Hemicentin Assembly in the Extracellular Matrix Is Mediated by Distinct Structural Modules

**Chun Dong, Joaquin M. Muriel, Sarah Ramirez, Harald Hutter, Edward M. Hedgecock, Leonid Breydo, Ilia V. Baskakov, and Bruce E. Vogel**

**From the Medical Biotechnology Center, University of Maryland Biotechnology Institute, Baltimore, Maryland 21201, the Max-Planck-Institut für Medizinische Forschung, D-69120 Heidelberg, Germany, and the Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218**

Hemicentins are conserved extracellular matrix proteins characterized by a single von Willebrand A (VWA) domain at the amino terminus, a long stretch (>40) of tandem immunoglobulin domains, multiple tandem epidermal growth factors (EGFs), and a single fibulin-like carboxyl-terminal module. In *Caenorhabditis elegans*, hemicentin is secreted from muscle and gonadal leader cells and assembles at multiple locations into discrete tracks that constrict broad regions of cell contact into adhesive and flexible line-shaped junctions. To determine hemicentin domains critical for function and assembly, we have expressed fragments of hemicentin as GFP tagged fusion proteins in *C. elegans*. We find that a hemicentin fragment containing the VWA domain can target to multiple assembly sites when expressed under the control of either endogenous hemicentin or after secretion as a result of proteolytic processing or interaction with cell surface receptors. Activation changes the structure of specific domains and results in increased homophilic interactions or heterophilic interactions with other ECM proteins as molecules assemble into higher order structures.

Hemicentins are a family of secreted ECM proteins first identified in *Caenorhabditis elegans* with two orthologs in most vertebrate genomes including human and mouse (6). Hemicentins are characterized by a single highly conserved von Willebrand A (VWA) domain near the amino terminus, followed by a long stretch (>40) of tandem immunoglobulin (Ig) domains, multiple tandem epidermal growth factors (EGFs), and a single fibulin-like carboxyl-terminal (FC) domain (6–8). In *C. elegans*, hemicentin is secreted by bodywall muscle and gonadal leader cells and assemblies at multiple distant locations, where it forms an oriented track-like geometry whose apparent function is to constrict broad regions of cell contact into discrete line-shaped junctions (6). There is nothing known about the mechanism of hemicentin recruitment to cell surfaces or about its assembly there.

Hemicentin tracks anchor mechanosensory neuron and uterine cells at hemidesmosome-mediated attachments and are required for cellularization of germ cells and migration of somatic male gonadal cells (6). Hemicentin expression by the proximal gonadal cell in hermaphrodites (i.e. the anchor cell) is controlled by the *fog-1* transcription factor, and accumulation of hemicentin on the anchor cell surface promotes anchor cell invasion into vulval epithelium to form a functional connection between the uterus and vulva (9).

To understand the role of individual hemicentin structural domains in hemicentin function and assembly, we have expressed hemicentin and hemicentin fragments as green fluorescent protein (GFP)-tagged fusion proteins in *C. elegans* and present data that suggest the function of the VWA domain is to target hemicentin to sites of assembly. The carboxyl-terminal EGF/FC domain contains a motif sufficient to bind existing hemicentin polymers in wild-type animals, suggesting that this is likely to constitute an assembly domain. Based on this data we suggest a multistep model for hemicentin assembly similar to that of other extracellular matrix components, involving cell binding, activation, and propagation (1).
Hemicentin Assembly Domains

Full-length *C. elegans* hemicentin comprises 5198 amino acids and contains a VWA domain followed by 48 tandem Ig modules, 3 EGF modules, and a FC module. To test the function of these domains in hemicentin assembly, we designed several hemicentin based constructs with various domains deleted or mutated (Fig. 1) and introduced them into hemicentin containing wild-type animals and hemicentin-deficient (*him-4*/*rh319*) mutant animals. Each construct contains a signal sequence followed by GFP and is under the control of endogenous *him-4* regulatory sequences (F15G9::12,861–14,017) with the exception of *unc-54:*1–433, which is under the control of the muscle-specific *unc-54* promoter (13). Constructs 1–433, *unc-54:*1–433 and D42A contain nucleotides encoding the initiator methionine up to but not including the first Ig module. Construct 1–1492 includes nucleotide sequences from the initiator methionine up to but not including the first Ig module. Construct 434–5198 includes all amino acids that follow Ig48 (Fig. 1). Constructs 1–1492 and D42A include nucleotides encoding the hemicentin initiator methionine to Ig11. Construct 434–5198 includes all amino acids and contains a VWA domain followed by 48 tandem Ig modules, 3 EGF modules, and a FC module. To test the function of these domains in hemicentin assembly, we designed several hemicentin based constructs with various domains deleted or mutated (Fig. 1) and introduced them into hemicentin containing wild-type animals and hemicentin-deficient (*him-4*/*rh319*) mutant animals. Each construct contains a signal sequence followed by GFP and is under the control of endogenous *him-4* regulatory sequences (F15G9::12,861–14,017) with the exception of *unc-54:*1–433, which is under the control of the muscle-specific *unc-54* promoter (13). Constructs 1–433, *unc-54:*1–433 and D42A contain nucleotides encoding the initiator methionine up to but not including the first Ig module. Construct 1–1492 includes nucleotide sequences from the initiator methionine up to but not including the first Ig module. Construct 434–5198 includes all amino acids that follow Ig48 (Fig. 1).

Expression of each of these constructs is detected by Western blot with anti-GFP antibodies and observed sizes match predicted sizes (Fig. 2). The observed size of full-length hemicentin::GFP matches the predicted molecular weight based on amino acid sequence of ~600 kDa, suggesting that the most abundant form of hemicentin present in wild-type animals is a full-length monomer that does not appear subject to either extensive proteolysis or post-translational modification.

**RESULTS**

Full-length *C. elegans* hemicentin comprises 5198 amino acids and contains a VWA domain followed by 48 tandem Ig modules, 3 EGF modules, and a FC module. To test the function of these domains in hemicentin assembly, we designed several hemicentin based constructs with various domains deleted or mutated (Fig. 1) and introduced them into hemicentin containing wild-type animals and hemicentin-deficient (*him-4*/*rh319*) mutant animals. Each construct contains a signal sequence followed by GFP and is under the control of endogenous *him-4* regulatory sequences (F15G9::12,861–14,017) with the exception of *unc-54:*1–433, which is under the control of the muscle-specific *unc-54* promoter (13). Constructs 1–433, *unc-54:*1–433 and D42A contain nucleotides encoding the initiator methionine up to but not including the first Ig module. Construct 1–1492 includes nucleotide sequences from the initiator methionine up to but not including the first Ig module. Construct 434–5198 includes all amino acids that follow Ig48 (Fig. 1). Expression of each of these constructs is detected by Western blot with anti-GFP antibodies and observed sizes match predicted sizes (Fig. 2). The observed size of full-length hemicentin::GFP matches the predicted molecular weight based on amino acid sequence of ~600 kDa, suggesting that the most abundant form of hemicentin present in wild-type animals is a full-length monomer that does not appear subject to either extensive proteolysis or post-translational modification.
As shown previously, hemicentin can assemble into fine, regularly spaced tracks that pivot and flex during foraging. These tracks are not associated with cell surfaces other than at attachment points at the pharyngeal basement membrane at one end and between lateral and medial bodywall muscle cells at the other end (6). Full-length GFP::hemicentin can assemble into tracks in the presence or absence of endogenous hemicentin (Fig. 3; Ref. 6). When we expressed constructs 434–5198 and 4838–5198 we found that these constructs are sufficient to bind to existing hemicentin tracks found in wild-type animals but cannot form these structures in the absence of endogenous hemicentin, i.e. hemicentin mutant background (C, F, and I). Also shown is full-length hemicentin (A) and truncated amino-terminal constructs (D and G).

Figure 3. Ability of constructs to bind to flexible track structures that connect pharynx and anterior bodywall muscles. Constructs shown in Figure 1 were expressed in wild-type (WT) or hemicentin mutant (him-4) animals. Constructs 434–5198, 4838–5198, and 1–4837 have the ability to bind to flexible tracks in wild-type animals (arrows in B, E, and H) but cannot form these structures in the absence of endogenous hemicentin, i.e. hemicentin mutant background (C, F, and I). Also shown is full-length hemicentin (A) and truncated amino-terminal constructs (D and G).

Although constructs located near the amino-terminal were not able to bind to hemicentin tracks (Fig. 3, D and G), we observed that they did have the ability to bind to cell surfaces where hemicentin tracks normally assemble in wild-type animals including mechanosensory neurons, uterine attachments, and germ cell surfaces (Fig. 4). This was true for constructs containing the first 1492 amino acids (Fig. 4, J–L) as well as constructs limited to the first 433 amino acids of hemicentin (Fig. 4, D–F). Even in hemicentin mutant (him-4 (rh319) and him-4 (hd999)) backgrounds, the first 433 amino acids were still able to bind to uterine attachments and germ cell surfaces (Fig. 4, H and I) suggesting that these hemicentin fragments were not binding to endogenous hemicentin but binding directly to cell surface molecules or extracellular molecules associated with those cell surfaces.

To determine whether the VWA motif or other sequences within hemicentin 1–433 were important for cell surface localization, we introduced a mutation in the conserved metal ion-dependent adhesion site (14). The conserved sequence DXTXS was mutated by substituting an alanine codon for the aspartate codon. When expressed in transgenic animals, the ability of this construct (D42A) to bind cell surfaces was dramatically reduced (cf Fig. 4, D–F and P–R), although similar protein levels were detected in both strains (Fig. 2).

Because the first 433 amino acids of hemicentin seemed to be sufficient for targeting to cell surfaces, we deleted this domain to determine whether this domain was necessary for hemicentin targeting. We found that a construct without the VWA domain (434–5198) retained the ability to target cell surfaces in the presence or absence of endogenous hemicentin, suggesting the presence of an additional cell binding domain in hemicentin (Fig. 4, S–U).
Hemicentin Assembly Domains

**Figure 4.** Binding of hemicentin fragments to surfaces of mechanosensory neurons, germ cells, and uterine attachments. A construct containing only amino acids 1–433 can bind to mechanosensory neuron surfaces (D and G), germ cell surfaces (E and H), and uterine attachments (F and I) in the presence of endogenous hemicentin, i.e. wild-type background (D–F) and in the absence of endogenous hemicentin, (him-4 mutant background, G–I). This is true when the construct is expressed under the endogenous him-4 promoter or when it is expressed from a muscle-specific unc-54 promoter (M–O). Binding is reduced when amino acid 42 is mutated from aspartate to alanine (P–R).

**DISCUSSION**

Extracellular matrix proteins assemble into networks with a variety of geometries that provide structural and regulatory information to cells to pattern tissue histogenesis and morphogenesis. In *C. elegans*, hemicentin assembles into discrete tracks that contract broad regions of cell contact into discrete line-shaped junctions. To understand the mechanism of hemicentin assembly, we expressed fragments of hemicentin as GFP fusion proteins in wild-type and hemicentin-deficient animals. Below, we discuss our findings and present a model for the assembly of hemicentin that incorporates this data.

The EGF/FC Modules Constitute an Assembly Domain of Hemicentin—A construct containing GFP fused to the EGF/FC modules of hemicentin (4838–5198) binds to hemicentin containing structures that are present in wild-type animals (Fig. 3). Hemicentin is found in flexible structures that extend from body wall muscle cell surfaces to the pharynx basement membrane and is not associated with cell surfaces other than at these attachment sites. Because hemicentin and fibulin are the only proteins that have been identified in these structures (15), this observation raises the possibility that one function of the EGF/FC modules is to mediate direct hemicentin–hemicentin interactions.

Even with the EGF/FC modules missing, other hemicentin–hemicentin binding sites are still present that can bind to EGF/FC modules present in endogenous hemicentin molecules that are found in the flexible tracks of wild-type animals. It is not surprising for an extracellular matrix molecule to have multiple self-assembly domains, because a minimum of two self-assembly domains are required for assembly into a polymer larger than a dimer. Fibronectin, for example, assembles into large, complex fibrillar structures and has at least four self-assembly domains (2). We speculate that the large number of Ig domains found in hemicentin reflect not only homophilic interactions but interactions with other binding partners. The Ig domains in perlecain, for example, are known to be involved in direct protein-protein interactions with multiple partners including nidogen and fibulin-2 (16).

Interestingly, the combination of EGF and FC modules seems to form a single functional unit, because we could not subdivide this region into smaller functional domains. This combination is also found in the two vertebrate hemicentins as well as all five fibulin family members (8). This striking similarity has resulted in occasional reference to human hemicentin-1 as fibulin-6. A *C. elegans* fibulin-1 ortholog co-assembles with hemicentin on mechanosensory neurons, uterine attachments, and flexible tracks (15). It is tempting to speculate that like hemicentin, fibulin uses the EGF/FC modules to interact directly with hemicentin as the 2 molecules co-assemble at these sites.

A mutation in the last EGF of human hemicentin-1 was found linked to the disease phenotype in a familial case of age-related macular degeneration (17). Based on data presented here, we suggest that the effect of this dominant mutation is likely to interfere with the hemicentin assembly process.

The VWA Domain of Hemicentin Is a Cell-targeting Domain—VWA domains are known to be frequently involved in mediating divalent cation-dependent interactions between proteins assembled into multicomponent complexes. A VWA domain containing fragment of hemicentin (1–433) targets to uterine attachments and germ cell surfaces in the presence or absence of endogenous hemicentin (Fig. 4). Targeting occurs when 1–433 is expressed under the control of endogenous hemicentin regulatory sequences or when expressed under the control of the muscle-specific unc-54 promoter. It is likely that this is a key domain in targeting hemicentin to different sites of assembly and that hemicentin binds to cell surface receptors or other extracellular matrix protein found at assembly sites. The binding is likely to be dependent on divalent cations because the cell-targeting activity of 1–433 is dramatically reduced when a point mutation is introduced into the highly conserved divalent cation binding sequences of the VWA domain metal-ion dependent adhesion site.

Structurally, the VWA domain is the most highly conserved region of hemicentin. This raises the possibility that the VWA domain has a similar function, that of targeting hemicentins to...
Hemicentin Assembly Domains

FIGURE 5. Model of hemicentin assembly. Step 1 (secretion), hemicentin is secreted from muscle cells as a monomer. Potential intramolecular interactions involving the carboxyl-terminal EGF/FC modules may result in newly secreted hemicentin forming compact "blocked" structures where the EGF/FC modules are sequestered, preventing hemicentin assembly in ectopic locations. Steps 2 and 3 (nucleation and activation), the blocked molecule encounters cell surface or other ECM molecules that bind to the VWA or Ig domain cause the blocked molecule to unfold, exposing homophilic binding sites in the carboxyl-terminal EGF/FC modules. Step 4 (propagation), carboxyl-terminal EGF/FC modules bind to homophilic intermolecular binding sites elsewhere in the molecule in either a parallel or anti-parallel orientation resulting in the assembly of hemicentin into polymeric structures of increasing length. Step 5, this assembly process is likely to incorporate fibulins that contain EGF and FC modules (blue/red) and possibly other ECM proteins.

cell surfaces in vertebrate organisms as well. Another molecule, G7c or NG37, has been identified in vertebrate genomes that has a VWA motif very similar to the VWA found in hemicentins (6, 7, 18) and may indicate that this domain targets the secreted G7c/NG37 molecule to hemicentin binding cell surfaces as well.

The observation that a hemicentin construct where amino acids 1–433 have been deleted still binds to urethane attachments and germ cell surfaces even in the absence of endogenous hemicentin suggests the possibility of a second cell binding site located elsewhere in the molecule. This is not unexpected for a large extracellular matrix molecule, many of which have multiple cell binding domains. Fibronectin, for example, has multiple cell binding sites that bind to distinct members of the integrin family of adhesion receptors (1).

Model for Hemicentin Assembly—We have summarized the results and incorporated these observations into a model of hemicentin assembly (Fig. 5). Step 1 (secretion): in C. elegans, hemicentin is secreted from muscle and gonadal leader cells. Potential intramolecular interactions between the carboxyl-terminal EGF/FC modules and other domains within hemicentin may result in newly secreted hemicentin forming compact “blocked” structures where the EGF/FC modules are sequestered, preventing hemicentin assembly in ectopic locations. Steps 2 (nucleation) and 3 (activation): the blocked molecule encounters a cell surface or ECM molecule that binds to the VWA and/or sites in the Ig domain, causing the monomer to unfold, exposing homophilic binding sites in the carboxyl-terminal EGF/FC modules and elsewhere in the molecule. Step 4 (propagation): additional hemicentin molecules are added to the nascent structure as carboxyl-terminal EGF/FC modules bind to homophilic binding sites. At this point, we cannot determine whether additional hemicentin molecules are added in a parallel or anti-parallel orientation. The next step in the assembly of hemicentin is likely to involve direct interactions with other hemicentin molecules and possibly other ECM components (Step 5). An obvious candidate to interact with hemicentin is fibulin, which, as mentioned above, also contains EGF/FC motifs and has been shown recently to co-assemble with hemicentin in flexible tracks and hemidesmosome-mediated uterine and mechanosensory neuron attachments (15).

Acknowledgments—We thank Dr. Xuehong Xu for valuable help and advice and Dr. A. Fire for cloning vectors. Some nematode strains used in this work were obtained from the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources.

REFERENCES
1. Schwarzbauer, J. E., and Sechler, J. L. (1999) Curr. Opin. Cell Biol. 11, 622–627
2. Wierzbicka-Patynowski, I., and Schwarzbauer, J. E. (2003) J. Cell Sci. 116, 3269–3276
3. Sasaki, T., Fassler, R., and Hohenester, E. (2004) J. Cell Biol. 164, 959–963
4. Yurchenco, P. D., Amenta, P. S., and Patton, B. L. (2004) Matrix Biol. 22, 521–538
5. Li, S., Liquari, P., McKee, K. K., Harrison, D., Patel, R., Lee, S., and Yurchenco, P. D. (2005) J. Cell Biol. 169, 179–189
6. Vogel, B. E., and Hedgecock, E. M. (2001) Development (Camb.) 128, 883–894
7. Whittaker, C. A., and Hynes, R. O. (2002) Mol. Biol. Cell 13, 3369–3387
8. Argraves, W. S., Greene, L. M., Cooley, M. A., and Gallagher, W. M. (2003) EMBO Rep. 4, 1127–1131
9. Sherwood, D. R., Butler, J. A., Kramer, J. M., and Sternberg, P. W. (2005) Cell 121, 951–962
10. Horton, R. M. (1995) Mol. Biotechnol. 3, 93–99
11. Chartier, C., Degryse, E., Gantzer, M., Dieiterle, A., Pavirani, A., and Mehtali, M. (1996) J. Virol. 70, 4805–4810
12. Andrade, M. A., Chacón, P., Merelo, J. J., and Morán, F. (1993) Proc. Natl. Acad. Sci. USA 90, 5383–5390
13. Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A., and Fire, A. (1993) EMBO J. 12, 385–404
14. Lee, I. O., Rieu, P., Arnaout, M. A., and Liddington, R. (1995) Cell 80, 631–638
15. Muriel, J. M., Dong, C., Hutter, H., and Vogel, B. E. (2005) Development (Camb.) 132, 4223–4234
16. Hopf, M., Gohring, W., Mann, K., and Timpl, R. (2001) J. Mol. Biol. 311, 529–541
17. Schultz, D. W., Klein, M. L., Humplik, A. J., Luzier, C. W., Persun, V., Schain, M., Mahan, A., Runckel, C., Cassera, M., Vittal, V., Doyle, T. M., Weleber, R. G., Francis, P. J., and Acott, T. S. (2003) Hum. Mol. Genet. 12, 3315–3323
18. Kumanovics, A., and Lindahl, K. F. (2001) Immunogenetics 53, 64–68