerico (5), y estas 151 residuos fueron suficientes para formar dimers cuando expresado en COS (6). Así, en uno de los 15 cisteína residuos en la forma de disulfuro en el proteína en el prototipo subunidades (5). El C-terminal 90 residuos de VWF (residuos 1961–2050, numerados del C-terminal de la forma de VWF subunit) son una forma que, al ser homologue to the cisteína cintina (CK) de familias de proteínas. Muchos CK proteínas son dimericas, aunque el tipo de dimerización varía (7). Los mutaciones C2008Y (8), C2010R (9), y C2043R (10) se han encontrado en pacientes con variantes de la enfermedad de von Willebrand tipo 2A que no pueden formar VWF multimeros grandes. También, en un mutante del C-terminal con la mutación C1991W (11) o C2010R (9) no puede dimerizar.

El VWF CK dominio contiene 11 cisteína residuos y es particularmente similar al CK dominio de varios epiteliales mucinas y Norrlander enfermedad proteína (norrin) que forman oligómeros a través de CK-C-terminal CK dominios (12–14). Estudios de estos relacionados CK dominios son consistentes con un papel conservado para los cisteína residuos en dimerización. Aminoácido residuo Cys95 de norrin corresponde a Cys2010 de VWF, y mutaciones C95A previenen la oligomerización de CK recombinante norrin (15). Residuos Cys19244 y Cys19246 de mucina submamilar porcina (PSM) corresponden a Cys2008 y Cys2010 de VWF, y la mutación C13244A o C13246A no puede prevenir la dimerización de un CK PSM dominio (16).

Estos resultados directamente implicanSeveral cisteína residuos del VWF CK dominio en la cola de cola de dimensión de VWF. Sin embargo, los específicos cisteína residuos no se identificaron. Para abordar este problema, se ha investigado la estructura de un recombinante CK dimers por parcial reducción y alquilación, química y proteolítica digestión, mass spectrometry, and amino acid sequencing.**

**EXPERIMENTAL PROCEDURES**

*Materials*—*N*-Ethylmaleimide (NEM), 4-vinylpyridine (4VP), and di-thiothreitol (DTT) fueron de Sigma. Tris(2-carboxyethyl)phosphine hidrocloruro (TCEP) fue de Pierce. Cyanogen bromide (CNBr) fue de Aldrich. Lysyl endopeptidase (Achromobacter protease I) fue de Wako Chemicals (Osaka, Japan). Thermolysin was from Roche Molecular Biochemicals. Pfu DNA Polymerase was from Stratagene (La Jolla, California).
CA. BigDye™ Terminator Cycle Sequencing Ready Reaction and GeneAmp® dNTPs were used from PE Biosystems (Foster City, CA).

**Construction of Expression Vectors Encoding the CK Domain of Human VWF—Plasmid pSVHVF1.** 17 was used as a template for expression vector construction by sequential PCR (18). The 5′ fragment was amplified with primers CT90KpnI and CT90KpnI (5′-cat gac tac aag gac gac gat gac aag gag gag cct gag tgc aa-3′) and CT906HAS (5′-atg atg atg atg ttct tctg caa gga gct gtc g3′). The 3′ fragment was made with primers CT906HIS (5′-cat cac cat cac cat gac tac g ac gat gac gac gat gac aag gag gag cct gag tgc aa-3′) and VWFpXbaI (5′-aag ggc ttc gga gac gat c-3′). These fragments were mixed and amplified with primers CT90KpnI and VWFxboI (full length product was digested with KpnI and XhoI and subcloned into pGEM-TZF (+) (Promega, Madison, WI), to give plasmid pGEM-7Zf(+)+VWFCK. The structure of the construct was confirmed by DNA sequencing.

**Expression and Purification of Recombinant Proteins—** Each baculovirus expression vector was cotransfected into Sf9 cells (PharMingen) with linearized BaculoGold® DNA (PharMingen) and high titer recombinant baculovirus was prepared by repeated infection. High Five™ cells (Invitrogen, Carlsbad, CA) were infected and grown in Express Five™ medium plus 18 % glutamine for 96 h. Media containing VWFCK or VWFCKM were dialyzed against TB (20 mM Tris-HCl, pH 7.9, 4 M urea, and 4 M guanidine HCl) over 20 min at 4 ml/min. The remaining disulfide bonds in partially reduced and alkylated VWFCKM was dialyzed against TB (20 mM Tris-HCl, pH 7.9, and 4 M urea, and 4 M guanidine HCl, and alkylated with excess 1 M NEM in dimethyl sulfoxide (ICN Biomedicals Inc., Costa Mesa, CA) containing VWFCK or VWFCKM were pooled and loaded directly onto a C8 column (10×300 mm) containing VWFCK or VWFCKM were digested with CNBr and Thermolysin Digestion of Partially Reduced and Alkylated VWFCKM was partially reduced, alkylated, and deglycosylated with trifluoromethane sulfonic acid (Aldrich) according to the manufacturer's instructions.

**Complete Reduction with DTT and Alklyation with NEM—** VWFCK or VWFCKM was partially reduced and alkylated as described previously (25). The 5′-ends of VWFCK (200 μg) were digested with 0.75 mM sodium acetate, pH 4.6, 6 μg guanidine HCl, and 8 μl of 0.1 M DTT was incubated at 45 °C for 20 min under N2. VWFCKM (200 μg) in 30 μl of 0.75 mM sodium acetate, pH 4.6, 6 μg guanidine HCl, and 4 μl of 0.1 M DTT was incubated at 45 °C for 20 min under N2. Then 3 μl (VWFCK) or 4 μl (VWFCKM) of 1 M NEM in dimethyl sulfoxide (ICN Biomedicals Inc., Costa Mesa, CA) was added and incubated at 37 °C for 60 min under N2. After alkylation, samples were loaded onto a RP-HPLC using a C4 column (4.6×250 mm, 5 μm, Vydac) at a mobile phase flow rate of 0.1 mL/min in 0.1% aqueous trifluoroacetic acid and eluted with a gradient of acetonitrile.

**RESULTS**

**Characterization of Recombinant VWF CK Domains—** Previous studies indicated that the C-terminal 151 amino acid residues of VWF contained cysteine residues that mediate dimer formation in the endoplasmic reticulum (5, 6). The last ~90 amino acid residues of VWF are homologous to CK motifs that are present in many proteins and often mediate their dimerization (7, 8). Therefore, proteins containing the CK domain of VWF were designed for structural characterization (Fig. 1A).

Construct VWFCK consists of the VWF signal peptide and four
VWF Cystine Knot Domain

The products of partial reduction with TCEP and alkylation with NEM were separated by RP-HPLC (Fig. 2), and peak fractions were analyzed by mass spectrometry. The increment in mass due to alkylation is 1 Da (for reduction) + 125.1 Da (for alkylation by NEM) = 126.1 Da. The mass of an unmodified monomer is taken to be the mass of the unmodified dimer × 0.5, which equals 13856.6 Da for VWFCK and 13893.2 Da for VWFCKM. The number of alkylated cysteines per monomer or dimer = (observed mass – unmodified mass)/126.1.

### Table I

| Sample            | Observed mass | Difference from unmodified protein | Alkylated cysteines per subunit |
|-------------------|---------------|-----------------------------------|---------------------------------|
|                   | m/z           |                                   |                                 |
| Unmodified VWFCK  | 27713.2       | (0)                               | (0)                             |
| NEM-VWFCK-1       | 27721.6       | 8.4                               | 0                               |
| NEM-VWFCK-2       | 14247.6       | 391                               | 3.1                             |
| NEM-VWFCK-3       | 15259.1       | 1402.5                            | 11.1                            |
| Unmodified VWFCKM | 27786.3       | (0)                               | (0)                             |
| NEM-VWFCK-M1      | 27771.9       | – 14.4                            | 0                               |
| NEM-VWFCK-M2      | 28029.7       | 243.4                             | 1.0                             |
| NEM-VWFCK-M3      | 14281.4       | 388.2                             | 0.1                             |
| NEM-VWFCK-M4      | 14794.6       | 901.4                             | 7.1                             |
| NEM-VWFCK-M5      | 14783.5       | 900.3                             | 7.1                             |
| NEM-VWFCK-M6      | 15292.3       | 1399.1                            | 11.1                            |

*NEM-VWFCK-M4 contained two major species with the indicated observed masses.

Amino acids of the propeptide (residues 1–26 of preproVWF), six histidines, an enteropeptidase cleavage sequence (NYKD-DDDK), and the C-terminal amino acid sequence of purified VWFCK or VWFCKM derivatives (Table I), whereas the predicted mass was 25729.2 Da. The difference of 1984 Da is due to glycosylation at Asn2027 because deglycosylation with TMSF reduced the mass of VWFCK to 25929.4 Da, which is close to the predicted value for the protein. After complete reduction and alkylation with NEM, the molecular mass of the VWFCK subunit was 15269.9 Da, which implies the modification of ~11.2 cysteines/subunit (1 Da per reduced cysteine + 125.1 Da per alkylation) compared with the theoretical value of 11 cysteines. Analysis of VWFCK gave similar results; the mass of dimeric VWFCKM was 27771.9 Da (Table I), and the mass of reduced and alkylated monomeric VWFCKM was 15294.8 Da.

Preparation of Monomeric VWF CK Domains by Partial Reduction and Alkylation—Potential intersubunit disulfide bonds were localized by selective reduction with TCEP to convert dimeric VWF CK domains into monomers, followed by alkylation with NEM to mark the positions of the reduced cysteines. TCEP is active at pH 4.6, and this low pH suppresses the base-catalyzed rearrangement of disulfide bonds that may occur during reduction (30, 31). NEM can be used effectively to alkylate thiols at pH 4.6 (20), and the combined use of TCEP with NEM at this pH avoided the exposure of proteins to alkaline conditions that could alter their disulfide structure. The partially reduced products obtained with VWFCK showed three major peaks upon analysis by RP-HPLC (Fig. 2A). The number of disulfide bond reduced and alkylated in each peak was determined by mass spectrometry (Table I).

The N-terminal amino acid sequence of purified VWFCK or VWFCKM was AEETHHHHHHHNYKDDDDK, demonstrating that the expected Cys-Ala bond was cleaved to release the VWF signal peptide. Results from mass spectrometry indicated that molecular mass of intact dimeric VWFCK was 27713.2 Da (Table I), whereas the predicted mass was 25729.2 Da. The difference of 1984 Da is due to glycosylation at Asn2027 because deglycosylation with TMSF reduced the mass of VWFCK to 25929.4 Da, which is close to the predicted value for the protein. After complete reduction and alkylation with NEM, the molecular mass of the VWFCK subunit was 15269.9 Da, which implies the modification of ~11.2 cysteines/subunit (1 Da per reduced cysteine + 125.1 Da per alkylation) compared with the theoretical value of 11 cysteines. Analysis of VWFCKM gave similar results; the mass of dimeric VWFCKM was 27771.9 Da (Table I), and the mass of reduced and alkylated monomeric VWFCKM was 15294.8 Da.

Preparation of Monomeric VWF CK Domains by Partial Reduction and Alkylation—Potential intersubunit disulfide bonds were localized by selective reduction with TCEP to convert dimeric VWF CK domains into monomers, followed by alkylation with NEM to mark the positions of the reduced cysteines. TCEP is active at pH 4.6, and this low pH suppresses the base-catalyzed rearrangement of disulfide bonds that may occur during reduction (30, 31). NEM can be used effectively to alkylate thiols at pH 4.6 (20), and the combined use of TCEP with NEM at this pH avoided the exposure of proteins to alkaline conditions that could alter their disulfide structure. The partially reduced products obtained with VWFCK showed three major peaks upon analysis by RP-HPLC (Fig. 2A). The number of disulfide bond reduced and alkylated in each peak was determined by mass spectrometry (Table I). The mass of NEM-VWFCK-M4 contained two major species with the indicated observed masses.

Preparation of Monomeric VWF CK Domains by Partial Reduction and Alkylation—Potential intersubunit disulfide bonds were localized by selective reduction with TCEP to convert dimeric VWF CK domains into monomers, followed by alkylation with NEM to mark the positions of the reduced cysteines. TCEP is active at pH 4.6, and this low pH suppresses the base-catalyzed rearrangement of disulfide bonds that may occur during reduction (30, 31). NEM can be used effectively to alkylate thiols at pH 4.6 (20), and the combined use of TCEP with NEM at this pH avoided the exposure of proteins to alkaline conditions that could alter their disulfide structure. The partially reduced products obtained with VWFCK showed three major peaks upon analysis by RP-HPLC (Fig. 2A). The number of disulfide bond reduced and alkylated in each peak was determined by mass spectrometry (Table I).

Preparation of Monomeric VWF CK Domains by Partial Reduction and Alkylation—Potential intersubunit disulfide bonds were localized by selective reduction with TCEP to convert dimeric VWF CK domains into monomers, followed by alkylation with NEM to mark the positions of the reduced cysteines. TCEP is active at pH 4.6, and this low pH suppresses the base-catalyzed rearrangement of disulfide bonds that may occur during reduction (30, 31). NEM can be used effectively to alkylate thiols at pH 4.6 (20), and the combined use of TCEP with NEM at this pH avoided the exposure of proteins to alkaline conditions that could alter their disulfide structure. The partially reduced products obtained with VWFCK showed three major peaks upon analysis by RP-HPLC (Fig. 2A). The number of disulfide bond reduced and alkylated in each peak was determined by mass spectrometry (Table I).

Preparation of Monomeric VWF CK Domains by Partial Reduction and Alkylation—Potential intersubunit disulfide bonds were localized by selective reduction with TCEP to convert dimeric VWF CK domains into monomers, followed by alkylation with NEM to mark the positions of the reduced cysteines. TCEP is active at pH 4.6, and this low pH suppresses the base-catalyzed rearrangement of disulfide bonds that may occur during reduction (30, 31). NEM can be used effectively to alkylate thiols at pH 4.6 (20), and the combined use of TCEP with NEM at this pH avoided the exposure of proteins to alkaline conditions that could alter their disulfide structure. The partially reduced products obtained with VWFCK showed three major peaks upon analysis by RP-HPLC (Fig. 2A). The number of disulfide bond reduced and alkylated in each peak was determined by mass spectrometry (Table I).
mass of NEM-VWFCK-2 with the mass of reduced and alkylated VWFCK (15269.9 Da) in which 11 cysteines are modified; the difference (1022.3 Da) corresponds to alkylation of 8.1 fewer cysteines. Because NEM-VWFCK-2 is monomeric, this result indicates that no more than three cysteines of VWFCK participate in intersubunit disulfide bonds.

Similar results were obtained for VWFCKM. Separation of partially reduced and alkylated products by RP-HPLC gave three major peaks and several minor peaks (Fig. 2B) like those observed for VWFCK (Fig. 2A). Analysis by mass spectrometry (Table I) indicated that NEM-VWFCK-M1 had a mass (27771.9 Da) similar to that of unmodified dimeric VWFCKM (27786.3 Da), and therefore contained no free cysteines. The mass of NEM-VWFCK-M6 (15292.3 Da) was similar to that of fully reduced and alkylated VWFCKM that was prepared independently (15294.8 Da), demonstrating that this last peak contains fully reduced and alkylated monomer. The mass of monomeric NEM-VWFCK-M3 (14281.4 Da) was consistent with the alkylation of three cysteines. The masses of the remaining peaks also correspond to integral numbers of reduced and alkylated cysteines per molecule (Table I). The results for VWFCK and VWFCKM agree and indicate that three or fewer cysteines form intersubunit disulfide bonds.

**Location of Potential Intersubunit Disulfide Bonds—NEM-VWFCK-2**

NEM-VWFCK-2 was completely reduced and alkylated with 4VP, cleaved with CNBr, and peptides were purified by RP-HPLC (Fig. 3). The amino acid sequences obtained from peaks 1–5 are shown in Table II. Peak 6 did not yield any amino acid sequence. All four expected products of CNBr cleavage were observed. Residue Cys^{2008}, Cys^{2010}, and Cys^{2048} were detected only as ES-Cys, the product of alkylation by NEM. Residue Cys^{1976}, Cys^{1987}, Cys^{2025}, and Cys^{2041} were detected only as PE-Cys, the product of alkylation by 4VP. The remaining cysteines, residues Cys^{1961}, Cys^{2011}, Cys^{2025}, Cys^{2041}, and Cys^{2043} were detected only as PE-Cys at positions Cys^{1987} and Cys^{1991} (Table III). The identification of ES-Cys at only three positions suggests that the dimerization of VWF CK domains is mediated by intersubunit disulfide bonds between one or more of the residues Cys^{2008}, Cys^{2010}, and Cys^{2048}.

The alkylation status of Cys^{1976} was not determined in this experiment because it was too far from the N terminus of the corresponding peptide and was not sequenced.

Similar studies of monomeric NEM-VWFCK-M3 confirmed that only ES-Cys was present at residues Cys^{2008}, Cys^{2010}, and Cys^{2048} and that only PE-Cys was present at residues Cys^{1961}, Cys^{1987}, Cys^{2025}, Cys^{2041}, and Cys^{2043}. In addition, only PE-Cys was found at positions Cys^{1987} and Cys^{1991} (Table III). The identification of ES-Cys at only three positions suggests that the dimerization of VWF CK domains is mediated by intersubunit disulfide bonds between one or more of the residues Cys^{2008}, Cys^{2010}, and Cys^{2048}.

**Intrasubunit Disulfide Bond Cys^{1976}–Cys^{2025} “a–c”**

Dimeric VWFCK or VWFCKM was sequentially digested with CNBr and lysyl endopeptidase. The products were separated by RP-HPLC and sequenced. Results for VWFCKM are shown in Fig. 4. Peak CL1 gave the sequences AEETHHHHHHHNYK and SEVEVD, corresponding to the N-terminal (His)6 tag sequence and VWF residues 1978–1983. Peak CL2 gave the sequences of two peptides (Table IV) consistent with expected cleavages.
after lysine and methionine (Fig. 1A). Residue Asn\textsuperscript{2027} is glycosylated (32) and was not detected. The peptides in fraction CL2 contain only two cysteine residues, and neither was detected, indicating the presence of a disulfide bond between Cys\textsuperscript{1976} and Cys\textsuperscript{2025}. Because Cys\textsuperscript{2025} is not alkylated in monomeric NEM-VWFCK-M3, this bond is an intrasubunit disulfide bond. Similar results were obtained starting with dimeric VW-FCK (data not shown).

**TABLE II**

Sequence results for CNBr-cleaved and 4VP-modified NEM-VWFCK-2 monomer

| Sample | Amino acid sequence | ES-Cys position | PE-Cys position |
|---|---|---|---|
| C1 | E\textsuperscript{2040}-Cp-K-Cp-S-P-R-K-Ce-S-K\textsuperscript{2050} | 2048 | 2041, 2043 |
| C2 | Q\textsuperscript{2020}-V-A-L-H-Cp-T\textsuperscript{2026} | | |
| C3 | A-E-E-T-H-H-H-H-H-N-Y-K | 2025 |
| C4 | Q\textsuperscript{2020}-V-A-L-H-Cp-T-[N]-G-S-V-Y-V\textsuperscript{2032} | | 2025 |
| C5 | Y\textsuperscript{1997}-S-I-D-I-N-D-V-Q-D-Q-Ce-S-Ce-Cp-S-P-T-R-T-E-P\textsuperscript{2018} | 2008, 2010 | 2011 |

**TABLE III**

Sequence results for CNBr-cleaved and 4VP-modified NEM-VWFCK-M3 monomer

| Sample | Amino acid sequence | ES-Cys position | PE-Cys position |
|---|---|---|---|
| MC1 | E\textsuperscript{2040}-Cp-K-Cp-S-P-R-K-Ce-S-S-P-R-K-Ce-S-K\textsuperscript{2050} | 2048 | 2041, 2043 |
| MC2 | H\textsuperscript{1985}-Y-Cp-Q-G-K-Cp-A-S-K-A\textsuperscript{1995} | | |
| MC3 | A-E-E-T-H-H-H-H-N-Y-K-D-D-D-D-K-E\textsuperscript{1957}-E-P-[E]-Cp-N-D\textsuperscript{1963} | 1961 |

**TABLE IV**

Identification of the Cys\textsuperscript{1976}-Cys\textsuperscript{2025} disulfide bond

Fraction CL2 (Fig. 4) was analyzed by amino acid sequencing without prior reduction. Residues are designated as described in the legend to Table II. Peak MC2 gave the same sequence data as peak MC1. Peaks MC4, MC5, and MC6 gave the same sequence data as peak MC3.

| Cycle | Residue | Residue |
|---|---|---|
| 1 | V (435.5) | Q (95.9) |
| 2 | G (283.1) | V (190.3) |
| 3 | S (109.2) | A (131.1) |
| 4 | [C\textsuperscript{2025}] | L (124.2) |
| 5 | K (257.9) | H (58.6) |
| 6 | C\textsuperscript{1976} | |
| 7 | T (67.5) | |
| 8 | [N\textsuperscript{2025}] | |
| 9 | G (53.6) | |
| 10 | S (25.4) | |
| 11 | V (37.9) | |
| 12 | V (85.4) | |
| 13 | Y (70.5) | |
| 14 | H (35.4) | |
| 15 | E (26.0) | |
| 16 | V (50.6) | |
| 17 | L (41.4) | |
| 18 | N (15.9) | |
| 19 | A (6.9) | |

**Fig. 4.** RP-HPLC separation of CNBr- and lysyl endopeptidase-digested VW-FCKM. VW-FCKM (200 μg) was digested with CNBr and dried. Then the sample was digested with lysyl endopeptidase in 0.1 M Tris-HCl, pH 7.9, 4 M urea. The sample was subjected to RP-HPLC (C18 monomeric, 4.6 × 250 mm, Vydac), in 0.1% aqueous trifluoroacetic acid and eluted with a gradient of acetonitrile. The solid line indicates the absorbance at 215 nm. The dashed line indicates the percentage of acetonitrile. Peaks CL1 and CL2 were analyzed by amino acid sequencing.

**Intrasubunit Disulfide Bond Cys\textsuperscript{1961}-Cys\textsuperscript{2011} “1–4”**—Monomeric NEM-VWFCK-M3 was digested sequentially with CNBr and thermolysin, and the products were separated by RP-HPLC (Fig. 5). Fraction MCT4 gave two amino acid sequences (Table V) that together are expected to contain four cysteine residues. One peptide corresponds to VQDQCeSCe-[C]SPTRTE (residue 2004–2017), where Ce indicates ES-Cys and the cysteine shown in brackets was not detected. The other peptide corresponds to YKDDDDKEEPE[C]Nd (enteropeptidase cleavage site and VWF residues 1957–1963). Residues Cys\textsuperscript{2008} and Cys\textsuperscript{2010} were alkylated and cannot be involved in pairing with either of the remaining cysteines. Mass spectrometry indicated that fraction MCT4 contained a single species with mass of 3496.96 Da, which is consistent with the predicted mass of 3519.43 Da for a dimer of the observed peptides. Therefore, the data indicate the presence of an intrasubunit disulfide bond between Cys\textsuperscript{1961} and Cys\textsuperscript{2011}.

**Intrasubunit Disulfide Bonds Cys\textsuperscript{1987}-Cys\textsuperscript{2041} “2–5”** and Cys\textsuperscript{1991}-Cys\textsuperscript{2043} “3–6”—The remaining two disulfide bonds were identified by differential alkylation. VW-FCKM was partially reduced with TCEP and alkylated with NEM (Fig. 2B), and fraction M4 contained two species with masses of 15042.1
or 14794.6 Da (Table I), consistent with the presence of one or two remaining intrasubunit disulfide bonds, respectively.

To locate these cysteines, fraction M4 was digested with CNBr, reduced completely with DTT, and alkylated with 4VP. The digest was separated by RP-HPLC (Fig. 5), and peak fraction MCT4 was analyzed by amino acid sequencing. Residues are designated as described in the legend for Table II. Detection of ES-Cys was not quantitated.

Identification of the Cys1961–Cys2011 (I–IV) disulfide bond

NEM-VWFCK-M3 was cleaved with CNBr, digested with thermolysin, separated by RP-HPLC (Fig. 5), and peak fraction MCT4 was analyzed by amino acid sequencing. Residues are designated as described in the legend for Table II. Detection of ES-Cys was not quantitative.

| Cycle | Residue | Residue |
|-------|---------|---------|
| 1     | V       | Y       |
| 2     | Q       | K       |
| 3     | D       | D       |
| 4     | Q       | D       |
| 5     | Ce      | D       |
| 6     | S       | D       |
| 7     | Ce      | K       |
| 8     | [C]     | E       |
| 9     | S       | E       |
| 10    | P       | P       |
| 11    | T       | E       |
| 12    | R       | [C]     |
| 13    | T       | N       |
| 14    | E       | D       |

These results allow the assignment of the two remaining intrasubunit disulfide bonds. All four cysteine residues in question were identified. Cys1987 and Cys2041 were detected exclusively as PE-Cys, whereas Cys1991 and Cys2043 were detected as both PE-Cys and ES-Cys. Monomeric NEM-VWFCK-M4 contains proteins with either one or two intrasubunit disulfide bonds (Table I), and the data of Table VI suggest that fraction M4 consists mainly of two specific proteins. Both proteins share a Cys1987–Cys2041 ‘‘2–5’’ intrasubunit disulfide bond, explaining the detection of only PE-Cys at these positions. In addition, one protein has a Cys1991–Cys2043 ‘‘3–6’’ intrasubunit bond, accounting for the presence of both PE-Cys and ES-Cys at residues Cys1991 and Cys2043.

Site-directed Mutagenesis of Cys2008, Cys2010, and Cys2048—
The role of cysteine residues at positions 2008, 2010, and 2048 in VWFCK dimerization was investigated by mutation to alanine and expression in COS-7 cells (Fig. 7). After pulse labeling, dimeric wild-type VWFCK was detected as an intracellular 34-kDa protein that was secreted into the medium within 3 h of chase. The single substitutions C2008A and C2010A, and the double mutation C2008A/C2048A, increased the electrophoretic mobility of intracellular and secreted VWFCK proteins to a position (,28 kDa) between that of dimeric VWFCK (34 kDa) and reduced VWFCK (,20 kDa) (Fig. 1). The subunit composition of these intermediate species was assessed by gel electrophoresis (Fig. 8) with VWFCK derivatives of known mass (Table I). Dimeric NEM-VWFCK1 comigrated with radiolabeled VWFCK, indicating that proteins at ,34 kDa are dimeric. As expected, the fully reduced and alkylated, monomeric NEM-VWFCKK3 comigrated with reduced radiolabeled VWFCK. Monomeric NEM-VWFCK2, in which three cysteines are reduced and alkylated, had an intermediate mobility similar to that of proteins with the mutations C2008A, C2010A, and C2008A/C2048A. The similar anomalous electrophoretic mobility of these four proteins probably is explained by decreased SDS binding compared with fully reduced VWFCK, and suggests that mutations C2008A, C2010A, or C2008A/C2048A cause the synthesis of monomers that retain some intrasubunit disulfide bonds. The mutation C2048A was compatible with the secretion of a dimeric 34-kDa species, although an intracellular band at ,28 kDa was detected that is consistent with the
Monomeric NEM-VWFCK-M4 was digested with CNBr, reduced with DTT, and alkylated with 4VP. Products were separated by RP-HPLC (Fig. 6), and peak fractions were analyzed. Residues predicted to be present from the known amino acid sequence of VWF, but not detected, are enclosed in brackets. Lowercase “m” indicates homoserine or homoserine lactone. (Fig. 6), and peak fractions were analyzed. Residues predicted to be present from the known amino acid sequence of VWF, but not detected, are enclosed in brackets. Lowercase “m” indicates homoserine or homoserine lactone.

Effect of cysteine to alanine mutations on the expression of VWFCK. COS-7 cells were transiently transfected with plasmids encoding VWFCK or mutants with the indicated cysteine to alanine substitutions. Cells were labeled with [35S]methionine for 30 min and chased with unlabeled medium for 0, 3, and 6 h. Chase medium and cell lysates were immunoprecipitated with polyclonal rabbit anti-human VWF antibody. Immunoprecipitates were separated by SDS-electrophoresis on 15% polyacrylamide gels under non-reducing conditions, stained with Coomassie Blue, dried, and subjected to autoradiography. The left side of the figure shows the autoradiographic pattern for radiolabeled VWFCK and VWFCK(C2010A). The positions of standard proteins with the indicated masses in kDa are shown without reduction (left side) and with reduction by 2.5% 2-mercaptoethanol (right side).

Electrophoretic mobility of dimeric and monomeric forms of VWFCK. Purified derivatives of VWFCK with specific numbers of alkylated cysteine residues were isolated by RP-HPLC (Fig. 3), and their mass was determined (Table I). NEM-VWFCK-1 is intact dimeric VWFCK with no cysteines alkylated. NEM-VWFCK-2 is monomeric with three cysteines alkylated. NEM-VWFCK-3 is monomeric with all 11 cysteines alkylated. COS-7 cells were transfected with plasmids encoding wild-type VWFCK or a variant with the mutation C2010A and radiolabeled with [35S]methionine. After 6 h of chase, samples of conditioned medium were immunoprecipitated with polyclonal rabbit anti-human VWF antibody. Purified proteins and immunoprecipitates were separated by SDS-electrophoresis on 15% polyacrylamide gels under non-reducing conditions (NR) or after reduction (R) with 2.5% 2-mercaptoethanol. The gel was stained with Coomassie Blue, dried, and subjected to autoradiography. The left half of the figure shows the mobility of purified proteins stained with Coomassie Blue. The right half of the figure shows the autoradiographic pattern for radiolabeled VWFCK and VWFCK(C2010A). The positions of standard proteins with the indicated masses in kDa are shown without reduction (left side) and with reduction by 2.5% 2-mercaptoethanol (right side).

For normal hemostatic function, VWF must be assembled into dimers and, subsequently, into multimers. Cysteine residues that mediate the C-terminal dimerization of VWF are within the last 151 amino acid residues of the subunit (5, 6), but their location has not been reported. The identification of a CK-like domain at the C terminus of norrin, VWF, and related epithelial mucins (28) was an important advance that focused attention on a substantially smaller segment of these proteins. CK domains comprise approximately 90 amino acids and include six cysteines arranged in a knot-like topology. Three amino acid residues including a central glycine usually separate the second and third cysteines: Cys-X-Gly-X-Cys. A single residue separates the fifth and sixth cysteines. Disulfide bonds between the second and fifth cysteines, and between the third and sixth cysteines, form a macrocyclic ring that is penetrated by a disulfide bond between the first and fourth cysteines. Many proteins with CK domains form homodimers or heterodimers, suggesting that CK domains function as dimerization motifs in several different contexts.

Studies of PSM demonstrated the potential of CK domains to mediate the oligomerization of epithelial mucins (16). PSM exhibits extensive homology to VWF and is assembled into oligomers by a similar mechanism; C-terminal domains dimerize in the endoplasmic reticulum, and N-terminal domains

### Table VI

| Sample   | Amino acid sequence                                                                 | Calculated mass (m/z) | Observed mass (m/z) | ES-Cys position | PE-Cys position |
|----------|-------------------------------------------------------------------------------------|-----------------------|---------------------|-----------------|-----------------|
| CV1      | H_{1985}Y-Cp-Q-G-K-Cp-A-S-K-A_{1995},[m]                                           | 1506.69*              | 1507.4              | 1987, 1991      |                 |
| CV2      | E_{2040}Cp-K-Cp-S-P-R-K-Ce-S-P-Ce-S_{2050},[m]                                     | 1603.85               | 1603.6              | 2048            | 2041, 2043      |
| CV3      | E_{2040}Cp-K-Ce-S-P-R-K-Ce-S-P_{2050}                                              | 1623.85               | 1622.8              | 1987, 1991      | 2043, 2048      |

* Mass of peptide with C-terminal homoserine.

# DISCUSSION

For normal hemostatic function, VWF must be assembled into dimers and, subsequently, into multimers. Cysteine residues that mediate the C-terminal dimerization of VWF are within the last 151 amino acid residues of the subunit (5, 6), but their location has not been reported. The identification of a CK-like domain at the C terminus of norrin, VWF, and related epithelial mucins (28) was an important advance that focused attention on a substantially smaller segment of these proteins. CK domains comprise approximately 90 amino acids and include six cysteines arranged in a knot-like topology. Three amino acid residues including a central glycine usually separate the second and third cysteines: Cys-X-Gly-X-Cys. A single residue separates the fifth and sixth cysteines. Disulfide bonds between the second and fifth cysteines, and between the third and sixth cysteines, form a macrocyclic ring that is penetrated by a disulfide bond between the first and fourth cysteines. Many proteins with CK domains form homodimers or heterodimers, suggesting that CK domains function as dimerization motifs in several different contexts.

Studies of PSM demonstrated the potential of CK domains to mediate the oligomerization of epithelial mucins (16). PSM exhibits extensive homology to VWF and is assembled into oligomers by a similar mechanism; C-terminal domains dimerize in the endoplasmic reticulum, and N-terminal domains

The triple mutation C2008A/C2010A/C2048A was associated with a transient 22-kDa intracellular species, and comigrated with fully reduced VWFCK (Fig. 7). Alternatively, these faster migrating species may reflect proteolytic degradation. The triple mutation C2008A/C2010A/C2048A was associated with a transient 22-kDa intracellular species, and comigrated with fully reduced VWFCK (Fig. 7). Alternatively, these faster migrating species may reflect proteolytic degradation.

Many proteins with CK domains form homodimers or heterodimers, suggesting that CK domains function as dimerization motifs in several different contexts.

Studies of PSM demonstrated the potential of CK domains to mediate the oligomerization of epithelial mucins (16). PSM exhibits extensive homology to VWF and is assembled into oligomers by a similar mechanism; C-terminal domains dimerize in the endoplasmic reticulum, and N-terminal domains

For normal hemostatic function, VWF must be assembled into dimers and, subsequently, into multimers. Cysteine residues that mediate the C-terminal dimerization of VWF are within the last 151 amino acid residues of the subunit (5, 6), but their location has not been reported. The identification of a CK-like domain at the C terminus of norrin, VWF, and related epithelial mucins (28) was an important advance that focused attention on a substantially smaller segment of these proteins. CK domains comprise approximately 90 amino acids and include six cysteines arranged in a knot-like topology. Three amino acid residues including a central glycine usually separate the second and third cysteines: Cys-X-Gly-X-Cys. A single residue separates the fifth and sixth cysteines. Disulfide bonds between the second and fifth cysteines, and between the third and sixth cysteines, form a macrocyclic ring that is penetrated by a disulfide bond between the first and fourth cysteines. Many proteins with CK domains form homodimers or heterodimers, suggesting that CK domains function as dimerization motifs in several different contexts.

Studies of PSM demonstrated the potential of CK domains to mediate the oligomerization of epithelial mucins (16). PSM exhibits extensive homology to VWF and is assembled into oligomers by a similar mechanism; C-terminal domains dimerize in the endoplasmic reticulum, and N-terminal domains

For normal hemostatic function, VWF must be assembled into dimers and, subsequently, into multimers. Cysteine residues that mediate the C-terminal dimerization of VWF are within the last 151 amino acid residues of the subunit (5, 6), but their location has not been reported. The identification of a CK-like domain at the C terminus of norrin, VWF, and related epithelial mucins (28) was an important advance that focused attention on a substantially smaller segment of these proteins. CK domains comprise approximately 90 amino acids and include six cysteines arranged in a knot-like topology. Three amino acid residues including a central glycine usually separate the second and third cysteines: Cys-X-Gly-X-Cys. A single residue separates the fifth and sixth cysteines. Disulfide bonds between the second and fifth cysteines, and between the third and sixth cysteines, form a macrocyclic ring that is penetrated by a disulfide bond between the first and fourth cysteines. Many proteins with CK domains form homodimers or heterodimers, suggesting that CK domains function as dimerization motifs in several different contexts.

Studies of PSM demonstrated the potential of CK domains to mediate the oligomerization of epithelial mucins (16). PSM exhibits extensive homology to VWF and is assembled into oligomers by a similar mechanism; C-terminal domains dimerize in the endoplasmic reticulum, and N-terminal domains
norrin, certain epithelial mucins, and VWF may constitute a subfamily of CK domains with a common structural basis for dimerization and additional intrasubunit or intersubunit disulfide bonds.

The disposition of these extra cysteine residues was addressed previously by computer modeling of norrin based on the crystal structure of TGF-β2 (28). The one extra cysteine conserved with TGF-β was proposed to form an intersubunit disulfide bond, and the four additional cysteines were paired in two intrasubunit disulfide bonds, linking Cysα–Cysδ and Cysβ–Cysγ. The observed covalent structure of VWF CK domains is consistent with this model. The core CK motif is present, as shown by the detection of conserved disulfide bonds Cysαβ–Cysβα (41), Cysγδ–Cysδγ (29) indicate that the CK domains of VWF, epithelial mucins, and (to a lesser extent) norrin differ from TGF-β2 homologs, mainly by having a markedly shorter segment corresponding to the α5 and γ3a helices of TGF-β2. This segment was modeled as a short α-helix for the VWF CK domain. The VWF CK model has the same disulfide connectivity as proposed for norrin (Fig. 10A) and includes a Cys2010–Cys2010 intersubunit disulfide bond analogous to that found in TGF-β2.

With the support of chemical evidence for the connectivity of all but three cysteine residues, we constructed a model for the dimeric VWF CK domain (Fig. 10A). Sequence alignments (Fig. 9) (29) indicate that the CK domains of VWF, epithelial mucins, and (to a lesser extent) norrin differ from TGF-β2 homologs, mainly by having a markedly shorter segment corresponding to the α5 and γ3a helices of TGF-β2. This segment was modeled as a short α-helix for the VWF CK domain. The VWF CK model has the same disulfide connectivity as proposed for norrin (Fig. 10A) and includes a Cys2010–Cys2010 intersubunit disulfide bond analogous to that found in TGF-β2.

This pattern is compatible with the results of our structural studies, which demonstrate the presence of a CK motif and limit the intersubunit disulfide bond candidates to Cys2008, Cys2010, and Cys2048. To avoid the occurrence of unpaired cysteine residues, which are not observed in VWF or recombinant VWFCK, the dimerization of the VWF CK domain must involve either one or three intersubunit disulfide bonds. However, the formation of two intersubunit Cys2008–Cys2048 bonds would require the dimer interface to deviate radically from that of TGF-β, and for this reason the model has a single intersubunit disulfide bond.

VWF CK dimers with three intersubunit disulfide bonds can be constructed easily if larger structural changes are accepted, particularly if the pairing of Cys2010 is altered. Because all cysteines are oxidized, such a dimer must include one pairing between the same cysteine residues in both subunits. There are four such patterns. One contains Cys2008–Cys2010, Cys2010–Cys2048, and Cys2048–Cys2048 bonds. The others correspond to the three structurally distinct combinations of Cysα–Cysδ, Cysβ–Cysγ, where Cysβ is Cys2008, Cys2010, or Cys2048. For example, the shift of one subunit by 5 Å enables the formation of two Cys2008–Cys2010 bonds (Fig. 10B). The (Cys2008–X–Cys2010)2 arrangement is similar to the antiparallel Cys-X-Cys/Cys-X-Cys cytochrome framework found in several influenza virus neuraminidases and in porcine leukocyte protegrin-1 (reviewed in Ref. 33). Completion of the structure requires the formation of a Cys2048–Cys2048 bond, which is accommodated easily by...
The various possible models for CK dimer structure might be distinguished by mutagenesis of cysteine residues, but this approach has been only partially successful. For the VWF CK domain, mutation of Cys\textsuperscript{2048} permitted the secretion of dimers and suggested the presence of one or two intersubunit disulfide bonds involving Cys\textsuperscript{2008}, Cys\textsuperscript{2010}, or both residues. However, no mutation affecting either Cys\textsuperscript{2008} or Cys\textsuperscript{2010} allowed the secretion of dimers, and the data therefore do not exclude a Cys\textsuperscript{2008}–Cys\textsuperscript{2010} intersubunit bond. Similar mutagenesis studies of the PSM CK domain are consistent with more than one intersubunit disulfide bond but do not identify their locations (16). In this case, single alanine substitutions for Cys\textsuperscript{13244} (Cys\textsubscript{b}), Cys\textsuperscript{13246} (Cys\textsubscript{b}), or Cys\textsuperscript{13283} (Cys\textsubscript{d}) caused the secretion mainly of PSM CK dimers, indicating that no one of these cysteines is necessary for intersubunit disulfide bond formation. Therefore, the available structural and mutagenesis data suggest that CK domains like those of VWF and PSM, which possess 11 conserved cysteine residues, may constitute a structural family of CK domains with a novel pattern of three intersubunit disulfide bonds. Although norrin has the same 11 conserved cysteines, it may represent a special case in which the same cysteine residues (Cys\textsubscript{b}, Cys\textsubscript{a}, and Cys\textsubscript{d}) form disulfide bonds among three rather than two monomers, resulting in the assembly of multimers rather than dimers.

Further study is needed to test this hypothesis. For dimeric proteins like the VWF CK domain, analytical methods that rely on fragmentation generally cannot distinguish intersubunit from intrasubunit disulfide bonds. Because we have not been able to isolate a monomeric VWF CK species with fewer than three reduced and alkylated cysteines, complete characterization of the disulfide bonds involving Cys\textsuperscript{2008}, Cys\textsuperscript{2010}, and Cys\textsuperscript{2048} probably will require a different methodology, such as x-ray crystallography.

Acknowledgments—We thank Dr. Mitsu Haniu (Amgen) for many helpful discussions on the determination of disulfide bond connectivity. We also thank Dr. Steve Smith (University of Texas at Galveston), Dr. Clive Slaughter and Carolyn Moomaw (University of Texas Southwestern), and Dr. Kenneth Walsh and Santosh Kumar (University of Washington, Seattle) for their outstanding help with protein sequencing and mass spectrometry.

REFERENCES
1. Sadler, J. E. (1998) Annu. Rev. Biochem. 67, 395–424
2. Jaffe, E. A., Hoyer, L. W., and Nachman, R. L. (1973) J. Clin. Invest. 52, 2757–2764
3. Nachman, R., Levine, R., and Jaffe, E. A. (1977) J. Clin. Invest. 60, 914–921
4. Sadler, J. E. (1991) J. Biol. Chem. 266, 22777–22780
5. Marti, T., Ronselet, S. J., Tiwani, R., and Walsh, K. A. (1987) Biochemistry 26, 8099–8109
6. Voorberg, J., Fontijn, R., Calafat, J., Jansen, H., van Mourik, J. A., and Pannekoek, H. (1991) J. Cell Biol. 115, 195–205
7. Isaacs, N. W. (1995) Curr. Opin. Struct. Biol. 5, 391–395
8. Enayat, M. S., Surdhar, G. K., Theophilus, B. D. M., Williams, D. M., and Hill, F. G. H. (1997) Thorab. Haemost. Suppl. 82, 655
9. Schneppenheim, R., Brasward, J., Krey, S., Budde, U., Kunicki, T. J., Holmberg, L., Ware, J., and Ruggeri, Z. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 3581–3586
10. Montgomery, R. R., Jozwiak, M. A., Hutter, J. J., Endres, J. L., Foster, P. A., and Friedman, K. D. (1999) Blood 94, Suppl. 1, 443a
11. Schneppenheim, R., Brasward, J., Obser, T., Brasward, E., Dweke, E., Krey, S., Oldenburg, J., Ruggeri, Z. M., Schwaab, R., and Ware, J. (1999) Blood 94, Suppl. 1, 453a
12. Chen, Z. Y., Hendriks, R. W., Jobling, A. M., Powell, J. F., Baekfle dean, X. O., Sims, K. B., and Craig, I. W. (1992) Nat. Genet. 1, 204–208
13. Eckhard, A. E., Timpte, C. S., DeLuca, A. W., and Hill, R. L. (1997) J. Biol. Chem. 272, 32304–32310
14. Lesuffleur, T., Roche, F., Hill, A. S., Lacaas, M., Fox, M., Swallow, D. M., Zweibaum, A., and Real, F. X. (1995) J. Biol. Chem. 270, 13665–13673
15. Perez-Vilar, J., and Hill, R. L. (1997) J. Biol. Chem. 272, 33410–33415
16. Perez-Vilar, J., and Hill, R. L. (1998) J. Biol. Chem. 273, 6982–6988
17. Randi, A. M., Jorieux, S., Tuley, E. A., Mazurier, C., and Sadler, J. E. (1995) J. Biol. Chem. 270, 21187–21192
18. Cormack, B. (1997) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) pp. 8.5.1–8.5.10, John Wiley & Sons, New York
19. Lu, D., Yuan, X., Zheng, X., and Sadler, J. E. (1997) J. Biol. Chem. 272, 31293–31300
20. Bures, R. J., Hai, J. O., Young, Y., Chow, D. T., Katta, V., Rohde, M. F., Zani, et al.
21. Haniu, M., Hui, J., Young, Y., Le, J., Katta, V., Lee, R., Shimamoto, G., and Rohde, M. F. (1996) *Biochemistry* **35**, 16799–16805

22. Haniu, M., and Arakawa, M. (1997) *Curr. Top. Pept. Protein Res.* **2**, 115–124

23. Kaufman, R. J., Davies, M. V., Wasley, L. C., and Michnick, D. (1991) *Nucleic Acids Res.* **19**, 4485–4490

24. Katsumi, A., Matsushita, T., Yamazaki, T., Suguiro, I., Kojima, T., and Saiito, H. (2000) *Thromb. Haemost.* **83**, 238–243

25. Katsumi, A., Senda, T., Yamashita, Y., Yamazaki, T., Hamaguchi, M., Kojima, T., Kobayashi, S., and Saiito, H. (1996) *Blood* **87**, 4164–4175

26. Daspin, S., Li, M., and Davies, D. R. (1993) *Proteins* **17**, 176–192

27. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950

28. Mettig, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M., and Murken, J. (1999) *Curr. Top. Pept. Protein Res.* **2**, 115–124

29. Perry-Vilar, J., and Hill, R. L. (1999) *J. Biol. Chem.* **274**, 31751–31754

30. Gray, W. R. (1993) *Protein Sci.* **2**, 1732–1748

31. Gray, W. R. (1993) *Protein Sci.* **2**, 1749–1755

32. Tiana, K., Kumar, S., Takio, K., Ericson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., and Fujikawa, K. (1986) *Biochemistry* **25**, 3171–3184

33. Tamaoki, H., Miura, R., Kasunoki, M., Kyogoku, Y., Kobayashi, Y., and Moroder, L. (1998) *Protein Eng.*, **11**, 649–659

34. Schlunegger, M. P., and Grutter, M. G. (1993) *J. Mol. Biol.* **231**, 445–458