THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 49, Issue of December 7, pp. 45882–45887, 2001
Printed in U.S.A.

Disulfide Connectivity of Recombinant C-terminal Region of Human Thrombospondin 2*

Received for publication, May 9, 2001, and in revised form, September 12, 2001
Published, JBC Papers in Press, October 4, 2001, DOI 10.1074/jbc.M104218200

Tina M. Misenheimer‡§, Allison J. Hahr‡, Amy C. Harms¶, Douglas S. Annis‡, and Deane F. Mosher‡

From the ‡Department of Medicine and the ¶Biotechnology Center, University of Wisconsin, Madison, Wisconsin 53706

The thrombospondin (TSP) family of extracellular glycoproteins consists of five members in vertebrates, TSP1 to -4 and TSP5/cartilage oligomeric matrix protein, and a single member in *Drosophila*. TSPs are modular multimeric proteins. The C-terminal end of a monomer consists of 3–6 EGF-like modules; seven tandem 23-, 36-, or 38-residue aspartate-rich, Ca^{2+}-binding repeats; and an ~230-residue C-terminal sequence. The Ca^{2+}-binding repeats and C-terminal sequence are spaced almost exactly the same in different TSPs and share many blocks of identical residues. We studied the C-terminal portion of human TSP2 from the third EGF-like module through the end of the protein (E3CaG2). E3CaG2, CaG2 lacking the EGF module, and Ca2 composed of only the Ca^{2+}-binding repeats were expressed using recombinant baculoviruses and purified from conditioned media of insect cells. As previously described for intact TSP1, E3CaG2 bound Ca^{2+} in a cooperative manner as assessed by equilibrium dialysis, and its circular dichroism spectrum was sensitive to the presence of Ca^{2+}. Mass spectrometry of the recombinant proteins digested with endoproteinase Asp-N revealed that disulfide pairing of the 18 cysteines in the Ca^{2+}-binding repeats and C-terminal sequence is sequential, *i.e.* a 1–2, 3–4, 5–6, etc., pattern.

The thrombospondin (TSP)[1] family of extracellular glycoproteins consists of five members in vertebrates, TSP1–4 and TSP5/cartilage oligomeric matrix protein (COMP) (1–5), and a single member in *Drosophila*, dTSP (6, 7). The temporal and spatial expression pattern of the TSP family members in vertebrates is distinctive (8–10). TSP1 and TSP2 are structurally similar homotrimeric proteins composed of three identical 150-kDa monomers connected by disulfide bridges (11). Each monomer contains an N-terminal globular module, followed by an oligomerization domain that has the cysteines that form the interchain disulfide linkages, a procollagen module, three type 1 or properdin modules, three type 2 or epidermal growth factor (EGF)-like modules, and seven tandem aspartate-rich repeats followed by a nonrepeating C-terminal sequence. TSP3, TSP4, and TSP5/COMP are pentamers of subunits that are composed of the seven aspartate-rich repeats and C-terminal sequence but lack procollagen and properdin modules and contain an extra EGF-like module. Four or six EGF-like modules are present in dTSP, which otherwise is most like vertebrate TSP5/COMP (6, 7). The presence of such a nonvertebrate homolog had been predicted based on sequence comparison of various vertebrate TSPs (12). Thus, the aspartate-rich repeats, which total 252–257 residues, and the C-terminal sequence of ~230 residues are common to all TSPs and have been extraordinarily well conserved.

The presence of Ca^{2+} alters the structure and function of TSPs. Rotary shadowing has demonstrated that, in the presence of Ca^{2+}, a globule composed of the C-terminal portion of TSP1 enlarges while the stalk connecting the oligomerization domain to the C-terminal globule shortens (13–15). This structural change was also observed for TSP3 (16), TSP4 (17), and TSP5/COMP (18). The structure of TSP2 is also Ca^{2+}-sensitive based on its susceptibility to trypsin proteolysis (19). Sedimentation velocity experiments of TSP1 (13) and recombinant Ca^{2+}-binding repeats of TSP5/COMP (20) revealed an increase in the sedimentation coefficient upon the addition of Ca^{2+}, which is consistent with a change in structure upon Ca^{2+} binding. Adhesion of cells to TSP1 (21, 22) and TSP2 (19) is Ca^{2+}-dependent, as is TSP1 inhibition of cathepsin G (23) and neutrophil elastase (24). TSP1 has been shown to interact with Ca^{2+} cooperatively (25, 26) and to bind 35 ± 3 Ca^{2+}/trimer (~12 Ca^{2+}/monomer) (26). Similarly, TSP5/COMP binds 11 Ca^{2+} per monomer (18). The presumptive Ca^{2+}-binding consensus sequence D*X*D*X*D*XX*GD*XX*D*X* occurs 12 times in a TSP monomer with only minor variations. Sequence comparison reveals that these aspartate-rich repeats are well conserved among TSP1, TSP2 (1, 2), TSP3 (3), TSP4 (4), TSP5/COMP (5), and dTSP (6), with the proposed Ca^{2+}-binding amino acids (Asp/Asn) being nearly always conserved. Mutations in the Ca^{2+}-binding region of TSP5/COMP have been identified in patients with two syndromes of skeletal dysplasia, pseudoachondroplasia and multiple epiphyseal dysplasia (27, 28). Recently, recombinant full-length TSP5/COMP and the Ca^{2+}-binding region of TSP5/COMP incorporating two of these TSP5/

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.This work was supported by National Institutes of Health Grant HL54462 and utilized the University of Wisconsin-Madison Biophysics Instrumentation Facility, which is supported by National Science Foundation Grant BIR-9512577.

§ To whom correspondence should be addressed: Dept. of Medicine, University of Wisconsin, 1300 University Ave., Madison, WI 53706. Tel.: 608-262-1375; Fax: 608-263-4969; E-mail: tmmisenh@facstaff.wisc.edu.

[1] The abbreviations used are: TSP, thrombospondin; hTSP, human TSP; Ca2, recombinant Ca^{2+}-binding repeats of TSP2 (residues 693–947; Fig. 1); Ca2 tun, recombinant Ca2 purified from tunicamycin-treated infected cells; CaG2, recombinant C-terminal portion of TSP2 from the Ca^{2+}-binding repeats to the end of TSP2 (residues 693–1172, Fig. 1); COMP, cartilage oligomeric matrix protein; EGF, epidermal growth factor; E3CaG2, recombinant C-terminal portion of TSP2 from the third EGF-like module to the end of TSP2 (residues 650–1172, Fig. 1); ESI, electrospray ionization; G2, recombinant C-terminal globule of TSP2 (residues 947–1172); LC, liquid chromatography; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; MS, mass spectrometry; tCaG2, recombinant C-terminal portion of hTSP2 from within the Ca^{2+}-binding repeats to the end of TSP2 (residues 718–1172, Fig. 1); HPLC, high pressure liquid chromatography; MOPS, 4-morpholinepropanesulfonic acid.

```
ADP(650)NPCKDKT HNCHKHAECI YLGHFSDPMY KCECQTGYAG DGLICG(692)
EDSDLDGWPNLNLVCATNATYHCIKDN (719)
↑                          ⇑
CPHLPNSGQEDFDKDGIGDACDDDDDNDGVTDEKDN
CQLLFNPRQADYDKDEVGDRCDN
CPYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERDN
CPYVYNTDQRDTDGDGVGDHCDN
CPLVHNPDQTDVDNDLVGDQCDNNEDIDDDGHQNNQDN
CPYISNANQADHDRDGQGDACDPDDDNDGVPDDRDN
CRLVFNPDQEDLDGDGRGDICKDDFDNDNIPDIDDV (947)
                                   ↑
CPENNAISET DFRNFQMVPL DPKGTTQIDP NWVIRHQGKE LVQTANSDPG
IAVGFDEFGS VDFSGTFYVN TDRDDDYAGF VFGYQSSSRF YVVMWKQVTQ
TYWEDQPTRA YGYSGVSLKV VNSTTGTGEH LRNALWHTGN TPGQVRTLWH
DPRNIGWKDY TAYRWHLTHR PKTGYIRVLV HEGKQVMADS GPIYDQTYAG
GRLGLFVFSQ EMVYFSDLKY ECRDI(1172)NAGGH HHHHH
```

FIG. 1. **Sequence of E3CaG2 and landmarks mentioned in this paper.** Residues 650–692 (numbering starting from the initiating methionine) comprise the third EGF-like repeat, 720–947 comprise the aspartate-rich Ca^{2+}-binding repeats, and 948–1172 comprise the nonrepeating C-terminal sequence of TSP2. The short tails are coded in the expression vector. Ca2 is composed of residues 693–947 of TSP2 (start and stop indicated by *thin* and *thick upward arrows*, respectively). CaG2 is composed of residues 693–1172 of TSP2 (start indicated by an *upward arrow*), while tCaG2 is composed of residues 718–1172 of TSP2 (start indicated by an *open upward arrow*). The seven Ca^{2+}-binding repeats are aligned to emphasize the sequentially paired disulfides predicted by our data. Aspartate residues that are potential targets for endoproteinase Asp-N are in *boldface type*. The 24 cysteines, the serine and valine residues that are cysteine in other TSPs, the eight tryptophans, and the consensus sequences for *N*-linked glycosylation are *underlined*.

COMP mutations have been expressed and compared with the wild type proteins in mammalian expression systems. Comparison of the Ca^{2+} binding properties of the wild type and mutant proteins revealed a significantly lower Ca^{2+} binding capacity in the mutant proteins (18, 20, 29).

In this report, we describe the expression, biophysical properties, and disulfide connectivity of E3CaG2, the carboxyl-region of TSP2 beginning at the third EGF-like module. We chose to concentrate on E3CaG2 because TSP2 is the only TSP with an even number of cysteines in the C-terminal region. The other TSPs have an additional unpaired cysteine, and TSP1 undergoes extensive thiol-disulfide isomerization in the C-terminal region (30). Like trimeric TSP1 (25, 26), monomeric E3CaG2 bound Ca^{2+} in a cooperative manner, and binding of Ca^{2+} affected the structure of the protein as measured by far UV CD. Mass spectrometry of E3CaG2 digested with endoproteinase Asp-N revealed that the disulfide pairing pattern of the Ca^{2+}-binding repeats and C-terminal sequence is sequential (*i.e.* a 1–2, 3–4, 5–6, etc., pattern). The same pairing patterns were observed in constructs lacking the EGF module or both the EGF module and the C-terminal 230 residues.

MATERIALS AND METHODS

Expression of E3CaG2, Ca2, CaG2, tCaG2, and G2—DNA encoding the third EGF-like domain through the C terminus of TSP2 (residues 650–1172 of full-length TSP2, Fig. 1), E3CaG2, was polymerase chain reaction-amplified from a full-length clone of human TSP2 and inserted into the *Bam*HI and *Pst*I sites (using the compatible cohesive end of *Nsi*I in the polymerase chain reaction product) of the pAcGP67.coco baculovirus transfer vector developed in our laboratory. This vector encodes a signal peptide 5′ to and a six-histidine tag (His tag) 3′ to the cloning site (31). Ca2 (residues 693–947), CaG2 (residues 693–1172), tCaG2 (residues 718–1172), and G2 (residues 947–1172) were made using the E3CaG2 plasmid for a further round of polymerase chain reaction to allow insertion of truncated constructs into pAcGP67.coco. The sites for Ca2 in pAcGP67.coco were *Bam*HI (5′) and *Pst*I (3′), while the sites for G2 were *Xba*I (5′) and *Pst*I (3′). CaG2 and tCaG2 utilized the 5′ *Bam*HI cloning site in pAcGP67.coco and a naturally occurring *Bam*HI site in the C-terminal globe region. The polymerase chain reaction products were digested with *Bam*HI and cloned into *Bam*HI-digested G2. Correct orientation and sequence of polymerase chain reaction-amplified DNA were verified in all cases by automated sequencing.

Protein Purification Using a Baculovirus Expression System and Nickel Chelate Chromatography—E3CaG2, Ca2, CaG2, tCaG2, and G2 were expressed by infecting High Five insect cells (Invitrogen) in SF900II serum-free medium at 27 °C with high titer virus (>10^8 plaque-forming units/ml) at a multiplicity of infection of 5. Conditioned medium was collected ~65 h postinfection, cells were spun down and removed, and sodium azide to 0.5% and phenylmethylsulfonyl fluoride to 2 mM were added. The medium was dialyzed into 10 mM Tris, 300 mM NaCl, 0.3 mM $CaCl_2$, pH 7.4, or 10 mM MOPS, 300 mM NaCl, 2 mM $CaCl_2$, pH 7.5, and then incubated with Ni^{2+}-nitrilotriacetic acid resin (Qiagen) overnight at 4 °C. The resin was washed with buffer containing 5–15 mM imidazole, and protein was eluted with buffer containing 300 mM imidazole. Purified protein was dialyzed into Tris or MOPS buffer containing 150 mM NaCl and 0.3 or 2 mM $CaCl_2$, stored in portions at −80 °C, and thawed in a bath at 25 °C prior to biochemical studies.

Yield was 80, 48, 18, or 14 mg/liter of conditioned medium for E3CaG2, Ca2, CaG2, or tCaG2, respectively. No detectable G2 could be found in conditioned medium. Protein concentration was determined using absorbance at 280 nm and a calculated extinction coefficient of 1.33, 0.524, 1.36, and 1.29 ml/mg cm for E3CaG2, Ca2, CaG2, and tCaG2, respectively (32). Mass was determined by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS) with a Bruker Biflex III at the University of Wisconsin Biotechnology Center Mass Spectrometry/Bioanalytical facility.

To study Ca2 in the absence of *N*-glycosylation, cells were treated with 5 μg/ml tunicamycin (ICN Biomedicals, Inc.) 24 h postinfection. Medium was harvested 48 h after tunicamycin treatment, and Ca2 tun (Ca2 purified from tunicamycin-treated infected cells) was purified as usual.

Disulfide Connectivity—E3CaG2, CaG2, tCaG2, or Ca2 was digested with endoproteinase Asp-N (Roche Molecular Biochemicals) (1:100, w/w) in 50 mM ammonium acetate, pH 6.1, at 37 °C for 4 h. The reaction was quenched with 1 mM EDTA and analyzed by liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS) at the University of Wisconsin Biotechnology Center Mass Spectrometry/Bioanalytical facility using C4 reverse-phase HPLC and a Micromass Q-TOF2 spectrometer.

Circular Dichroism—CD studies were performed on an Aviv CD spectrometer, model 62A DS. Far UV measurements were taken in a 0.1-cm path length cuvette using protein dialyzed into 1 mM K_2HPO_4, 85 mM NaCl, 0.3 mM $CaCl_2$, pH 7.4. The resulting spectra were converted to mean residue weight ellipticity and smoothed using Aviv's Igor Pro software.

Calcium Binding—The ability of purified E3CaG2 to bind Ca^{2+} was measured directly using equilibrium dialysis with $^{45}CaCl_2$ (Amersham Pharmacia Biotech) as described previously (26). Briefly, E3CaG2 was dialyzed into 10 mM Tris, 150 mM NaCl, pH 7.4 (TBS), containing 0–0.3 mM $CaCl_2$ plus $^{45}CaCl_2$ for 4 h at 4 °C. After dialysis, protein and buffer samples were analyzed for radioactivity, and protein concentration was determined using the BCA protein assay (Pierce) with bovine albumin as the standard.

Intrinsic Fluorescence—Fluorescence emission spectra of E3CaG2 in TBS plus 0.3 mM $CaCl_2$ were obtained at 25 °C with an SLM-8000C fluorimeter with excitation at 295 nm. EDTA to 0.4 mM was added as needed to chelate calcium. Three scans were averaged for each measurement and base line-subtracted.

Other Methods—N-terminal sequencing was generously provided by Johan Stenflo (University of Lund, Sweden). The presence of glycosylation was tested using a GelCode glycoprotein staining kit (Pierce) as directed by the manufacturer. Ellman's reagent, 5,5′-dithiobis-(2-nitrobenzoic acid), was used to assay for the presence of free thiols (33).

RESULTS

Expression and Characterization of E3CaG2—The recombinant baculovirus directed High Five cells to express and secrete large quantities of E3CaG2. The yield was ~80 mg/liter conditioned medium. Purified protein was kept in 0.3 or 2 mM $CaCl_2$ to maintain stability. The concentration of purified protein was usually about 5 mg/ml, but concentrations up to 10 mg/ml have been achieved without aggregation. N-terminal sequencing verified that E3CaG2 starts with ADPNP, indicating that the signal sequence derived from viral GP67 was cleaved at the predicted site. Reactivity with Ellman's reagent indicated that less than 0.1 mol of free thiol/mol of E3CaG2 was present. Under reducing and nonreducing conditions, the protein migrated relative to markers on SDS-polyacrylamide gel electrophoresis at ~90 and 60 kDa, respectively (Fig. 2). We surmise

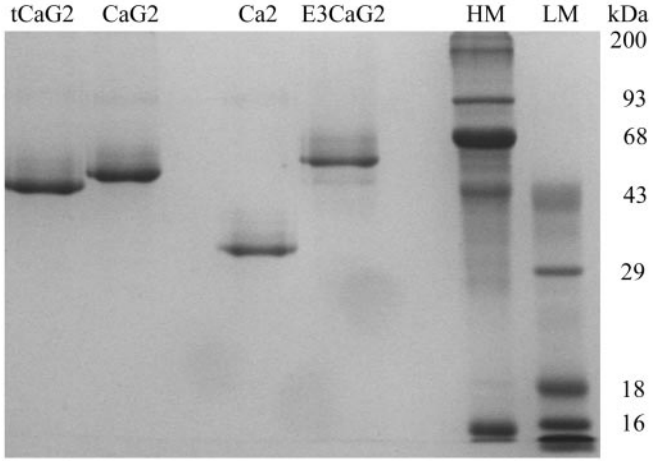

FIG. 2. **SDS-polyacrylamide gel electrophoresis of E3CaG2, CaG2, tCaG2, and Ca2.** 12% polyacrylamide gels of nonreduced E3CaG2, CaG2, tCaG2, Ca2, and molecular weight markers (high, HM, and low, LM) were run and stained with Gel-Code Blue.

that the protein migrated more rapidly when nonreduced due to the presence of constraining disulfides that prevent the protein from adapting an extended SDS-protein complex. The anomalous apparent size of reduced protein is probably due to the unusual amino acid composition (17% aspartate). The mass determined by MALDI-TOF MS was 62,122 Da. The difference between 62,122 Da and the expected mass of 60,150 Da can be explained by *N*-linked glycosylation of E3CaG2 at consensus sites at residues 710 and 1069. E3CaG2 stained positively with the GelCode glycoprotein staining kit. Treatment of infected cells with tunicamycin, an inhibitor of *N*-glycosylation, caused greatly reduced secretion of a protein of 60,109 Da. These results indicate that E3CaG2 is indeed glycosylated.

To compare the structure of E3CaG2 with TSP1 purified from platelets, the protein was analyzed by CD spectroscopy in the presence and absence of Ca^{2+}. The far UV CD spectra were similar in shape to those reported for full-length TSP1 (25), with one negative band at 202 nm that was more negative in the presence of Ca^{2+} than in the presence of EDTA (data not shown). The SOPM program for secondary structure prediction (34, 35) predicts 7% α-helix and 33% β-structure for E3CaG2, with the α-helix and β-structure localized predominantly in the C-terminal sequence. Deconvolution of the CD spectrum of Ca^{2+}-replete E3CaG2 using the K2d program (36, 37) suggests 8% α-helix and 41% β-structure. In the presence of EDTA, the CD spectrum predicts 8% α-helix and 45% β-structure.

The environments around the eight tryptophans of E3CaG2 were analyzed by intrinsic fluorescence spectroscopy. The emission spectrum of E3CaG2 after excitation at 295 nm had a wavelength of maximum fluorescence (λ_{max}) of 337 nm in buffer containing 0.3 mM $CaCl_2$, and upon the addition of 0.4 mM EDTA, λ_{max} consistently shifted to 340 nm (data not shown). These findings indicate that the average tryptophan environment is nonpolar in the presence of Ca^{2+}, and the addition of EDTA increases the polarity of the average environments.

Ca^{2+} binding to E3CaG2 was quantified by equilibrium dialysis at 4 °C with 0–0.3 mM $CaCl_2$ plus $^{45}CaCl_2$. E3CaG2 maximally bound 9.1 ± 0.7 Ca^{2+} with an apparent K_d of 51 μM and a Hill coefficient of 1.7 (data not shown). This K_d is similar to the transition midpoints of 45, 50, and 52 μM reported previously for Ca^{2+}-dependent tryptic susceptibility of TSP1 (25), Ca^{2+}-dependent monoclonal antibody binding to TSP1 (38), and equilibrium Ca^{2+} binding to TSP1 (26), respectively. These K_d values all were determined at 0–4 °C.

Taken together, these data indicate that recombinant

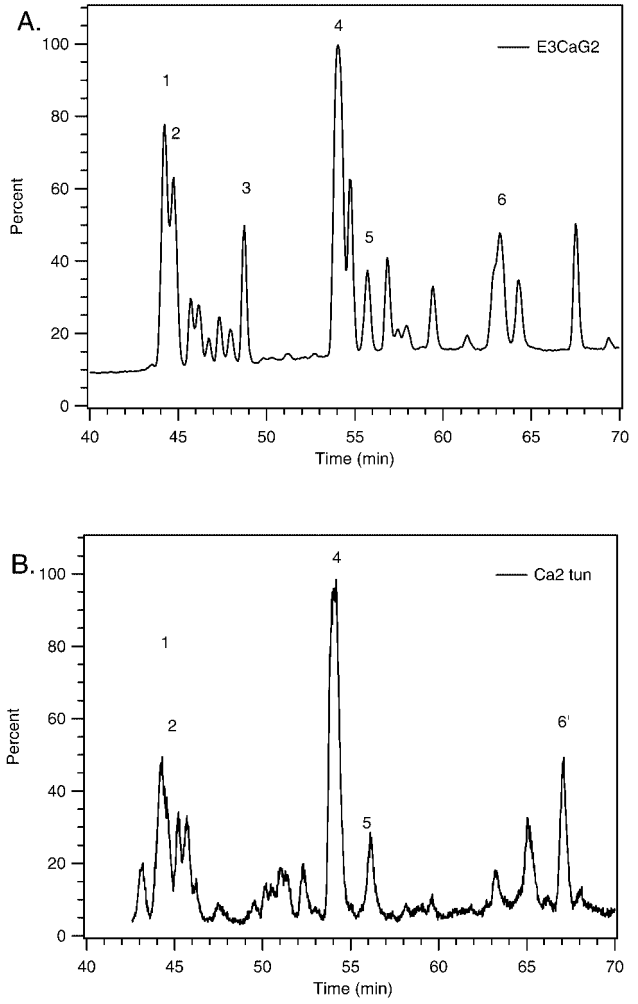

FIG. 3. **Disulfide connectivity of E3CaG2 and Ca2 tun.** HPLC profiles are shown from endoproteinase Asp-N digests of E3CaG2 (*A*) and Ca2 tun (*B*). *Peaks labeled 1–6* and *6′* contained mass peaks corresponding to the disulfide-linked peptides listed in Table I.

E3CaG2 is a Ca^{2+}-binding protein with a Ca^{2+}-sensitive structure that is similar to the structure of the homologous C-terminal region of TSP1 purified from human platelets.

Disulfide Connectivity—The disulfide connectivity in E3CaG2 was determined by cleavage with endoproteinase Asp-N followed by analysis via LC/ESI-MS using a Micromass Q-TOF2 spectrometer. Due to the large number of aspartate residues (Fig. 1), endoproteinase Asp-N can potentially cleave E3CaG2 into 92 peptides, including 17 that contain the cysteines in the aspartate-rich region and C-terminal sequence of TSP2. Potential Asp-N cleavage sites occur between all pairs of cysteine residues except for Cys^{707} and Cys^{715}. The LC profile of the digest was complex (Fig. 3*A*). Tabulation of the possible disulfide arrangements and the corresponding masses was made and compared with observed masses within each peak. Peak 1 contained two peptides that correspond to peptides disulfide-linked by Cys^{815} and Cys^{835} and by Cys^{838} and Cys^{858} (Figs. 3 and 4; Table I). Peak 2 contained two peptides that correspond to peptides disulfide-linked by Cys^{720} and Cys^{740} and by Cys^{876} and Cys^{896}. Peak 3 contained peptides that correspond to peptides disulfide-linked by Cys^{948} and Cys^{1170}. Peak 4 contained two peptides that correspond to peptides disulfide-linked by Cys^{756} and Cys^{776} and by Cys^{779} and Cys^{799}. Peak 5 contained peptides that correspond to peptides disulfide-linked by Cys^{912} and Cys^{932}. The only observed mass that was potentially due to different pairs was 1613.6. This mass

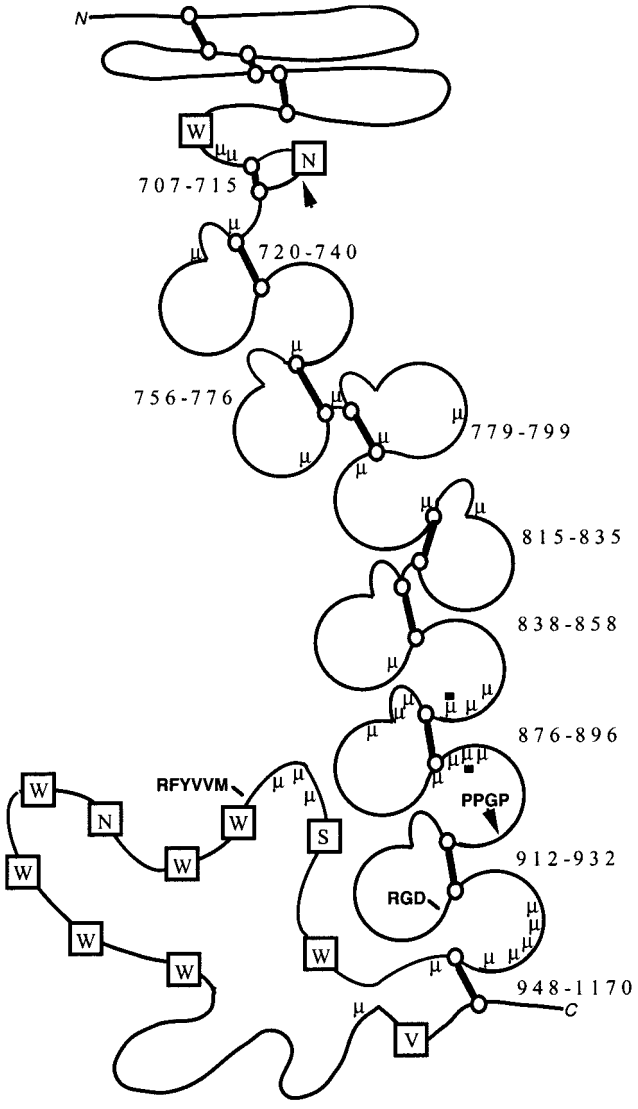

FIG. 4. **Schematic diagram of *E3CaG*.** The model of E3CaG2 shows the consecutive disulfide pairing of the Ca^{2+}-binding repeats and C-terminal sequence as deduced by MS. The *open circles* represent the cysteines, and the *connecting thick lines* represent the disulfide bonds. The residue numbers of the cysteines in the Ca^{2+}-binding repeats and C-terminal sequence are noted. Disulfides of the N-terminal *E3* module are depicted with the 1–3, 2–4, 5–6, pattern observed in other EGF-like modules. The *arrowheads* point out a site of *N*-glycosylation that is deleted in TSP3, TSP4, TSP5/COMP, and dTSP and an insertion of PPGP that occurs in TSP3 and TSP4. *Boxed* are the locations of the eight tryptophan residues (*W*), two potential *N*-linked glycosylation sites (*N*), the serine (*S*) in TSP2 that is a cysteine in TSP1, and the valine (*V*) in TSP2 that is a cysteine residue in TSP3, TSP4, TSP5/COMP, and dTSP. The locations of the reported mutations in TSP5/COMP that are found in patients with pseudoachondroplasia/multiple epiphyseal dysplasia are indicated by μ. The *filled boxes* indicate the sites of mutations in TSP5/COMP that have been studied as recombinant proteins and are discussed in this paper. The RGD and RFYVVM binding sites for integrins and CD47 are also indicated.

was assigned to peptides linked by Cys^{876} and Cys^{896} (expected mass 1613.6), but if Asp^{741} was lost from peptide Asp^{738}–Asp^{741}, the mass observed could be due to peptides linked by Cys^{740} and Cys^{876} (expected mass 1613.6). However, peptides linked by Cys^{720} and Cys^{896} would then be expected, but the expected mass of 1614.6 was not observed. Also, disulfide-linked pairing of residues Asp^{718}–Glu^{729} and Asp^{738}–Asp^{741} is the only peptide that fits the observed mass of 1729.6. Thus, these data indicate that consecutive cysteines are connected together: Cys^{720}–Cys^{740}, Cys^{756}-Cys^{776}, Cys^{779}–Cys^{799}, Cys^{815}–Cys^{835}, Cys^{838}–Cys^{858}, Cys^{876}–Cys^{896}, Cys^{912}–Cys^{932}, and Cys^{948}–Cys^{1170} (Table I, Fig. 4). By exclusion, Cys^{707} must pair with Cys^{715} and be present in Asp^{698}–Lys^{717} without a cleavage site for endoproteinase Asp-N. The unmodified mass of this peptide is 2230.0, but the peptide contains one of the consensus sequences for *N*-linked glycosylation and therefore is of unknown mass. A peptide with a mass of 3252.4 was present in peak 6 (Fig. 3, Table I). This mass is 1022.4 greater than the mass of 2230 expected for unglycosylated peptide Asp^{698}–Lys^{717} and is consistent with a glycosylated form of Asp^{698}–Lys^{717}. *N*-Glycosylation with $mannose_2$-$GlcNac_2$-$fucose_2$ corresponds to an additional mass of 1023. This oligosaccharide has been identified on membrane proteins of Sf21 insect cells (39).

An additional set of recombinant proteins was constructed based on the deduced disulfide connectivity of E3CaG2 (Fig. 1): Ca2, the Ca^{2+}-binding region of hTSP2 (Glu^{693}–Val^{947}); CaG2, the C-terminal region of hTSP2 without the EGF domain (Glu^{693}–Ile^{1172}); and tCaG2, a truncated version of CaG2 that begins at Asp^{718} and is therefore lacking the EGF domain as well as the first two cysteine residues (Cys^{707} and Cys^{715}) and the glycosylation site at Asn^{710} (Figs. 1 and 2). We also tried to express the C-terminal globular region of hTSP2, G2, by itself but detected no expressed protein in either conditioned medium or the cell pellet. MALDI-TOF MS analysis of the purified proteins revealed masses of 57716, 53373, and 30626 Da for CaG2, tCaG2, and Ca2, respectively, whereas the expected masses were 55372, 52542, and 29566, respectively. The differences can be explained by *N*-glycosylation of the two sites in CaG2 and the single sites in Ca2 and tCaG2. Peptides consistent with sequential disulfide connectivity deduced for E3CaG2 were observed in digests of Ca2, CaG2, and tCaG2 (Table I). A peak corresponding to peak 3 of the E3CaG2 digest, containing a peptide with a mass of 2213.9 corresponding to the Cys^{948}–Cys^{1170} linkage, was found in digests of E3CaG2, CaG2, and tCaG2, but not in the digest of Ca2, which lacks residues 947–1172. As with E3CaG2, a peptide with a mass of 2230 expected for unglycosylated peptide Asp^{698}–Lys^{717} was not observed. A peak corresponding to peak 6 of the E3CaG2 digest was observed in endoproteinase Asp-N digests of Ca2 and CaG2 and contained a peptide with a mass of 3252.4, presumed to be *N*-glycosylated Asp^{698}-Lys^{717}. This peptide was not found in tCaG2 lacking Asp^{698}–Lys^{717} (Table I).

To identify positively the peptide Asp^{698}–Lys^{717} containing Cys^{707} and Cys^{715}, cells infected with Ca2 virus were treated with tunicamycin, an inhibitor of *N*-glycosylation, and the unglycosylated form of Ca2, Ca2 tun, was isolated as described above. Tunicamycin treatment greatly reduced Ca2 secretion and resulted in secretion of a protein with a mass of 29596 Da as determined by MALDI-TOF MS. This is consistent with the expected unglycosylated Ca2 mass of 29,566 Da. The LC/ESI-MS patterns of the endoproteinase Asp-N digest of Ca2 tun were similar to those of E3CaG2 in regard to peaks 1, 2, 4, and 5 (Fig. 3*B*, Table I) but lacked peaks 3 and 6. A new peak (*peak 6′* in Fig. 3*B*) contained a peptide with a mass of 2229.9, which corresponds to the unglycosylated mass of Asp^{698}–Lys^{717}.

DISCUSSION

A baculovirus expression system allowed expression of large quantities of the Ca^{2+}-binding C-terminal region of hTSP2, E3CaG2, and of constructs lacking the E3 EGF module and the C-terminal 230 residues. Equilibrium dialysis, CD, and fluorescence spectroscopy indicated that E3CaG2 has the Ca^{2+} binding properties and Ca^{2+}-sensitive structure previously deduced for intact platelet TSP1 (25, 26). Estimates of binding at Ca^{2+} concentrations between 0 and 0.3 mM indicated that E3CaG2 maximally binds 9.1 ± 0.7 Ca^{2+} with an apparent K_d of 51 μM and a Hill coefficient of 1.7. TSP1 has been shown to

TABLE I

Disulfide connectivity of E3CaG2, CaG2, tCaG2, Ca2, and Ca2 tun

Mass spectrometry results from endoproteinase Asp-N digestion of E3CaG2, CaG2, tCaG2, Ca2, and Ca2 tun in 50 mM ammonium acetate, pH 6.1, at 37 °C for 4 h indicate a sequential disulfide connectivity. Peaks labeled 1–6 and 6′ correspond to HPLC peaks in Fig. 3. Peak 6′ was found only in Ca2 tun. The observed masses are those found in digests of Ca2 tun (peak 6′) or E3CaG2 (peaks 1–6). Masses in other digests varied from the masses in the table by <0.01%.

Disulfide-linked peptides	Residue nos.	Cysteines	Peak	Predicted mass	Observed mass	Mass found in
				Da	*Da*	
DGWPNLNLV**C**ATNATYH**C**IK	698–717	707 + 715	6′	2230.0	2229.9	Ca2 tun
DGWPNLNLV**C**ATNATYH**C**IK	698–717	707 + 715	6	2230.0+	3252.4	E3CaG2, Ca2, CaG2
DN**C**PHLPNSGQE	718–729	720 + 740	2	1729.6	1729.6	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DA**C**D	738–741					
DN**C**QLLFNPRQA	754–765	756 + 776	4	1807.8	1807.8	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DR**C**	774–776					
DN**C**PYVHNPAQI	777–788	779 + 799	4	1860.8	1860.8	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DA**C**SV	797–801					
DN**C**PYVYNT	813–821	815 + 835	1	1458.5	1458.5	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DH**C**	833–835					
DN**C**PLVHNP	836–844	838 + 858	1	1369.5	1369.5	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DQ**C**	856–858					
DN**C**PYISNANQA	874–885	876 + 896	2	1613.6	1613.6	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DA**C**	894–896					
DN**C**RLVFNP	910–918	912 + 932	5	1551.7	1551.7	E3CaG2, Ca2, Ca2 tun, CaG2, tCaG2
DI**C**K	930–933					
DV**C**PENNAISET	946–957	948 + 1170	3	2214.0	2213.9	E3CaG2, CaG2, tCaG2
DLKYE**C**R	1165–1171					

bind 11–12 Ca^{2+}/monomer cooperatively with an apparent K_d of 52 μM (26). Because the second EGF-like module of TSP1 is predicted to bind a single Ca^{2+}, the C-terminal region of TSP1 is predicted to contain 10–11 Ca^{2+} binding sites/monomer. The difference between the number of Ca^{2+}-binding sites in the C-terminal region of TSP1 and E3CaG2, therefore, is within experimental error.

TSP2 is the only TSP with an even number of cysteines. The disulfide connectivity of E3CaG2 can serve as a starting point for consideration of other TSPs. Endoproteinase Asp-N digestion of E3CaG2, CaG2, tCaG2, Ca2, and Ca2 tun followed by LC/ESI-MS analysis revealed a sequential disulfide connectivity of the 18 cysteines in the C-terminal region of hTSP2: Cys^{707}–Cys^{715}, Cys^{720}–Cys^{740}, Cys^{756}–Cys^{776}, Cys^{779}–Cys^{799}, Cys^{815}–Cys^{835}, Cys^{838}–Cys^{858}, Cys^{876}–Cys^{896}, Cys^{912}–Cys^{932}, and Cys^{948}-Cys^{1170} (Table I, Fig. 4). The fact that we were able to express constructs lacking the Cys^{707}–Cys^{715} or Cys^{948}–Cys^{1170} disulfide at high levels supports the deduced consecutive bonding order. Our failure to express isolated C-terminal globule sequence, G2, in the same system in which CaG2 was secreted efficiently indicates that although the Ca^{2+}-binding domain can fold independently of G2, G2 only folds correctly in the context of CaG2.

The consecutive bonding order fits the observed behaviors of TSPs in the presence and absence of Ca^{2+} as studied by rotary shadowing electron microscopy. In the presence of Ca^{2+}, TSPs have a short C-terminal stalk and large C-terminal globule, while in the absence of Ca^{2+}, the stalk is elongated at the expense of the size of the globule (13–15, 18, 38). If the disulfides were not consecutive, the structure would be intertwined and would not be expected to elongate as readily in the absence of Ca^{2+}. The free thiol in TSP1 is known to engage in a remarkably degenerate thiol-disulfide exchange such that a small fraction of each of 12 cysteines in the C-terminal region can be free rather than in a disulfide (30). Such exchange must be initiated by attack of disulfides by the cysteine in TSP1 that is at the position of Ser^{994} in TSP2 (Fig. 4). Whether the unpaired cysteine in TSP3, TSP4, TSP5/COMP, or dTSP, which is at the position of Val^{1154} in TSP2 (Fig. 4), can also initiate exchange is not known.

The C-terminal region, including the most C-terminal of the tandem EGF-like modules, is the most highly conserved region of the TSP family. Only two differences in spacing are observed in this region of TSPs. TSP3, TSP4, TSP5/COMP, and dTSP all lack the N*X*T glycosylation sequence between the ultimate EGF-like module and the first Ca^{2+}-binding sequence (Fig. 4). The second difference is insertion of PPGP in the sequence connecting the sixth and seventh 20-residue Ca^{2+}-binding sequences of TSP3 and TSP4 (Fig. 4). It will be interesting to learn whether these differences cause TSP3, TSP4, TSP5, and dTSP to have a different disulfide connectivity compared with TSP2.

The Ca^{2+}-binding region is composed of a series of 12 repeats, D*X*D*X*D*XX*GD*XX*D, that fit the consensus sequence for the Ca^{2+}-binding loop in EF-hands (Fig. 1). EF-hands bind Ca^{2+} within a helix-loop-helix (40). The C-terminal region of TSPs, however, is not predicted to have the helices on either side of the consensus sequences necessary for forming the EF-hand. Instead, a disulfide linkage closes 7 of the 12 presumptive sequences and may serve to stabilize the Ca^{2+}-binding sequences (Fig. 4). Each of the disulfide-containing sequences contains 23 residues and has a proline- and asparagine-containing sequence that has been predicted to form a β-bend (depicted as a "thumb" in Fig. 4) (21, 22). We know of no precedent for these types of Ca^{2+}-binding motifs. BM-40 (also known as SPARC or osteonectin) is an extracellu-

lar protein that has been found to contain a pair of EF-hand Ca^{2+}-binding sites in which a disulfide bond is present to stabilize the loop (41), but the disulfide bond is between the E and F helices that are not present in E3CaG2.

Mutations in TSP5/COMP at over 30 sites have been identified that lead to skeletal dysplasias. These sites are in both the aspartate-rich Ca^{2+}-binding repeats (27, 28, 42, 43) and the C-terminal nonrepeating sequence (43–46) (Fig. 4). One proposed mechanism by which these mutations cause disease is accumulation of unstable protein intracellularly (47–49). The E3CaG2 expression strategy, which allows production of large amounts of functional protein at 27 °C, and knowledge of the sequential disulfide connectivity of the region should facilitate well controlled studies of the effects of mutations inside and outside the Ca^{2+}-binding repeats on protein structure and function at physiological temperatures.

REFERENCES

1. Bornstein, P., O'Rourke, K., Wikstrom, K., Wolf, F. W., Katz, R., Li, P., and Dixit, V. M. (1991) *J. Biol. Chem.* **266,** 12821–12824
2. LaBell, T. L., McGookey Milewicz, D. J., Disteche, C. M., and Byers, P. H. (1992) *Genome* **12,** 421–429
3. Vos, H. L., Devarayalu, S., de Vries, Y., and Bornstein, P. (1992) *J. Biol. Chem.* **267,** 12192–12196
4. Lawler, J., Duquette, M., Whittaker, C. A., Adams, J. C., McHenry, K., and DeSimone, D. W. (1993) *J. Cell Biol.* **120,** 1059–1067
5. Oldberg, A., Antonsson, P., Lindblom, K., and Heinegard, D. (1992) *J. Biol. Chem.* **267,** 22346–22350
6. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Colayne, J. D., *et al.* (2000) *Science* **287,** 2185–2195
7. Adolph, K. W. (2001) *Gene* (*Amst.*) **269,** 177–184
8. Iruela-Arispe, M. L., Liska, D. J., Sage, E. H., and Bornstein, P. (1993) *Dev. Dyn.* **197,** 40–56
9. Tooney, P. A., Sakai, T., Sakai, K., Aeschlimann, D., and Mosher, D. F. (1998) *Matrix Biol.* **17,** 131–143
10. Tucker, R. P., Adams, J. C., and Lawler, J. (1995) *Dev. Dyn.* **203,** 477–490
11. Bornstein, P. (1992) *FASEB J.* **6,** 3290–3299
12. Adams, J., and Lawler, J. (1993) *Curr. Biol.* **188,** 190–193
13. Lawler, J., Chao, F. C., and Cohen, C. M. (1982) *J. Biol. Chem.* **257,** 12257–12265
14. Galvin, N. J., Dixit, V. M., O'Rourke, K. M., Santoro, S. A., Grant, G. A., and Frazier, W. A. (1985) *J. Cell Biol.* **101,** 1434–1441
15. Lawler, J., Derick, L. H., Connolly, J. E., Chen, J.-H., and Chao, F. C. (1985) *J. Biol. Chem.* **260,** 3762–3772
16. Qabar, A., Derick, L., Lawler, J., and Dixit, V. (1995) *J. Biol. Chem.* **270,** 12725–12729
17. Lawler, J., McHenry, K., Duquette, M., and Derick, L. (1995) *J. Biol. Chem.* **270,** 2809–2814
18. Chen, H., Deere, M., Hecht, J. T., and Lawler, J. (2000) *J. Biol. Chem.* **275,** 26538–26544
19. Chen, H., Sottile, J., O'Rourke, K. M., Dixit, V. M., and Mosher, D. F. (1994) *J. Biol. Chem.* **269,** 32226–32232
20. Maddox, B. K., Mokashi, A., Keene, D. R., and Bachinger, H. P. (2000) *J. Biol. Chem.* **275,** 11412–11417
21. Lawler, J., Weinstein, R., and Hynes, R. O. (1988) *J. Cell Biol.* **107,** 2351–2361
22. Sun, X., Skorstengaard, K., and Mosher, D. F. (1992) *J. Cell Biol.* **118,** 693–701
23. Hogg, P. J., Owensby, D. A., and Chesterman, C. N. (1993) *J. Biol. Chem.* **268,** 21811–21818
24. Hogg, P. J., Owensby, D. A., Mosher, D. F., Misenheimer, T. M., and Chesterman, C. N. (1993) *J. Biol. Chem.* **268,** 7139–7146
25. Lawler, J., and Simons, E. R. (1983) *J. Biol. Chem.* **258,** 12098–12101
26. Misenheimer, T. M., and Mosher, D. F. (1995) *J. Biol. Chem.* **270,** 1729–1733
27. Briggs, M. D., Hoffman, S. M. G., King, L. M., Olsen, A. S., Mohrenweiser, H., Leroy, J. G., Mortier, G. R., Rimoin, D. L., Lachman, R. S., Gaines, E. S., Cekleniak, J. A., Knowlton, R. G., and Cohn, D. H. (1995) *Nat. Genet.* **10,** 330–336
28. Hecht, J. T., Nelson, L. D., Crowder, E., Wang, Y., Elder, F. F. B., Harrison, W. R., Francomano, C. A., Prange, C. K., Lennon, G. G., Deere, M., and Lawler, J. (1995) *Nat. Genet.* **10,** 325–329
29. Thur, J., Rosenberg, K., Nitsche, D. P., Pihlajamaa, T., Ala-Kokko, L., Heinegard, D., Paulsson, M., and Maurer, P. (2001) *J. Biol. Chem.* **276,** 6083–6092
30. Speziale, M. V., and Detwiler, T. C. (1990) *J. Biol. Chem.* **265,** 17859–17867
31. Misenheimer, T. M., Huwiler, K. G., Annis, D. S., and Mosher, D. F. (2000) *J. Biol. Chem.* **275,** 40938–40945
32. Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) *Anal. Biochem.* **200,** 74–80
33. Ellman, G. L. (1959) *Arch. Biochem. Biophys.* **82,** 70–77
34. Geourjon, D., and Deleage, G. (1994) *Protein Eng.* **7,** 157–164
35. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) *Trends Biochem. Sci.* **25,** 147–150
36. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng.* **6,** 383–390
37. Merelo, J. J., Andrade, M. A., Prieto, A., and Moran, F. (1994) *Neurocomputing* **6,** 443–454
38. Dixit, V. M., Galvin, N. J., O'Rourke, K. M., and Frazier, W. A. (1986) *J. Biol. Chem.* **261,** 1962–1968
39. Kubelka, V., Altmann, F., Kornfeld, G., and Marz, L. (1994) *Arch. Biochem. Biophys.* **308,** 148–157
40. Lewit-Bentley, A., and Rety, S. (2000) *Curr. Opin. Struct. Biol.* **10,** 637–643
41. Hohenester, E., Maurer, P., Hohenadl, C., Timpl, R., Jansonius, J. N., and Engel, J. (1996) *Nat. Struct. Biol.* **3,** 67–73
42. Ikegawa, S., Ohashi, H., Nishimura, G., Kim, K. C., Sannohe, A., Kimizuka, M., Fukushima, Y., Nagai, T., and Nakamura, Y. (1998) *Hum. Genet.* **103,** 633–638
43. Deere, M., Sanford, T., Francomano, C. A., Daniels, K., and Hecht, J. T. (1999) *Am. J. Med. Genet.* **85,** 486–490
44. Briggs, M. D., Mortier, G. R., Cole, W. G., King, L. M., Golik, S. S., Bonaventure, J., Nuytinck, L., De Paepe, A., Leroy, J. G., Biesecker, L., Lipson, M., Wilcox, W. R., Lachman, R. S., Rimoin, D. L., Knowlton, R. G., and Cohn, D. H. (1998) *Am. J. Hum. Genet.* **62,** 311–319
45. Deere, M., Sanford, T., Ferguson, H. L., Daniels, K., and Hecht, J. T. (1998) *Am. J. Med. Genet.* **80,** 510–513
46. Holden, P., Meadows, R. S., Chapman, K. L., Grant, M. E., Kadler, K. E., and Briggs, M. D. (2001) *J. Biol. Chem.* **276,** 6046–6055
47. Maddox, B. K., Keene, D. R., Sakai, L. Y., Charbonneau, N. L., Morris, N. P., Ridgway, C. C., Boswell, B. A., Sussman, M. D., Horton, W. A., Bachinger, H. P., and Hecht, J. T. (1997) *J. Biol. Chem.* **272,** 30993–30997
48. Hecht, J. T., Montufar-Solis, D., Decker, G., Lawler, J., Daniels, K., and Duke, P. J. (1998) *Matrix Biol.* **17,** 625–633
49. Delot, E., Brodie, S. G., King, L. M., Wilcox, W. R., and Cohn, D. H. (1998) *J. Biol. Chem.* **273,** 26692–26697