Cell. Mol. Life Sci. 65 (2008) 672 – 686
1420-682X/08/050672-15
DOI 10.1007/s00018-007-7484-1

Cellular and Molecular Life Sciences

Structures of thrombospondins

C. B. Carlson[a], J. Lawler[b] and D. F. Mosher[a,*]

[a] Departments of Biomolecular Chemistry and Medicine, 1300 University Avenue, 4285 Medical Sciences Center, University of Wisconsin, Madison, Wisconsin 53706 (USA), Fax: +1 608 263 4969, e-mail: dfmosher@wisc.edu
[b] Division of Cancer Biology and Angiogenesis, Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, 99 Brookline Avenue, Research North 270C, Boston, Massachusetts 02215 (USA), Fax: +1 617 667 3591, e-mail: jlawler@bidmc.harvard.edu

Online First 12 January 2008

Abstract. Thrombospondins are large secreted, multimodular, calcium-binding glycoproteins that have complex roles in mediating cellular processes. Determination of high-resolution structures of thrombospondins has revealed unique and interesting protein motifs. Here, we review this progress and discuss implications for function. By combining structures of modules from thrombospondins and related extracellular proteins it is now possible to prepare an overall model of the structure of thrombospondin-1 and thrombospondin-2 and discern features of other thrombospondins. (Part of a multi-author Review)

Keywords. Thrombospondin, extracellular protein structure, cartilage oligomeric matrix protein, calcium.

Introduction

The first thrombospondin (THBS), now called THBS-1, was described by Baenziger et al. in 1971 [1]. THBS-1 was identified as a glycoprotein released from the α-granules of platelets in response to stimulation with thrombin, hence the name [1, 2]. THBSs have now been characterized as secreted glycoproteins in organisms as diverse as *Drosophila*, *Ciona* and vertebrates (five genes encoding THBS-1 through THBS-5 in humans). There are two major classifications of THBSs. Group A THBSs (THBS-1 and THBS-2) form trimers of subunits that are composed of an N-terminal domain (THBS-N), an oligomerization domain, a von Willebrand Factor type C (VWC) domain, three properdin-like repeats or thrombospondin repeats (TSRs), and a signature domain comprising three epidermal growth factor (EGF)-like repeats, a calcium-binding wire and a lectin-like C-terminal globe (Figs. 1A, B). Group B THBSs (THBSs 3–5 and arthropod THBSs) form pentamers, lack VWC modules and TSRs, and have additional EGF-like repeats [3] (Fig. 1A). These proteins are conserved, with the signature domain having the highest identity (53–82 % identity across the whole family) [4].

The large size, complex multi-modular nature, and post-translational modifications of proteins such as THBS make solving the structure of entire molecules impossible by nuclear magnetic resonance (NMR) and problematic by X-ray crystallography. Because of these obstacles, the problem has been tackled by studying recombinant fragments of THBSs comprising a limited number of modules. These modules have been identified based on a clear pattern of repeated amino acid sequence, the position of intron-exon boundaries, and homology to sequences in other proteins. In many cases, the distinct structural domains have been shown to have unique functional properties. Structures of each of the modules found in THBSs have been solved for THBS or a related protein [5–11]. These structures provide a working model of group A THBSs in high calcium (Fig. 1B). In addition, electron microscopic and low-resolution biophysical studies provide insight into the dynamic nature of THBSs and how THBS structure changes in low calcium.

* Corresponding author.

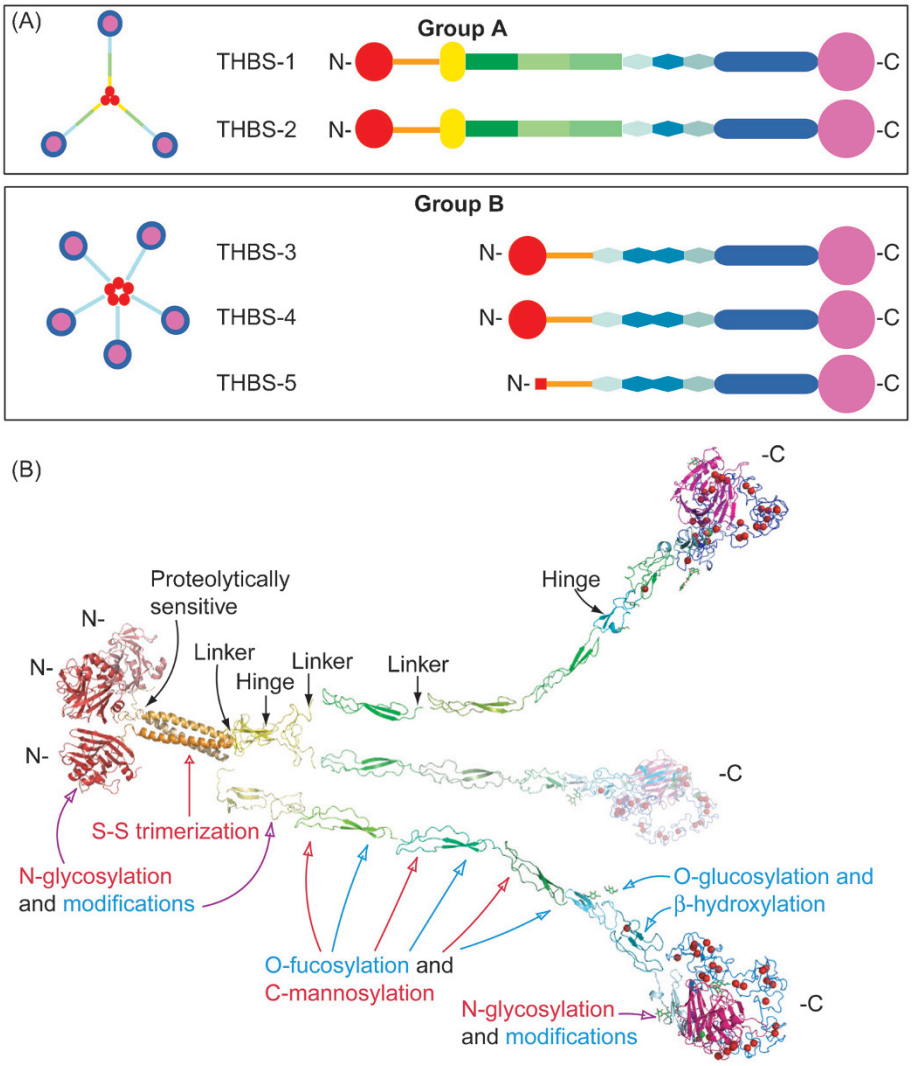

Figure 1. Full-length THBSs. (*A* and *B*) Modular structures of THBS family members. THBS-1 and -2 comprise Group A and form trimers. THBS-3, -4, and -5 comprise Group B and form pentamers. The modules are colored as follows: THBS-N (red), oligomerization coiled-coil (orange), VWC (yellow), TSR1–3 (shades of green), EGF-like repeats (shades of blue-green, with predicted or known calcium-binding repeats colored in the darkest shade), calcium-binding wire (blue), lectin-like module (purple). The N- and C-termini are labeled. (*B*) Model of the complete structure of a Group A THBS. Computer-generated model prepared using SYBYL 7.0 (Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA). This model follows the modular structure of a Group A THBS and uses PDB files 1Z78, 1AQ5, 1U5M, 1LSL and 1YO8. TSR1 is modeled from 1LSL. In the third chain, the VWC module has been modeled with the alternate N-terminal domain conformation. Elements are colored as above and: calcium ions (red spheres), signature domain sugars [green Corey-Pauling-Kultin (CPK) sticks]. Possible sites of structural flexibility (black), and sites of co-translational modifications (red) and post-translational modifications (blue) are also labeled.

N-terminal domain

Overview

THBS-N is found in all members of the human THBS family, except for THBS-5 (also called cartilage oligomeric matrix protein or COMP). In addition, a homologous domain is also found in NELL-1 and -2, kielin, TSPEAR, and several members of the collagen family [5]. THBS-N is the most divergent module of the THBS family, with low sequence identity among THBSs [5]. THBS-N has been found to have many binding partners and perhaps is best known for its heparin-binding ability [12]. Two crystal structures of THBS-N from THBS-1 (THBS-N-1) have been solved [5]: THBS-N-1 at 1.8 Å resolution (R_{work}/R_{free} = 23.6/25.9 %) and THBS-N-1 binding a synthetic heparin ligand at 1.9 Å resolution (R_{work}/R_{free} = 25.0/27.8 %) (Fig. 2A).

Classification

Homology-searching algorithms based on sequence indicated that THBS-N belongs to the laminin G superfamily [13]. Although THBS-N-1 forms a β-sandwich like other members of this superfamily, its structure is distinct from the structure of laminin G domains. It does not contain a calcium ion, as is typical for laminin G domains. In addition, the disulfide bond arrangement and the N- and C-terminal ends are distinct from the laminin G domain superfamily. A structure homology search placed THBS-N-1 in the concanavalin A-like lectins/glucanases superfamily [5]. Structures in the concanavalin A-like lectins/glucanase superfamily are characterized by a concave β-sheet and a convex β-sheet. Many members of this superfamily bind carbohydrates using a cleft-like motif found within the concave β-sheet. This cleft-like motif is not fully accessible in the THBS-N-1 module, and instead the glycosaminoglycan binding site is present on the bottom of the domain (Fig. 2A).

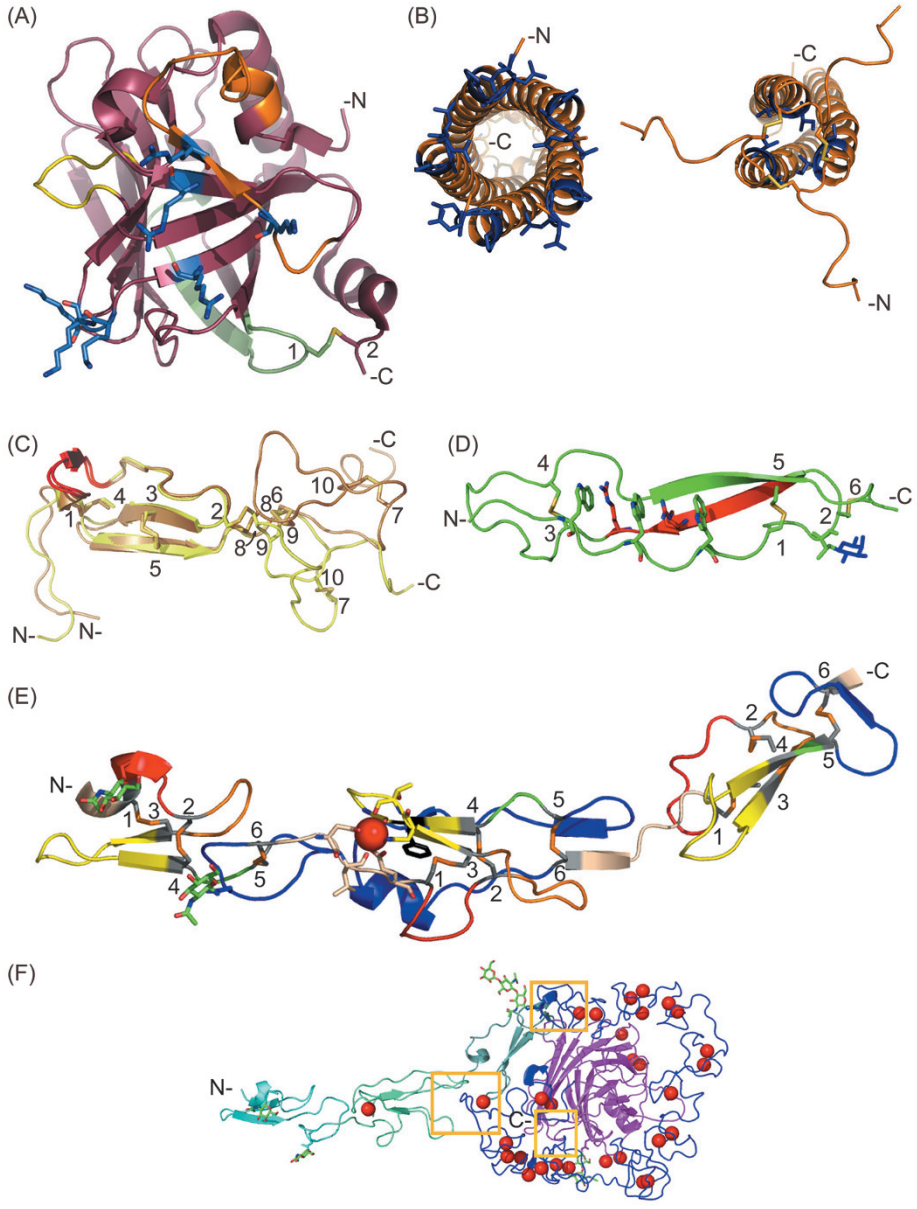

Figure 2. The modules of THBS. The N- and C-termini and order of cysteines are labeled except when indicated. (*A*) The THBS-N-1 crystal structure (PDB code: 1Z78) and binding sites are colored as follows: heparin-binding region (blue CPK sticks), fibrinogen and α4β1 binding sites (green), calreticulin binding site (orange), β13-β14 loop covering potential carbohydrate-binding cleft (yellow), disulfide bond (CPK sticks). (*B*) The oligomerization domain. Right: THBS-5 pentameric domain (PDB code: 1VDF). Left: Chicken matrilin-1 trimeric domain (PDB code: 1AQ5). These domains are viewed from the top of the coiled-coil and are colored as follows: disulfides (not-numbered CPK sticks), amino acid side chains involved in 'knobs-into-holes' packing (blue sticks). (*C*) The VWC NMR structure from procollagen A2 (PDB code: 1U5M). The two models are shown in which the N-terminal sub domain is fixed and the C-terminal sub-domain is in its two preferred conformations: 1 (yellow) and 2 (sand). Residues homologous to the THBS-1 anti-angiogenic peptide N302GVQYRN are colored in red. Disulfides are shown as CPK sticks. The cysteine residues are numbered 1–10, starting with the most N-terminal cysteine. For the N-terminal sub-domain, there is one set of cysteine numbers, but for the C-terminal sub-domain, there are two sets of cysteine numbers. (*D*) The THBS-1 TSR2 crystal structure from the TSR2-TSR3 structure (PDB code: 1LSL) is colored as follows: disulfide bonds (CPK sticks), tryptophan-arginine stacking motif (CPK sticks), and fucose (blue sticks). The cysteine residues are numbered from 1–6. The GVITRIR sequence in strand B that is mimicked by the ABT-510 peptide is colored in red. (*E*) The THBS-2 EGF1-EGF3 crystal structure from the signature domain of THBS-2 (PDB code: 1YO8) is colored as follows: disulfide bonds (grey CPK sticks), calcium-binding residues (CPK sticks), calcium ions (red sphere), sugars (green CPK sticks), conserved Tyr/Phe residue (black sticks), residues before the first cysteine (wheat), residues between the first and second cysteine (red), residues between the second and third cysteine (orange), residues between the third and fourth cysteine (yellow), residues between the fourth and fifth cysteine (green), residues between the fifth and sixth cysteine (blue). Cysteine residues are numbered 1–6 in each repeat. (*F*) The signature domain of THBS-2 (PDB code: 1YO8) is colored as follows: EGF-like repeats (blue-green), calcium-binding wire (blue), lectin-like module (purple), calcium ions (red spheres), sugar moieties (green CPK sticks). Disulfide bonds (sequential in the wire and lectin-like module) are not shown. Interaction sites are highlighted with yellow boxes.

Secondary structure

The globular β-sandwich structure of THBS-N-1 has 13 antiparallel β-strands and one unique strandlike segment (β4′), which is separate from the other β-strands. THBS-N-1 also contains six α-helices. Helix α3 is the largest and crosses over the top of the two β-sheets. There are several interactions among the α-helices and the β-sheets. There is an additional α-helix located near the C-terminal end of the domain. This helix contains Cys214, which forms a disulfide bond with Cys153 found in the β11-β12 loop. The disulfide bond brings the C terminus into close proximity to the rest of the THBS-N domain (Fig. 2A). THBS-N domains are followed by linker sequences that are protease-sensitive, and likely of variable length (Fig. 1B). We hypothesize that these linker regions allow the THBS-N domains to adopt multiple orientations to facilitate ligand binding.

Functional sites

A wide variety of receptors function to sequester THBS-1 at the cell surface. The binding sites for some of these receptors have been mapped to THBS-N-1 using synthetic peptides. Establishing that these sequences are in solvent-accessible regions of the folded domain helps to validate the function of these sites. The structure of THBS-N-1 indicates that it is organized such that a β-sandwich provides the structural framework and the binding sites for the various ligands are on the edges, top and bottom of the domain. THBS-N-1 bound to Arixtra (fondaparinus sodium), a synthetic heparin compound, is similar to the ligand-free THBS-N-1. However, there is a significant difference between the structures in the α1-β2 loop. Although the Arixtra molecule was not completely resolved in the crystal structure, sulfate groups were closely associated with Arg29, Arg42, and Arg77. Arg29 shifts dramatically between the two structures, suggesting it is a key residue in binding to heparin and modifying protein activity. The sulfate-binding residues comprise a large positively charged surface area that also includes residues Lys32, Lys80, Lys81, and Lys106 (Fig. 2A). In addition, there is a smaller patch of positive charge on the top of THBS-N-1 comprising residues Arg65, Lys68, and Arg171. Comprising the large positively charged area, residues Arg29, Arg42, and Arg77 are conserved in THBS-N-2, further supporting the importance of these residues. However, these residues are not well conserved in other THBS-N modules. THBS-N-3 and THBS-N-4 have sequence gaps that suggest that they lack the edge β-strands 2 and 3. Despite this, sequences of THBS-N modules indicate at least one positively charged surface area in all, suggesting retention of heparin-binding ability across the domain family. Since the THBS-N domain is almost always found at the N-terminal of proteins, it may serve to anchor the proteins to proteoglycans at one end and leave the other end of the protein available for interaction with cell membrane proteins.

The THBS-N-1 structure also provides insight into sequences of THBS-N-1 that have been deduced to participate in protein-protein binding sites, using peptide-based experiments (Fig. 2A). The proposed fibrinogen binding site lies within the β12 strand and nearby loop structures [5]. This binding site is present in a negatively charged patch that includes Glu127, Glu154, Glu157, Glu160, and Asp162. The A159ELDVP presumptive α4β1 integrin site also lies within this region of THBS-N-1. The identified site for calreticulin binding (residues 17–35) lies in a region that includes the α1-β2 loop, and is exposed and available for protein-protein binding. The fibrinogen, calreticulin, and α4β1 binding sites are exposed and available in the crystal structure (Fig. 2A). The presumptive α3β1 binding site, however, lies within the β14 strand and is partially obscured. Of the important residues for binding, only Arg198 is fully surface-accessible. Lastly, the presumptive α6β1 binding site is found within the β6 strand and accompanying loop, and is not surface-accessible. The residue deduced to be essential for α6β1 binding, Glu90, forms salt bridges with Arg178, Arg180, and Lys183 and is not available for binding [5].

Oligomerization domain

Overview

All THBSs contain an oligomerization domain of one of two distinct groups. Group A oligomerizes into trimers, whereas group B oligomerizes into pentamers. Although both domains are based on α-helical coiled-coils, the structures are very different (Figs. 1A and 2B).

Pentameric oligomerization domain

The oligomerization domain of THBS-5 was the first THBS domain to be crystallized. This coiled-coil structure of the rat THBS-5 oligomerization domain was solved to a resolution of 2.05 Å (Fig. 2B) [6]. This structure revealed five α-helix chains of 46 residues each, with an average length of 73 Å and an average diameter of 30 Å. The coiled-coil bundle creates an axial pore (radius = 2–6 Å), which is predominantly lined with aliphatic side chains. The five α-helices wrap together to form a left-handed superhelix, using interchain disulfide bonds (Cys71 and Cys68 of a neighboring peptide) and hydrophobic packing. Notably, there is a circular ring of hydrogen bonds that

form among the side chain nitrogen and oxygen groups of Gln54 residues from each chain. Also, the amide group of Asn41 is hydrogen bonded to the carboxylic acid group of Glu36 and the hydroxyl of Thr40 from a neighboring peptide. Additionally, salt bridges were observed between Asp46 and Arg48 from a neighboring peptide, and Glu57 and Lys62 from a neighboring peptide. This structure exhibits 'knobs-into-holes' packing, where side chains act as knobs that project into holes formed by gaps between residues on the neighboring helix (Fig. 2B) [14].

Trimeric oligomerization domain

The structure of a group A THBS oligomerization domain has not been solved. However, the NMR structure of a 46-residue stretch including the trimeric oligomerization domain of chicken cartilage matrix protein, also known as matrilin-1, provides insight into the structure of the group A THBS oligomerization domains, which form trimers [7]. The matrilin domain forms a coiled-coil structure, and each chain contains two cysteines that create stabilizing interchain disulfide bonds (Fig. 1B). These disulfide bonds are not essential to the formation of the matrilin trimer, but act to stabilize the structure once formed [7]. The same is true of THBS-1 [15]. Other interchain bonds include a hydrogen bond between the amide of Cys7 and the carbonyl of Cys5. Also, Ser9-Glu6 form a hydrogen-bonded type I β-turn, whereas Glu6-Asp3 are in an open type I β-turn. The core of the domain is hydrophobic and contains 'knobs-into-holes' packing as described for THBS-5, and also may be rigid. The surface of the domain is hydrophilic and relatively disordered.

Flexibility

In addition to the linkage between the N-terminal module and the oligomerization domain (~25 residues) described above, the linkage between the oligomerization domain (11–17 residues) and the VWC module is also a site of flexibility and protease sensitivity [16] (Fig. 1B). Thus, there likely is flexibility at both ends of the oligomerization domain.

vWF type C domain

Overview

Group A THBSs contain a single VWC module, also known as a chordin-like cysteine-rich repeat or an N-terminal fibrillar procollagen module, located C-terminal to the oligomerization domain. These modules are found in many extracellular proteins [8]. Each contains 10 cysteine residues, and several VWC modules have been found to bind TGF-β superfamily members. VWC modules do not have high sequence conservation, with the exception of CXXCXC (starting at the second cysteine) and CCXXC (starting at the eighth cysteine) motifs [8]. Although the atomic structure of this module has not been solved from THBS, insights can be gained from low resolution biophysical studies of the VWC module of THBS-1 [17] and from the structure of this module from procollagen IIA, solved by NMR [8] (Fig. 2C).

Solution studies of THBS-1 VWC

Mass spectrometry analysis of the VWC module of THBS-1 revealed that this module is glycosylated and that all of the cysteine residues are involved in disulfide bonds. No clear melting transitions were evident by differential scanning calorimetry [17]. The disulfide bonds stabilize the protein, as demonstrated by circular dichroism in the presence and absence of dithiothreitol. Lastly, the module is most likely monomeric in solution, as ultracentrifugation experiments did not reveal any oligomerization and far UV circular dichroism experiments did not show any changes in the structure of the VWC module when in a trimeric verus monomeric environment [17].

NMR structure of VWC from procollagen IIA

The NMR structures of the non-glycosylated VWC module from procollagen IIA (residues 32–70) consists of two sub-domains, with a flexible area between the domains (Fig. 2C). The disulfides form a 1–4, 2–8, 3–5, 6–9, 7–10 pairing scheme. The structure of the N-terminal sub-domain contains a double-stranded anti-parallel β-sheet formed by Cys34-Val35 and Arg40-Tyr41 and a triple-stranded anti-parallel β-sheet formed by Val46-Trp47, Arg53-Asp58 and Thr61-Asp66. The structure of the C-terminal sub-domain consists of a more flexible loop structure. These two sub-domains are held together by the 2–8 disulfide bond and a short (six residues) linker (Fig. 2C). The disulfide bond and linking sequence apparently act like a hinge between the N-terminal and C-terminal sub-domains. Thus, if two of the NMR-deduced structures of the N-terminal sub-domain are overlayed, the C-terminal sub-domain appears to have two preferred conformations (Fig. 2C). On a larger scale, the flexibility between the two sub-domains may provide one of only a few opportunities for bending and twisting of the stalk of group A THBSs (Fig. 1B).

Functional sites

The VWC module of THBS-1 has been shown to have anti-angiogenic activity [18]. This activity was mapped to N302GVQYRN using peptide-blocking studies [18]. The homologous sequences in the VWC module

from procollagen IIA are exposed in the NMR structures, and available for binding (Fig. 2C) [8].

Homology to fibronectin

Interestingly, the structure of the N-terminal subdomain of VWC demonstrates a high structural identity with fibronectin type 1 module (FN1) structure [8]. FN1 modules have not been recognized in invertebrates. The similarity between FN1 modules and VWC modules suggests that there is an evolutionary link between the module types, in that FN1 may have arisen by truncation of a VWC module. The two structures overlay with a root mean squared deviation (RMSD) of 2.0 Å over 40 residues, and residues that form the core of FN1 modules are conserved in VWC modules [8]. FN1 modules have also been found to bind TGF-β family members, which suggests a possible functional conservation [8].

Thrombospondin type 1 repeats

Overview

Three tandem TSRs are found in group A THBSs directly C-terminal to the VWC module (Fig. 1). TSRs are found in over 40 human proteins, both as tandem repeats and alone [19, 20]. Six TSRs comprise the entire structure of properdin, and thus an alternative name for this repeat is 'properdin repeat'. TSRs are not presentin group B THBSs. Although the name TSR might therefore be construed as misleading, it is used widely and likely will not be replaced. TSRs are conserved over evolution, and 14 *Drosophila* and 27 *Caenorhabditis elegans* proteins contain TSRs. The TSRs are characteristically extracellular (either in extracellular proteins or extracellular portions of transmembrane proteins), about 60 amino acids long, and contain around 12 conserved residues comprising six cysteines (for which there are two pairing strategies), two conserved arginine residues, two conserved glycine residues, and two to three tryptophan residues separated by two to four amino acids [19, 20] (Fig. 2D). Functions associated with the TSRs found in THBS-1 include cell attachment, angiogenesis inhibition, protein-protein interactions, and protein-glycosaminoglycan interactions [12]. The module is subject to two unusual carbohydrate modifications, introduction of a mannose onto a conserved tryptophan, and of a fucose-glucose onto a conserved serine or threonine [21] (Fig. 2D).

THBS-1 TSR2-TSR3 crystal structure

The structure of TSR2-TSR3 from THBS-1 has been solved to 1.9 Å resolution ($R_{work}/R_{free} = 23.8/28.2\,\%$) [19] (Fig. 2D). The crystals arose from a construct that contained all three TSRs and lost TSR1 due to cleavage. Two fucose moieties were bound to the complex at Thr432 and Thr489. Predicted C-mannosylation was not observed, as the S2 cell expression system is incapable of this modification [19, 21]. However, it was observed that the Cδ1 atoms of the predicted tryptophan modification sites are exposed in the structure. The structure revealed that each TSR is approximately $15 \times 20 \times 55$ Å and comprises a three-stranded anti-parallel design, with Pro472 at the intersection between the repeats. A twist between the repeats creates a 180° turn such that TSR3 faces directly opposite to TSR2. Stabilized by the disulfide bond Cys433-Cys471 and a hydrophobic interface formed by Ala470, Cys471, Ile473, Phe508, and Gly509, the linkage between TSR2 and TSR3 is most likely rigid and relatively inflexible. It is predicted that TSR1 has similar structure to TSR2 and TSR3. However, the linking sequence between TSR1 and TSR2 is four residues longer than between TSR2 and TSR3, which may affect the flexibility of the connection between TSR1 and TSR2 (Fig. 1B) and likely contributed to proteolytic loss of TSR1 during crystallization [19].

The Arg-Trp stacking motif

Of the three strands in the structure, only the second two strands of the TSR have a typical β-strand architecture (Fig. 2D). The first strand, strand A, has a 'rippled' appearance and contains the conserved motif WXXWXXW. The tryptophan side chains create a striking layering effect, with conserved arginine guanidinium groups from the neighboring strand (strand B) filling in the spaces between the layers. This packing likely acts to stabilize the structure via cation-π interactions. In addition, the arginine residues have extensive interactions with residues in the third strand, named strand C. Since the tryptophan and arginine layers are capped on the ends by disulfides, the TSR repeats have a CWR layer creating a Cys1-Trp1-Arg1-Trp2-Arg2-Trp3-Arg3-Cys2 pattern from the top to bottom of each strand A-strand B pairing. At the bottom of the repeat, a cysteine at the C-terminus of strand A creates a disulfide bond with a cysteine found at the C-terminus of strand C (Fig. 2D). This disulfide closes off the bottom of the repeat. At the top of the repeat, located between strands B and C, there are two jar handle structures. The first, comprising residues Asn445-Ser448 (of TSR2), form direct hydrogen bonds with the N-terminus strand A, using the amide group of Ser446 and the carbonyl group of Pro447. This acts to stabilize the top of the domain. In TSR3, the residue at the position of Pro447 is also a proline; however, this is not the case in TSR1. This may cause a different

conformation in TSR1 at this location. The second jar handle comprises residues Cys456-Ala460 of TSR2, including a conserved glycine at position 3 (Gly458), and creates a hydrogen bond to the Trp layer of strand A (Trp1), using the carbonyl of Glu457. This allows for the side chain of Trp1 to sit within the jar handle, while Trp2 and Trp3 side chains are exposed. This amino acid arrangement is conserved in TSR1 and TSR3. The buried and exposed tryptophans are suggested to create a 'recognition' face of the molecule, where the Trp2 and Trp3 exposed residues and exposed arginines in the R layers are available for binding interactions.

Cysteine pairing schemes

Using database analyses, Tan et al. 2002 and Huwiler et al. 2002 surveyed the disulfide arrangement of TSRs found in diverse proteins, suggesting that there are two major groups of TSRs [9, 20]. Group 1 TSRs, including those found in THBS-1 and THBS-2, have a disulfide pattern where the top disulfide (using the C layer) is created by a cysteine in strand B and a cysteine in strand C. However, group 2 TSRs, including those found in F-spondin, have a disulfide pattern where the top disulfide is created by a cysteine in strand A and a cysteine in strand C. This indicates that the N-terminus of group 2 TSRs is stabilized via the disulfide bond, rather than via the first jar handle, as seen in the THBS TSRs. The difference in structure between group 1 and group 2 TSRs may help to explain functional differences observed between TSRs from different proteins.

Carbohydrate modifications

As mentioned above, two unusual types of glycosylation have been recognized in the TSRs of THBSs. C-mannosylation is co-translational and is mediated by dolichol-P-mannose. It occurs within the TSR WXXW sequence, and specifically Trp368, Trp420, Trp423, and Trp480 of THBS-1 each form a C-mannose [21]. Other proteins with a WXXW sequence are also subject to C-mannosylation [21]. O-fucosylation occurs within the CSX(S/T)C sequence of TSRs, where the underlined Ser or Thr is fucosylated. This is a post-translational modification that requires that the module be folded and is mediated by O-fucosyltransferase 2 [22]. Studies of ADAMTS13 and ADAMTS-like-1/Punctin-1 have demonstrated that O-fucosylation of TSRs is critical for protein maturation and secretion [22, 23]. Interestingly, this modification is so far unique for TSRs and is, for instance, distinct from O-fucosylation of EGF-like repeats, which is catalyzed by O-fucosyltransferase 1 [22, 23].

Functional sites

The TSRs mediate interactions of THBS-1 with CD36 and integrins [24–26]. CD36 has been shown to mediate the anti-angiogenic activity of the TSRs [27]. The TSRs of THBS-1 and -2, as well as several other proteins, inhibit angiogenesis. Furthermore, an anti-angiogenic compound, designated ABT-510, which is based on the THSB-1 sequence, is currently in clinical trials as a treatment for cancer [27]. ABT-510 is a substituted derivative of the peptide GVITRIR, which is found in the second β strand of the second TSR of THBS-1 (Fig. 2D). These data are consistent with the hypothesis that the positively charged recognition face of the TSRs is involved in binding to CD36.

The WSHWSPW and RFK sequences in THBS-1 TSR2 mediate binding and activation of TGF-β, respectively. Since the appropriate spacing of these two sequences is specific to THBS-1, it appears to be unique in its ability to activate TGF-β. It has been proposed that in latent TGF-β, the L54SKL sequence in the N-terminal region of the latency associated peptide (LAP) interacts with the sequence R94KPK in the active portion of TGF-β. As a first step in activation, the WSHWSPW sequence of THBS-1 binds to the VLAL sequence in the active portion of TGF-β. This interaction positions the THBS-1 RFK sequence so that it can compete with the TGF-β RKFK sequence for binding to the LAP LSKL sequence. The binding of the THBS-1 RFK sequence to the LAP LSKL sequence induces a conformational change that renders the TGF-β able to bind to its receptor and initiate signaling [28, 29].

EGF-like repeats

Overview

There are at least three tandem EGF-like repeats in all THBSs. These repeats are the most N-terminal portion of the signature domain and have the potential to propagate changes in the signature domain to the rest of the protein. The EGF-like repeats from THBS have been solved within the context of the signature domain (2.6 Å resolution with $R_{work}/R_{free} = 21.8/28.4$) (Fig. 2E,F) [11]. THBSs have a mixture of calcium-binding and non-calcium-binding EGF-like repeats. All have a presumptive 1–3, 2–4, 5–6 disulfide pattern. The consensus sequence for calcium-binding EGF-like repeats is ZXZZ-C-$X_{variable}$-C-$X_{variable}$-C-X-D/N-X_4-Y/F-X-C, where Z is an aspartate, glutamate, asparagine, or glutamine [30]. THBS-2 has been shown to have one calcium-binding EGF [11]. There is one EGF-like repeat with the calcium-binding consensus sequence in THBS-1, two in vertebrate

group B THBSs, and two in *Drosophilia* THBS. Shrimp (*Marsupenaeus japonicus*) THBS has at least six EGF-like repeats, with five repeats having a calcium-binding consensus sequence, except for an arginine at the position of the conserved Y/F [31]. It will be interesting to learn whether or not these repeats bind calcium.

Most EGF-like repeats have one residue between the fourth and fifth cysteine [30, 32]. Interestingly, several THBS EGF-like repeats have the canonical six cysteine pattern, but with two residues between the fourth and fifth cysteines. NMR studies of an EGF-like repeat from Protein S demonstrated flexibility about the single residue that separates the fourth and fifth cysteine such that the N-terminal part of the repeat containing the 1–3 and 2–4 disulfides acted as a subdomain, as did the C-terminal part of the repeat containing the 5–6 disulfide [33]. Having two residues between the fourth cysteine and the fifth cysteine, as seen in THBSs, may increase the possibility for such structural flexibility. Examination of the structure of the loops found between disulfide bonds in EGF-like repeats from THBS-2 reveals that in EGF1, the extra length between the fourth and the fifth cysteines may allow the N-terminal portion of the repeat to bend, while this seems unlikely in EGF2, because of the elongated 5–6 loop that wraps back to embrace the N-terminal portion of the repeat (Figs. 1B and 2E). THBS-2 EGF3 has only one residue between the fourth and fifth cysteines and has multiple interactions with the wire and the lectin-like module, and therefore likely has limited movement (Figs. 2E, F).

EGF1 and EGF2

EGF1 and EGF2 form an elongated stalk in the crystal structure of the signature domain of THBS-2 [11]. Each has the expected 1–3, 2–4, 5–6 disulfide connectivity of modules in the cEGF subtype 1 [32], despite the unusual two-cysteine separation of the fourth and fifth cysteines. EGF1 contains a saccharide, modeled as glucose, via Ser555 (Fig. 2E), which is in the sequence CxSxPC that is modified by a glucose-xylose dissacharide in bovine THBS-1 and other EGF-like modules containing this recognition sequence [34]. The proximal portion of a carbohydrate modification at Asn584 is also visible (Fig. 2E). EGF2 binds a single calcium ion and is classified as a class A calcium-binding EGF-like module [30]. The calcium ion is bound at the interface between EGF1 and EGF2 (Fig. 2E) and likely stabilizes the interactions between the modules, as has been noted in other examples of tandem calcium-binding EGF modules [30]. The calcium coordination involves the side chains of Asp590, Asn612, and Glu593 and the main chain carbonyl groups of Leu591, Thr613, and Gly616. In addition, EGF2 contains the unusually long loop of 23 residues between the fifth and sixth cysteines. Arg625, at the bottom of this loop, juts into and appears to mediate the interactions with more C-terminal aspects of the signature domain (Fig. 2F).

EGF3

EGF3, also of the cEGF subtype 1 grouping, has a length of 30.5 Å and 1–3, 2–4, 5–6 disulfide connectivity. The module is aligned approximately 60° from the axis through EGF1 and EGF2 and has extensive interactions with EGF2, the N- and C-terminal portions of the calcium-binding wire, and the lectin-like module (Fig. 2F).

Signature domain

Overview

At the C-terminus, all THBSs contain a 'signature domain', which includes the tandem EGF-like repeats described above, 13 calcium binding repeats comprising the calcium-binding wire, and a C-terminal lectin-like module (Fig. 2F). In addition to being found in all THBS family members, this portion of the proteins contains the highest sequence conservation (53–82% identity across the whole family) [4]. The largest difference in the signature domains among THBSs is in the number of EGF-like repeats. There are three in THBS-1 and THBS-2, four in THBS-3, THBS-4 and THBS-5, and more in various non-human THBSs [4]. Other differences include two sites of short insertions within the calcium-binding wire that are conspicuous because spacing in the signature domain is otherwise maintained rigorously (Fig. 3). The first insert, insert A, is located in repeat 1C of the wire. Group B THBSs lack the 3-residue NAT glycosylation recognition sequence found in insert A. In the structure of the THBS-2 signature domain, the insertion does not perturb the folding of repeat 1C, which overlays the other C-type repeats nearly exactly [11]. The second insert, designated insert B, is found only in THBS-3 and THBS-4, is located in repeat 11C of wire, and consists of four amino acids. An insertion this short may not be able to be accommodated without considerable disturbance of the structure of repeat 11C. Lastly, the C-terminal tail of the lectin-like globe is longer in the group B THBSs than in the group A THBSs [4, 11]. In the THBS-2 structure, the C-terminal tail is on the exterior of the protein and seems to be moving away into space. However, a longer C-terminal tail might interact with another part of the protein and could lead to differences in the structures of group B versus group A THBSs. One may speculate that such variations could impart

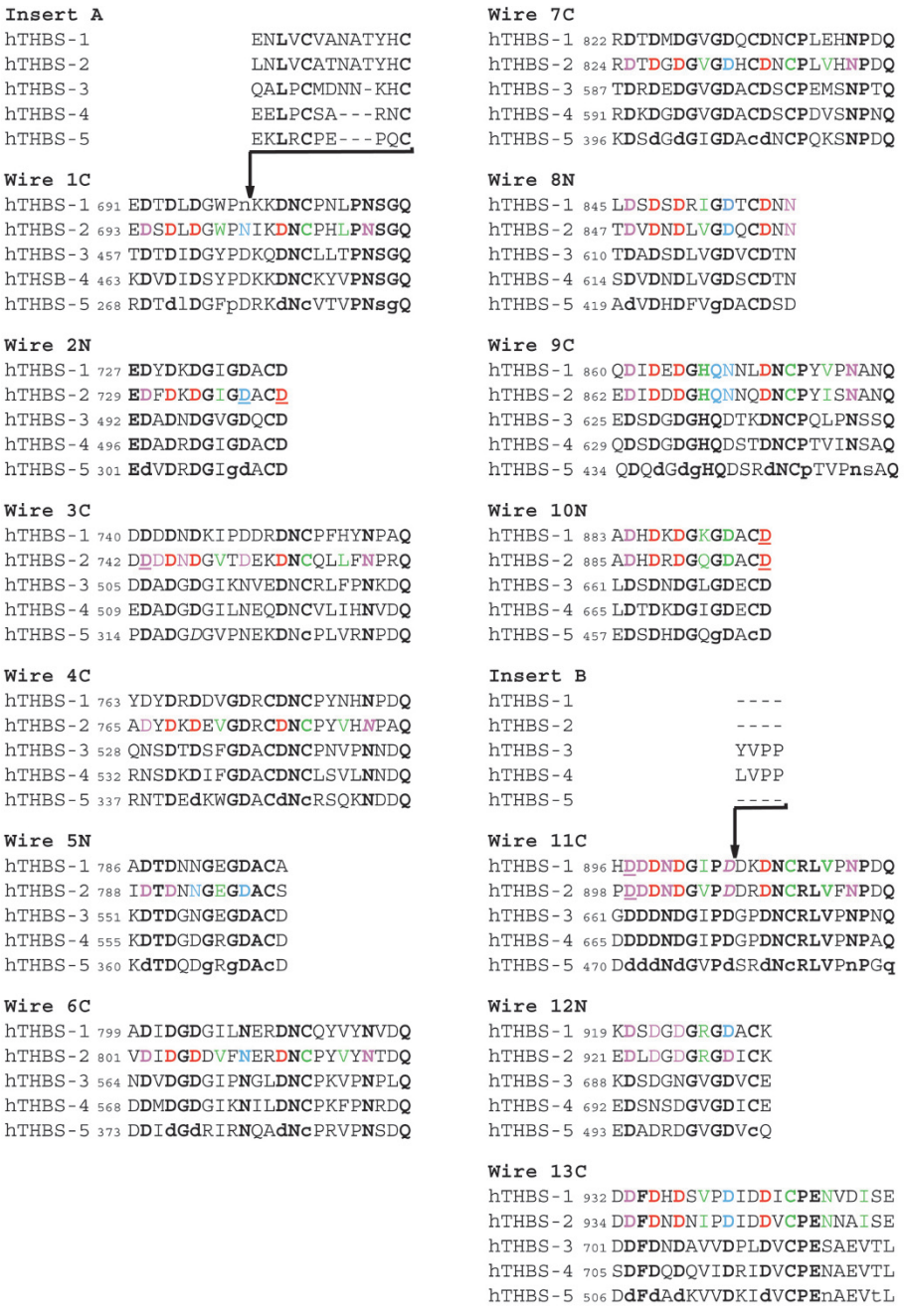

Figure 3. Sequence alignment of the 13 repeats of the wire module. Alignment of 13 calcium-binding repeats in the wire. The calcium coordination scheme present in the THBS-2 and THBS-1 structures (PDB codes: 1YO8 and 1UX6), is colored-coded as follows: side chain coordination of two calcium ions (red), side chain coordination of one calcium ion (purple), main chain carbonyl coordination of one calcium ion (green, or underlined for residues that also coordinate in other ways), and water-mediated coordination (blue, or italicized for residues that also coordinate in other ways). Residues involved in mutations in THBS-5 and the polymorphism in THBS-1 are shown in lowercase letters. Amino acids conserved among all human THBSs are shown in bold.

specialized functions upon the different THBS family members.

The signature domain and disease

Additional interest in this portion of the protein arises from the large number of disease-associated genetic alterations occurring within the signature domain. Over 100 signature-domain mutations in THBS-5 (also known as cartilage oligomeric matrix protein or COMP) have been associated with the bone and joint diseases pseudoachondroplasia (PSACH) and multiple epiphyseal dysplasia (MED) [35, 36]. The endoplasmic reticula of chondrocytes from patients with PSACH are dilated and contain precipitated proteins consisting of THBS-5, type IX collagen, BiP, and calreticulin [37, 38]. Phenocopies of the MED syndrome occur in patients with mutations of the α1, α2, or α3 chains of type IX collagen or matrilin-3 [39]. The likely mechanism by which these mutations cause disease is through the accumulation of unstable or misfolded protein in the endoplasmic reticulum of the chondrocytes within the growth plate [39]. Over time, the chondrocytes become engorged with inclusions that are derived from the endoplasmic reticulum and

die. In PSACH patients, the death of growth plate chondrocytes leads to a premature cessation of growth and significantly shortened stature.

In addition to the THBS-related diseases PSACH and MED, nonsynonymous single-nucleotide polymorphisms in the coding region of the THBS-1 and -4 genes have been linked to a familial risk for premature coronary artery disease [40]. The THBS-4 polymorphism (Ala387Pro) is located in the third EGF-like repeat of THBS-4, which is predicted to bind calcium. The Asn700Ser THBS-1 polymorphism is located in repeat 1C of the calcium-binding wire (Figs. 2F and 3) and has been associated with subtle changes in the kinetics of binding of calcium to repeat 1C and adjacent repeats [41, 42].

Overview of structures of the signature domain

The structures of a portion of the THBS-1 signature domain (1.9 Å resolution and $R_{crystal}/R_{free} = 19.9/22.7\,\%$) and the structure of the entire THBS-2 signature domain (2.6 Å resolution and $R_{work}/R_{free} = 21.8/28.4\,\%$) have been solved [10, 11]. These two structures show high conservation, as could be hypothesized from the sequence identity between the two proteins. The THBS-1 structure comprised repeats 7C–13C of the calcium-binding wire and the entire C-terminal lectin-like module, whereas the THBS-2 included the three EGF-like repeats (EGF1–3), calcium-binding wire, and C-terminal lectin-like module that comprise the THBS-2 signature domain. The following description of the structure of the signature domains of THBSs is based on the THBS-2 structure (Fig. 2F) and is extended to discuss implications of sequence variations found among the THBSs. The structure of the human THBS-2 signature domain contains stalk, wire, and globe elements. The stalk, comprising EGF1 and EGF2, extends 60 Å from the base of the molecule. The wire, which has a contour length of ~170 Å, is formed by 13 repetitive C- or N-type aspartate-rich sequences coordinating 26 calcium ions. EGF3 serves as a clasp to hold the stalk and wire elements together. The globe consists of the calcium-binding lectin-like module suspended beneath the wire and EGF3. This arrangement results in a 14×40 Å opening between the wire and the globe. The overall structure contains a large number of intramolecular interactions that suggest that the signature domain of THBSs constitutes a single calcium-sensitive structural and functional unit (Fig. 2F).

Stalk

Stalk modules EGF1 and EGF2 extend away from the rest of the domain, presumably connecting with the C-terminus of the elongated TSRs. Through the rigid EGF1-EGF2 connection and the unusually long loop of EGF2, which interacts with the hairpin turn comprising repeats 12N and 13C of the C-terminal part of the wire (Fig. 2F), EGF1 and EGF2 modulate the structure and stability of the signature domain, i.e., differential scanning calorimetric analyses have demonstrated increased protein stability upon introduction of EGF1 and EGF2 to the more C-terminal portions of the signature domain [43].

Clasp

EGF3 acts as a clasp to bring together the stalk, wire, and globe elements. EGF3 has extensive interactions with EGF2, the N- and C-terminal portions of the calcium-binding wire, and the lectin-like module (Fig. 2F).

The wire

The structure of the calcium-binding wire can best be described as a protein chain draping itself around 26 calcium ions to form a tortuous polypeptide tube with metal on the inside. The wire comprises 13 aspartate-rich repeats that form individual globular folds connected in a disulfide-stabilized arrangement. These 13 DXDXD/N-containing repeats fall into the N-types and C-types proposed by Kvansakul et al.[10] (Fig. 3). N-type repeats are 13 or 15 residues in length, and their tertiary structures superimpose with an RMSD of ~0.35 Å for all C_α atoms. C-type repeats are 23 residues long and superimpose with an RMSD ~0.36 Å.

The structures of the THBS-1 and THBS-2 signature domains [10, 11] are rich in information about calcium coordination (Fig. 3). On average, each repeat binds two calcium ions. C-type repeats bind these calcium ions using aspartate residues at positions 2, 4, 6, and 13, and the main chain carbonyl at position 8 coordinates the first calcium ion. At position 10, there is variable coordination of the first calcium ion by aspartate or asparagine, either directly, via a water molecule, or not at all. The second calcium ion is coordinated by aspartates at positions 4, 6, and 13, main chain carbonyls of the cysteine at position 15 and the residues at positions 18, and the asparagine or isoleucine at position 20 or 21 (Fig. 3). N-type repeats use the same residues to bind the first calcium ion, but employ different strategies to bind the second. Repeat 8N, which binds two calcium ions, is longer than the other N-type repeats and uses an asparagine side chain in position 15 for coordination. Also, a glutamine in postion 8 of repeat 9C contributes to the coordination by binding the calcium ion via a water molecule. Repeats 5N and 12N, which lack an aspartate at position 13 and are shorter than repeat 8N, bind only one calcium ion. The remaining N-type repeats, 2N

and 10N, are involved in the coordination of three calcium ions. This coordination is achieved by use of residues, most strikingly the DDDND sequence, from repeats 3C or 11C. The aspartate at position 2 and the asparagine at position 5 of the C-type repeat directly contribute to the coordination of the second calcium ion. The third calcium ion is coordinated by main chain contributions of aspartates in positions 10 and 13 of the N-type repeat and side chain contributions of aspartates at positions 3 and 10 of the C-type repeat. Most of this calcium-binding scheme appear to be conserved in the group B THBSs. However, there likely are some differences. Repeat 3C of group B THBSs lacks an aspartate at position 3 and an aspargine at position 5 of the DDDND sequence. This makes it unlikely that a third calcium ion can be bound between repeats 2N and 3C. A second difference is that unlike THBS-1 and THBS-2, there is an aspartate at position 13 of repeat 5N of THBS-3, THBS-4, and THBS-5. Although this residue may allow repeat 5N to bind a second calcium ion, the repeat is still only 13 residues long and it lacks the ability to use a residue at position 15, as in repeat 8N. Also, repeat 6C does not have an asparagine at position 5 to help contribute to coordination of the second ion. Thus, only one calcium ion likely binds to repeat 5N of group B THBSs, using different coordination geometry than the binding of the single calcium ion by repeat 5N of group A THBSs. The critical residues for coordination of two calcium ions are conserved in repeat 11C found in all THBSs. However, the impact of the four-residue insert B of THBS-3 or THBS-4 wire repeat 11C is unknown. Insert A in repeat 1C, although shorter in group B THBSs, can likely be accommodated (Fig. 3).

Insert A forms a structured loop that is stabilized by a disulfide bond and interacts extensively with EGF3. An N-glycosylation recognition sequence is present at the apex of the inserted loop, and the proximal three saccharides are resolved in the structure. Repeats 10C–13C form a hairpin turn, which tacks this region of the wire to the C-terminal portions of EGF2 and EGF3 (Fig. 2F). This hairpin turn appears to be important for the folding and stability of the wire, as truncations of the THBS-1 [10] and THBS-2 [D. S. Annis and D. F. Mosher, unpublished] signature domain not including the hairpin turn revealed poor protein solubility.

The globe

The lectin-like module is composed of two β-sheets, each comprising seven β-strands, in a jelly-roll packing arrangement that is nearly identical in the THBS-1 and THBS-2 constructs [10, 11]. Of the two β-sheets, one has a convex curve on the surface of the protein and the other has a more severe bend toward the interior of the protein. A short α-helix lies within the cleft formed by the second β-sheet (Fig. 2F). Although the homology was not recognized by sequence analysis, structure homology studies placed this module in the L-type lectin domain family [10]. Other examples of proteins in this family include the lectin-like module of p58/ERGIC-53, which shuttles blood coagulation factors V and VIII from the ER to the Golgi, and a plant lectin from *Griffonia simplicifolia* [10]. THBS-2 contains a serine at the homologous site of an unpaired Cys992 found in THBS-1. The residue was changed to Ser992 in the truncated THBS-1 protein used for the crystal structure [10]. In both THBS-1 and THBS-2 structures, this serine is buried [10, 11]. Cys1167 of THBs-1 or Cys1169 of THBS-2, near the C-terminus of the module, forms a disulfide bond with Cys946 of THBS-1 or Cys948 of THBS-2 in wire repeat 13C. Thus, the C-teminus of the lectin-like module lies close to its N-terminus.

Despite the overall similarities between the lectin-like module structures of THBS-1 and THBS-2, there are differences. There are four bound calcium ions in the THBS-1 lectin-like module, but only three in THBS-2. The residues that coordinate a fourth calcium ion in the THBS-1 structure are present in THBS-2, but are part of a meandering loop in contrast to the tight circular loop observed in THBS-1. The addition of a single calcium ion in the THBS-1 construct could be due to the higher calcium concentration used for crystallization (5 mM $CaCl_2$ used in THBS-1 [10] versus 2 mM $CaCl_2$ used in THBS-2 [11]). Calcium-binding studies of THBS-1 and THBS-2 have not shown a difference in numbers of calcium bound between these proteins, but a difference of one calcium ion would fall within the experimental error [41, 44].

Interactions within the structure

Extensive interactions occur among the EGF-like modules, the ends of the wire, and the lectin-like module. EGF3 forms a clasp in the signature domain through its interactions with the wire (repeats 1C, 8N, 9C, 12N, and 13C), EGF2, and the lectin-like module (Fig. 2F). Combined, these interactions create an intermodular complex where the 170 Å-long wire wraps around the 30–40 Å-wide lectin-like module, interacting at its ends with EGF3 and the lectin-like module. EGF2 and the lectin-like module make additional direct contacts with the wire. Inasmuch as attempts to produce soluble lectin-like module by itself from THBS-1 [10, 45] or -2 [46] have failed, it is likely that the more N-terminal modules must fold first to facilitate folding of the lectin-like module. Thus, one can envision a series of events in the

endoplasmic reticulum in which a third EGF-like module and wire support folding of the remainder of the THBS signature domain. This model helps explain the structural basis of disease-linked THBS-5 mutations, since these appear to affect protein folding *in vivo*.

Mapping disease

Of the disease-linked THBS-5 mutations, 80 are missense mutations that change residues at 62 different positions in the signature domain. Strikingly, all but one of these maps to either the wire (53 sites) or to positions on the globe that form interfaces with the wire (9 sites). Of the 53 involved residues in the wire, 50 are invariant between THBS-2 and THBS-5, and all but 8 coordinate calcium, form disulfides, or interact with the lectin-like module. Twenty-two additional disease-linked mutations result in insertions or deletions of residues in the wire. Several examples of THBS-5 with mutations in the wire have been found to bind less calcium and to differ from wild-type THBS-5 in their stability and binding properties [47–51]. Thus, the sites of disease-causing mutations in THBS-5 target residues that appear to be critical for either the stability of the calcium-sensitive structure of the wire or the ability of the wire to stabilize the structure of the lectin-like module.

While the THBS-4 mutation is in a domain not present in THBS-2, the Asn700Ser THBS-1 polymorphism maps to position 10 of repeat 1C (Fig. 3). The THBS-2 signature domain structure shows that the analogous residue in THBS-2, Asn702, coordinates a calcium ion (Fig. 3). The Ser700 polymorphism is associated with subtle changes in the kinetics of binding of calcium to repeat 1C and adjacent repeats [41, 42]. In THBS-5, there are three known disease-associated mutations in the wire at the aspartates at position 10 (Asp310Val in repeat 2N, and Asp479His and Asp479Tyr in repeat 11C) (Fig. 3). Similar to Asn702, the THBS-2 residues at these positions, Asp738 and Asp907, coordinate a calcium ion through a water molecule. However, Asp738 and Asp907 also coordinate an additional calcium ion. Asp738 provides main chain coordination for the calcium ion found between repeats 2N and 3C, while Asp907 provides a side chain coordination for the calcium ion found between repeats 10N and 11C. Thus, the position 10 changes in THBS-5 have the potential to perturb calcium binding and protein structure more profoundly than the THBS-1 polymorphism.

Functional sites

The calcium-replete THBS-2 signature domain structure provides insight into the several proposed protein-protein binding sites within this region. The RGD sequence implicated in αvβ3 and αIIbβ3 integrin binding [52, 53] is located in wire repeat 12N, but is mostly buried and not accessible [10, 11]. Cathepsin G has been shown to bind to two sequences in the wire of THBS-1 [54]. The first sequence, N753CPFHYNP, is located in wire repeat 3C and the second sequence, N812CQYVYNV, is located in wire repeat 6C. Neutrophil elastase has also been shown to bind to the second sequence [54]. Although the first sequence is fully accessible, the second sequence is only partially accessible. R1034FYVVM, the proposed site of CD47-binding in THBS-1, is located in the lectin-like module [55] and is buried in the calcium-replete structure. The monoclonal antibody A6.1, for which the epitope has been mapped to wire repeat 1C (residues 692–717) of THBS-1 [56], blocks THBS-1 binding to collagen V [57]. This region is surface accessible in the calcium-replete structure. Residue Cys992 in THBS-1, which is unpaired and free, mediates thiol-disulfide exchange of THBS-1 and von Willebrand factor [58]. This residue is located in the lectin-like module and is partially accessible. Most of these binding interactions are enhanced in the absence of calcium, or the presence of reducing agent, implying that a different conformation of THBS than is found in the calcium-replete structure is needed for binding. Major challenges for the future are developing better understandings of calcium-depleted structures of THBSs and the extent and pairings of disulfide rearrangement.

Effects of calcium

Effects of calcium on THBS structure have been demonstrated by rotary shadowing electron microscopy and far UV circular dichroism. In the rotary shadowing images of each of the human THBSs, THBSs are shown to exist as trimers or pentamers with long stalks extending from the oligomerization site and ending in C-terminal globes. When calcium was removed by EDTA treatment, the stalks of the protein elongated and the C-terminal globes shrunk [59]. These results suggested that calcium keeps the signature domains of THBSs in a more compact conformation that unravels when calcium is lost. In addition, the circular dichoic spectra changed with the removal or addition of calcium, indicating again that calcium causes a conformational change [44, 48, 60]. Interestingly, the circular dichroic spectrum of a THBS-2 construct comprising solely its wire repeats displays a calcium-dependent ellipticity minimum at 203 nm that is characteristic of the circular dichroic spectra of all THBSs studied to date [44]. The wire-containing protein is denatured only slightly more

readily than the full signature domain [44]. Thus, the wire forms a stable, calcium-dependent structure that is independent of the rest of the signature domain. In the crystal structure of the signature domain of THBS-2 [11], the wire makes extensive interactions with the stalk and globe elements as described above (Fig. 2F). Thus, the wire is positioned to transmit allosteric changes induced by calcium ions to other parts of THBSs.

Shape and dimensions of the modules of THBS gained from the NMR and crystallographic structures are compatible with and allow for a more exact interpretation of images of intact THBS-2 obtained by rotary shadowing electron microscopy [61, 62]. Models based on conformation-sensitive antibody-binding experiments suggest two possible results of calcium loss. These models are based on antibody evidence that wire repeat 12N remains bound to EGF2 when calcium is removed [62]. In the minimal model, wire repeat 13C loses its structure, and all interactions between the wire and the lectin-like module are lost. In the second case, repeats 11C and 13C lose their structure, but the interactions between repeats 8N and 9C of the wire and the lectin-like module are maintained [62]. In both models, interactions of the lectin-like module and EGF3 are lost, allowing the lectin-like module to fall away and elongating the stalk by 60 Å [62].

Model of a full-length group A THBS

The current NMR and crystal structures can be compiled to create a model of a full-length type A THBS molecule (Fig. 1B). This model assumes propagation of geometry of the trimeric domain throughout the length of the protein, C-terminal to N-terminal sequential interactions between adjacent modules through EGF-1, and no interactions between modules, except extensive interactions within the signature domain. In addition, it assumes a straight connection and 180° turn between TSR1 and TSR2. This model allows consideration of a THBS in its entirety. For example, it visualizes locations where flexibility may be present in the stalk due to linkers between modules or the possible subdomain structure of von Willebrand factor C and EGF-1 (Fig. 1B). This flexibility may enable THBSs to bind and organize multiple ligands. In one chain of the trimeric model, the VWC module has been modeled with the alternate N-terminal domain conformation, although no major changes were observed as a result (Fig. 1B).

As described above, there are many co- and post-translational modifications in the various THBS modules. The full-length THBS model is useful in thinking about these modifications (Fig. 1B). The modifications dot the surface of THBSs and likely act as beacons to guide chaperones to particularly hard-to-fold structures in order to ensure the proper maturation of this complicated molecule. For example, the O-fucosylation sites in the TSRs lie at the interface between modules, where it may be critical to have the quality control that has been attributed to the glycosylation machinery [22, 23].

Wrap-up

There are many questions left to answer. Despite regions of high sequence identity among the THBSs, these proteins have a wide diversity of functions. Comparative investigation into the structures of THBS family members may illuminate how THBS structure is coupled to function. It is unclear how the lack of an N-terminal module, the addition of an EGF-like repeat, the lack of an N-glycosylation site in repeat 1C of the wire, and insertions in wire repeat 11C and at the C-terminus impact the overall fold of various group B THBSs. It is certain, however, that the group B THBSs do have significant structural differences from group A THBSs, as evidenced by biophysical studies on homologous segments of THBS-2 and THBS-4 [43].

Many binding partners have been identified for THBSs, too many to consider comprehensively in this review. These interactions need to be reinvestigated taking into account what is now known about structure. This is particularly true for the signature domain, which was unexpectedly found to contain epitopes for multiple monoclonal antibodies that block frunctions of THBS-1 [56]. Several interaction sites deduced by peptides are obscured in the published structures, raising the possibility of alternate structures in which these sequences are exposed. Future investigations into THBS-protein interactions may include examination of high-resolution structures of THBS-protein complexes. Finally, more investigation is needed to understand fully the dynamic nature of THBS structure as a response to calcium concentration and disease-linked genetic alterations.

Acknowledgements. We would like to thank Dr. Ken Satyshur for his help in creating the trimeric model of a group A THBS. This work was supported by US National Institutes of Health grants HL54462 (D. F. M.) and HL49081 (J. L.). C. B. C. was supported by US National Institutes of Health training grant HL07899.

1 Baenziger, N. L., Brodie, G. N. and Majerus, P. W. (1971) A thrombin-sensitive protein of human platelet membrane. Proc. Natl. Acad. Sci. USA 68, 240–243.

2 Lawler, J. W., Slayter, H. S. and Coligan, J. E. (1978) Isolation and characterization of a high molecular weight glycoprotein from human blood platelets. J. Biol. Chem. 253, 8609–8616.
3 Adams, J. C. and Lawler, J. (2004) The thrombospondins. Int. J. Biochem. Cell. Biol. 36, 961–968.
4 Adams, J. C. (2004) Functions of the conserved thrombospondin carboxy-terminal cassette in cell-extracellular matrix interactions and signaling. Int. J. Biochem. Cell. Biol. 36, 1102–1114.
5 Tan, K., Duquette, M., Liu, J. H., Zhang, R. G., Joachimiak, A., Wang, J. H. and Lawler, J. (2006) The structures of the thrombospondin-1 N-terminal domain and its complex with a synthetic pentameric heparin. Structure 14, 33–42.
6 Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T. and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? Science 274, 761–765.
7 Dames, S. A., Kammerer, R. A., Wiltscheck, R., Engel, J. and Alexandrescu, A. T. (1998) NMR structure of a parallel homotrimeric coiled coil. Nat. Struct. Biol. 5, 687–691.
8 O'Leary, J. M., Hamilton, J. M., Deane, C. M., Valeyev, N. V., Sandell, L. J. and Downing, A. K. (2004) Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand factor type c module) from collagen IIA. J. Biol. Chem. 279, 53857–53866.
9 Tan, K. M., Duquette, M., Liu, J. H., Dong, Y. C., Zhang, R. G., Joachimiak, A., Lawler, J. and Wang, J. H. (2002) Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. J. Cell Biol. 159, 373–382.
10 Kvansakul, M., Adams, J. C. and Hohenester, E. (2004) Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. EMBO 23, 1223–1233.
11 Carlson, C. B., Bernstein, D. A., Annis, D. S., Misenheimer, T. M., Hannah, B.-I. A., Mosher, D. F. and Keck, J. L. (2005) Structure of the calcium-rich signature domain of human thrombospondin-2. Nat. Struct. Mol. Biol. 12, 910–914.
12 Chen, H., Herndon, M. E. and Lawler, J. (2000) The cell biology of thrombospondin-1. Matrix Biol. 19, 597–614.
13 Elzie, C. A. and Murphy-Ullrich, J. E. (2004) The N-terminus of thrombospondin: the domain stands apart. Int. J. Biochem. Cell Biol. 36, 1090–1101.
14 Harbury, P. B., Kim, P. S. and Alber, T. (1994) Crystal structure of an isoleucine-zipper trimer. Nature 371, 80–83.
15 Sottile, J., Selegue, J. and Mosher, D. F. (1991) Synthesis of truncated amino-terminal trimers of thrombospondin. Biochemistry 30, 6556–6562.
16 Lee, N. V., Sato, M., Annis, D. S., Loo, J. A., Wu, L., Mosher, D. F. and Iruela-Arispe, M. L. (2006) ADAMTS1 mediates the release of antiangiogenic polypeptides from TSP1 and 2. EMBO J. 25, 5270–5283.
17 Misenheimer, T. M., Huwiler, K. G., Annis, D. S. and Mosher, D. F. (2000) Physical characterization of the procollagen module of human thrombospondin 1 expressed in insect cells. J. Biol. Chem. 275, 40938–40945.
18 Tolsma, S. S., Volpert, O. V., Good, D. J., Frazier, W. A., Polverini, P. J. and Bouck, N. (1993) Peptides derived from two separate domains of the matrix protein thrombospondin-1 have anti-angiogenic activity. J. Cell Biol. 122, 497–511.
19 Tan, K., Duquette, M., Liu, J. H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J. and Wang, J. H. (2002) Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. J. Cell. Biol. 159, 373–382.
20 Huwiler, K. G., Vestling, M. M., Annis, D. S. and Mosher, D. F. (2002) Biophysical characterization, including disulfide bond assignments, of the anti-angiogenic type 1 domains of human thrombospondin-1. Biochemistry 41, 14329–14339.
21 Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F. and Peter-Katalinic, J. (2001) C-mannosylation and O-fucosylation of the thrombospondin type 1 module. J. Biol. Chem. 276, 6485–6498.
22 Ricketts, L. M., Dlugosz, M., Luther, K. B., Haltiwanger, R. S. and Majerus, E. M. (2007) O-Fucosylation is required for ADAMTS13 secretion. J. Biol. Chem. 282, 17014–17023.
23 Wang, L. W., Dlugosz, M., Somerville, R. P. T., Raed, M., Haltiwanger, R. S. and Apte, S. S. (2007) O-Fucosylation of thrombospondin type 1 repeats in ADAMTS-like-1/Punctin-1 regulates secretion: implications for the ADAMTS superfamily. J. Biol. Chem. 282, 17024–17031.
24 Simantov, R. and Silverstein, R. L. (2003) CD36: a critical anti-angiogenic receptor. Front. Biosci. 8, s874–s882.
25 Calzada, M. J., Annis, D. S., Zeng, B., Marcinkiewicz, C., Banas, B., Lawler, J., Mosher, D. F. and Roberts, D. D. (2004) Identification of novel beta1 integrin binding sites in the type 1 and type 2 repeats of thrombospondin-1. J. Biol. Chem. 279, 41734–41743.
26 Short, S. M., Derrien, A., Narsimhan, R. P., Lawler, J., Ingber, D. E. and Zetter, B. R. (2005) Inhibition of endothelial cell migration by thrombospondin-1 type-1 repeats is mediated by {beta}1 integrins. J. Cell Biol. 168, 643–653.
27 Markovic, S. N., Suman, V. J., Rao, R. A., Ingle, J. N., Kaur, J. S., Erickson, L. A., Pitot, H. C., Croghan, G. A., McWilliams, R. R., Merchan, et al. (2007) A phase II study of ABT-510 (thrombospondin-1 analog) for the treatment of metastatic melanoma. Am. J. Clin. Oncol. 30, 303–309.
28 Young, G. D. and Murphy-Ullrich, J. E. (2004) The tryptophan-rich motifs of the thrombospondin type 1 repeats bind VLAL motifs in the latent transforming growth factor-{beta} complex. J. Biol. Chem. 279, 47633–47642.
29 Young, G. D. and Murphy-Ullrich, J. E. (2004) The tryptophan-rich motifs of the thrombospondin type 1 repeats bind VLAL motifs in the latent transforming growth factor-beta complex. J. Biol. Chem. 279, 47633–47642.
30 Boswell, E. J., Kurniawan, N. D. and Downing, A. K. (2004) Calcium-binding EGF-like domains. In: Handbook of Metalloproteins vol. 3, pp. 553–570, Messerschmidt, A., Bode, W. and Cygler, M. (ed.), John Wiley, Chichester.
31 Yamano, K., Qiu, G. F. and Unuma, T. (2004) Molecular cloning and ovarian expression profiles of thrombospondin, a major component of cortical rods in mature oocytes of penaeid shrimp, Marsupenaeus japonicus. Biol. Reprod. 70, 1670–1678.
32 Wouters, M. A., Rigoutsos, I., Chu, C. K., Feng, L. L., Sparrow, D. B. and Dunwoodie, S. L. (2005) Evolution of distinct EGF domains with specific functions. Protein Sci. 14, 1091–1103.
33 Drakenberg, T., Ghasriani, H., Thulin, E., Thamlitz, A.-M., Muranyi, A., Annila, A. and Stenflo, J. (2005) Solution structure of the Ca2+-binding EGF3–4 pair from vitamin K-dependent protein S: identification of an unusual fold in EGF3. Biochemistry 44, 8782–8789.
34 Nishimura, H., Yamashita, S., Zeng, Z., Walz, D. A. and Iwanaga, S. (1992) Evidence for the existence of O-linked sugar chains consisting of glucose and xylose in bovine thrombospondin. J. Biochem. (Tokyo) 111, 460–464.
35 Kennedy, J., Jackson, G. C., Barker, F. S., Nundlall, S., Bella, J., Wright, M. J., Mortier, G. R., Neas, K., Thompson, E., Elles, R. and Briggs, M. D. (2005) Novel and recurrent mutations in the C-terminal domain of COMP cluster in two distinct regions and result in a spectrum of phenotypes within the pseudoachondroplasia – multiple epiphyseal dysplasia disease group. Hum. Mutat. 25, 593–494.
36 Posey, K. L., Hayes, E., Haynes, R. and Hecht, J. T. (2004) Role of TSP-5/COMP in pseudoachondroplasia. Int. J. Biochem. Cell Biol. 36, 1005–1012.
37 Vranka, J., Mokashi, A., Keene, D. R., Tufa, S., Corson, G., Sussman, M., Horton, W. A., Maddox, K., Sakai, L. and Bachinger, H. P. (2001) Selective intracellular retention of extracellular matrix proteins and chaperones associated with pseudoachondroplasia. Matrix Biol. 20, 439–450.
38 Hecht, J. T. Hayes, E., Snuggs, M., Decker, G., Montufar-Solis, D., Doege, K., Mwalle, F., Poole, R., Stevens, J. and Duke, P. J. (2001) Calreticulin, PDI, Grp94 and BiP chaperone proteins are associated with retained COMP in pseudoachondroplasia chondrocytes. Matrix Biol. 20, 251–262.

39 Dinser, R., Zaucke, F., Kreppel, F., Hultenby, K., Kochanek, S., Paulsson, M. and Maurer, P. (2002) Pseudoachondroplasia is caused through both intra-and extracellular pathogenic pathways. J. Clin. Invest. 110, 505–513.
40 Topol, E. J., McCarthy, J., Gabriel, S., Moliterno, D. J., Rogers, W. J., Newby, L. K., Freedman, M., Metivier, J., Cannata, R., O'Donnell, C. J. et al. (2001) Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. Circulation 104, 2641–2644.
41 Hannah, B. L., Misenheimer, T. M., Pranghofer, M. M. and Mosher, D. F. (2004) A polymorphism in thrombospondin-1 associated with familial premature coronary artery disease alters Ca2+ binding. J. Biol. Chem. 279, 51915–51922.
42 Hannah, B. L., Misenheimer, T. M., Annis, D. S. and Mosher, D. F. (2003) A polymorphism in thrombospondin-1 associated with familial premature coronary heart disease causes a local change in conformation of the Ca2+-binding repeats. J. Biol. Chem. 278, 8929–8934.
43 Misenheimer, T. M. and Mosher, D. F. (2005) Biophysical characterization of the signature domains of thrombospondin-4 and thrombospondin-2. J. Biol. Chem., M504696200.
44 Misenheimer, T. M., Hannah, B. L., Annis, D. S. and Mosher, D. F. (2003) Interactions among the three structural motifs of the C-terminal region of human thrombospondin-2. Biochemistry 42, 5125–32.
45 Adams, J. C. and Lawler, J. (1994) Cell-type specific adhesive interactions of skeletal myoblasts with thrombospondin-1. Mol. Biol. Cell 5, 423–437.
46 Misenheimer, T. M., Hahr, A. J., Harms, A. C., Annis, D. S. and Mosher, D. F. (2001) Disulfide connectivity of recombinant C-terminal region of human thrombospondin 2. J. Biol. Chem. 276, 45882–45887.
47 Kleerekoper, Q., Hecht, J. T. and Putkey, J. A. (2002) Disease-causing mutations in cartilage oligomeric matrix protein cause an unstructured Ca2+ binding domain. J. Biol. Chem. 277, 10581–10589.
48 Thur, J., Rosenberg, K., Nitsche, D. P., Pihlajamaa, T., Ala-Kokko, L., Heinegard, D., Paulsson, M. and Maurer, P. (2001) Mutations in cartilage oligomeric matrix protein causing pseudoachondroplasia and multiple epiphyseal dysplasia affect binding of calcium and collagen I, II and IX. J. Biol. Chem. 276, 6083–6092.
49 Chen, H., Deere, M., Hecht, J. T. and Lawler, J. (2000) Cartilage oligomeric matrix protein is a calcium-binding protein, and a mutation in its type 3 repeats causes conformational changes. J. Biol. Chem. 275, 26538–26544.
50 Maddox, B. K., Mokashi, A., Keene, D. R. and Bachinger, H. P. (2000) A cartilage oligomeric matrix protein mutation associated with pseudoachondroplasia changes the structural and functional properties of the type 3 domain. J. Biol. Chem. 275, 11412–11417.
51 Holden, P., Meadows, R. S., Chapman, K. L., Grant, M. E., Kadler, K. E. and Briggs, M. D. (2001) Cartilage oligomeric matrix protein interacts with type IX collagen, and disruptions to these interactions identify a pathogenetic mechanism in a bone dysplasia family. J. Biol. Chem. 276, 6046–6055.
52 Lawler, J. and Hynes, R. O. (1989) An integrin receptor on normal and thrombasthenic platelets that binds thrombospondin. Blood 74, 2022–2027.
53 Chen, H., Sottile, J., O'Rourke, K. M., Dixit, V. M. and Mosher, D. F. (1994) Properties of recombinant mouse thrombospondin 2 expressed in Spodoptera cells. J. Biol. Chem. 269, 32226–32232.
54 Hogg, P. J., Jiménez, B. M. and Chesterman, C. N. (1994) Identification of possible inhibitory reactive centers in thrombospondin 1 that may bind cathepsin G and neutrophil elastase. Biochemistry 33, 6531–6537.
55 Gao, A. G. and Frazier, W. A. (1994) Identification of a receptor candidate for the carboxyl-terminal cell binding domain of thrombospondins. J. Biol. Chem. 269, 29650–29657.
56 Annis, D. S., Murphy-Ullrich, J. E. and Mosher, D. F. (2006) Function-blocking antithrombospondin-1 monoclonal antibodies. J. Thromb. Haemost. 4, 459–68.
57 Galvin, N. J., Vance, P. M., Dixit, V. M., Fink, B. and Frazier, W. A. (1987) Interaction of human thrombospondin with types I-V collagen: direct binding and electron microscopy. J. Cell Biol. 104, 1413–1422.
58 Pimanda, J. E., Annis, D. S., Raftery, M., Mosher, D. F., Chesterman, C. N. and Hogg, P. J. (2002) The von Willebrand factor-reducing activity of thrombospondin-1 is located in the calcium-binding/C-terminal sequence and requires a free thiol at position 974. Blood. 100, 2832–2838.
59 Adams, J. C., Tucker, R. P. and Lawler, J. (1995) In: The Thrombospondin Gene Family, R. G. Landes, Austin.
60 Lawler, J. and Simons, E. R. (1983) Cooperative binding of calcium to thrombospondin. The effect of calcium on the circular dichroism and limited tryptic digestion of thrombospondin. J. Biol. Chem. 258, 12098–12101.
61 Chen, H., Aeschlimann, D., Nowlen, J. and Mosher, D. (1996) Expression and initial characterization of recombinant mouse thrombospondin 1 and thrombospondin 3. FEBS Lett. 387, 36–41.
62 Annis, D. S., Gunderson, K. A. and Mosher, D. F. (2007) Immunochemical analysis of the structure of the signature domains of thrombospondin-1 and thrombospondin-2 in low calcium concentrations. J. Biol. Chem., M703804200.