The CCN (cyr61, ctgf, nov) proteins (CCN1–6) are an important family of matricellular regulatory factors involved in internal and external cell signaling. They are central to essential biological processes such as adhesion, proliferation, angiogenesis, tumorigenesis, wound healing, and modulation of the extracellular matrix. They possess a highly conserved modular structure with four distinct modules that interact with a wide range of regulatory proteins and ligands. However, at the structural level, little is known although their biological function(s) seems to require cooperation between individual modules. Here we present for the first time structural determinants of two of the CCN family members, CCN3 and CCN5 (expressed in Escherichia coli), using small angle x-ray scattering. The results provide a description of the overall molecular shape and possible general three-dimensional modular arrangement for CCN proteins. These data unequivocally provide insight of the nature of CCN protein(s) in solution and thus important insight into their structure-function relationships.

The CCN proteins are an important family of matricellular proteins that modulate cellular signaling and play a role in regulating critical cellular functions (for recent reviews, see Refs. 1 and 2). There are now six recognized members of the CCN family. The first three to be discovered, Cyr61 (cysteine-rich protein 61) (CCN1) (3), CTGF (connective tissue growth factor) (CCN2) (4), and Nov (nephroblastoma overexpressed gene) (CCN3) (5), gave rise to the CCN acronym (6). The remaining three members are the Wnt-inducible proteins Wisp1–3 (CCN4–6) (7). The standardized nomenclature (8) was necessary as many of the proteins had other names dependent on observed function; for example, CCN1 was previously known as Cyr61, CTGF-2, IGFBP10, and IGFBP-rP4 (8).

The CCN proteins are heavily involved in angiogenesis, chondrogenesis, proliferation, adhesion, wound healing, tumorigenesis, and cell survival (9–11). Although all of the CCN proteins have been shown to play some role in basic functions such as proliferation, adhesion, and migration, the role of certain CCNs play in more complex biological functions, such as CCN1 and CCN2 in angiogenesis or CCN4 and CCN6 in cancer, is more pronounced (11–13). However, the biochemical basis, and importantly, the underlying protein-protein interactions that lead to these functions are not understood. CCN3, like all CCN proteins, is involved in many biological functions including cancer (14), regulation of angiogenesis (15), and cell adhesion (16). CCN5 was first identified in the WISP pathway and has been implicated heavily for its role in several types of cancers, most importantly breast (17) and colon (7) cancers, and for its role in bone turnover and chondrogenesis (18).

The complexity of CCN behavior is due to their large repertoire of ligands. These include the integrins, which are considered to be the “functional receptors of the CCN family”; the small cysteine knot growth factors (vascular endothelial growth factor (VEGF), transforming growth factor (TGF), and bone morphogenic protein (BMP)); extracellular matrix proteins (collagen, fibulin, and Notch); heparin sulfated proteoglycans; LDL receptor protein-1; and insulin growth factor (11, 12, 19). In some cases, the module that these receptors bind is known. VEGF isoforms bind to domains 3 and 4 of CCN2 (20) and domain 2 of CCN3 binds BMP2 (21). Aggrecan binds to the N terminus of CCN2 (either domain 1 or domain 2) (22), and domain 3 and domain 4 facilitate binding to heparin sulfated proteoglycans (23) and fibulin (16). For some functions, only a single domain is needed, but for others, multiple domains acting in concert are required (24). This raises questions about how the shape of the molecule influences its biological function.

The CCN proteins, like many extracellular matrix (ECM) proteins, are constructed from a series of distinct domains (25) that have evolved through exon shuffling (6, 26, 27). The building blocks that make up the CCN family are: 1) an insulin-like growth factor binding domain (IGFBP); 2) a von Willebrand type C repeat (VWC); 3) a thrombospondin type I repeat (TSP); and 4) a cysteine knot containing a carboxyl-terminal domain (CT) (6). A schematic diagram is shown in Fig. 1. These four domains have been shown to be discrete structural entities and can be produced as individual truncates that still possess biological activity (24). The CCN family is highly conserved at the primary structure level with ~30–50% amino acid identity (and 40–60% similarity) (28, 29). CCN5 is an exception in that it
lacks the CT domain but still maintains the first three highly conserved domains, and CCN6 lacks four normally conserved cysteine residues in the VWC domain. Protease-susceptible linkers separate the domains with the N- and C-terminal halves separated by a longer linker region that varies in both length and composition between the proteins. These linker regions can be cleaved by a range of matrix metalloproteases (MMP-1, MMP-2, MMP-3, MMP-7, MMP-9, and MMP-13) (30) and proteases such as elastase and plasmin (30–32). Currently there is no structural information known about the CCN proteins. The high number of cysteines in each protein (~10% by mass) makes obtaining pure material difficult (12), and previous attempts to purify these proteins have resulted in insoluble aggregates (33, 34). There has been some progress, with microgram quantities of active CCN2 produced in bacteria (35). Production of individual domains or truncated sections has been more successful (35–38), although it has not led to a greater understanding of the structure of the CCN proteins or their individual domains.

Although the structure of the individual domains can be hypothesized through homology modeling of related ECM domains (1), it does not yield any clues as to the overall shape of the molecule and any interdomain interactions that may occur. It has been hypothesized that the proteins exist as a globular bundle until positive or negative regulators act upon them to unfold and expose their active domains or cleave them at their protease-susceptible linkers into biologically active truncated fragments (9). Given the high degree of sequence identity between the proteins, it is likely that domain-domain interactions may be one of the factors that control the subtle differences in behavior and binding of these proteins (1, 2). In the case of CCN2, domain cooperativity has already been seen to influence ligand binding with domain 4 binding to VEGF_{121} and both domain 3 and domain 4 being required to bind the longer VEGF_{165} (20). Experiments done by Kubota et al. (24) also demonstrated that certain biological effects were only seen when multiple domains were involved. Movement about linker regions has been seen to be important in other mosaic proteins (39), and the secondary structure, if any, of the linker regions can also have an effect on interdomain communication (40).

To begin to understand how structure can lead to function, we have expressed and purified CCN3 and CCN5 in high yields. We have elucidated their low resolution solution structures using small angle scattering to give us a first look at the hitherto unseen CCN protein family members.

---

**FIGURE 1.** A schematic diagram of the CCN family. A, the six members of the CCN family illustrating the four conserved domains. The conserved cysteine bonds are also shown in each domain with CCN6 lacking two of these in the VWC domain. CCN5 can also be seen to be lacking the CT domain (6). B, a schematic of the recombinant (r) constructs of CCN3 and CCN5 with an N-terminal Trx tag, an N-terminal hexa-histidine affinity tag, and a factor Xa cleavage site.

**FIGURE 2.** Circular dichroism data. CD spectra of CCN3 (gray triangles) and CCN5 (black squares) are shown. CD spectra were collected on a Jasco-J600 between 190 and 290 nm. Analysis in the program JFIT suggested that ~50% of the protein was composed of β-strands, ~40% consisted of random coil, and ~10% consisted of α-helix. The large proportion of β-strand and coil is consistent with the homology modeling with the small proportion of α-helix likely from the Trx tag.
EXPERIMENTAL PROCEDURES

Cloning—The cDNA clones of CCN3 (IRATp970B0612D) and CCN5 (IRATp970H0284D) were obtained from imaGenes and cloned into the pET32-Xa/LIC vector at the ligation independent cloning site (Merck Biosciences), excluding the predicted signal sequence (residues 1–31 and 1–23, respectively). The pET32-rCCN vector was then transformed into Rosetta-Gami2 cells (Merck Biosciences) for expression.

Expression and Purification—Single colonies were used to inoculate 10 ml of LB supplemented with 100 μg/ml ampicillin and grown overnight at 37 °C. This was then used to inoculate 1 liter of LB supplemented with 100 μg/ml ampicillin and 35

Figure 3. Scattering curves for CCN3 and CCN5. Top, A, the scattering curve of CCN3 (blue circles). The fit to the ab initio data is shown in red and was calculated using the program DAMMIN (41). B, the radius of gyration ($R_g$) was calculated using the Guinier approximation (61) and was found to be 57.7 Å. C, the interatomic distance function ($P(r)$) gave a maximum dimension of 170 Å. These values were calculated in GNOM (44). Bottom, A, the scattering curve of CCN5 (blue circles). The fit to the ab initio data is shown in red and was calculated using the program DAMMIN (41). B, the radius of gyration ($R_g$) was calculated using the Guinier approximation (61) and was found to be 48.8 Å. C, the interatomic distance function ($P(r)$) gave a maximum dimension of 160 Å. These values were calculated in GNOM (44).
μg/ml chloramphenicol and was grown for ~8 h at 37 °C before expression was induced with 0.2 mm isopropyl-1-thio-β-D-galactopyranoside and left for ~16 h at 20 °C before cells were harvested through centrifugation. Cell pellets were frozen overnight at ~20 °C before being resuspended in lysis buffer (25 mM Hepes, pH 7.3, 500 mM NaCl, 1 mM PMSF, EDTA-free protease inhibitor tablet (Roche Applied Science); 10% v/v Bugbuster; Benzonase; and lysozyme (Merck Biosciences) and incubated for ~20 min at room temperature before being lysed using a French press. Lysed cells were spun at 24,000 rpm for ~30 min, and the inclusion bodies were collected. Inclusion bodies were resuspended in ~20 ml solubilization buffer (25 mM Hepes, pH 7.3, 500 mM NaCl, 1 mM PMSF, 7.2 mM urea) before being applied to a 1-ml His-Trap crude column (GE Healthcare) equilibrated in 25 mM Hepes, pH 7.3, 500 mM NaCl, 20 mM imidazole, and 7.2 mM urea. The CCN proteins were eluted in a single step with elution buffer (as above, except 0.5 mM imidazole). The partially purified CCN proteins were dripped slowly into ~100 volumes of refolding buffer (50 mM Tris, pH 8.2, 250 mM NaCl, 10 mM KCl, 0.05% PEG3350) and left at 4 °C for >24 h. The refolded CCN proteins were then subjected to a two-step purification with an AKTA Express (GE Healthcare) that involved an Ni2+ affinity column equilibrated with refolding buffer and eluted with an ascending imidazole gradient (0–500 mM) followed by gel filtration on a Superdex 75 column (GE Healthcare) equilibrated in 25 mM Hepes, pH 8.2, 200 mM NaCl. Proteins were then concentrated for use in biochemical studies. The concentration was calculated using absorbance at 280 nm. Once concentrated, the purity was checked by SDS-PAGE and Western blot using anti-His antibody targeting the N-terminal fused His tag. Samples were stored at 4 °C.

Circular Dichroism (CD)—CD data were recorded on a Jasco J600 spectropolarimeter at 25 °C from 190 to 290 nm with a 50 nm/min scanning speed using quartz cuvette with 5-mm path length. For both samples, eight runs were performed and averaged. A buffer blank (50 mM Tris, pH 8.2, 200 mM NaCl) was subtracted. An estimation of the secondary structure was performed by the program JFIT (41). CD data of CCN3 and CCN5 can be seen in Fig. 2. CD data of CCN3 and CCN5 are shown in Fig. 2. The small proportion of helical content (~10%) is likely due to the presence of the attached Trx tag that is composed primarily of α-helices (Protein Data Bank (PDB) ID 1KEB) (48). This high ratio of β-strands and random coil is in agreement with homology modeling (1) and secondary structure prediction software. These suggested that domains 1–3 are composed entirely of β-strands and disulfide-stabilized random coil and that only the CT domain may contain a small proportion of α-helix. For CCN3 that lacks the CT domain, all of the observed helical content is likely due to the α-helices present in the Trx tag.

There have been suggestions that the CCN proteins may exist as dimers or higher order oligomers in solution due to the VWC and CT domains (10). The calculated mass (calculated from the distance distribution function (P(r) curve) gave a mass of 50.2 kDa for CCN3 and 38.1 kDa for CCN5, which is close to the calculated masses of the fusion proteins of 51.2 and 39.9 kDa, respectively. The monomeric state was also supported by
Figure 4. Model fitting to small angle x-ray scattering data. A, the ab initio models generated by DAMMIN (45) for CCN3 and CCN5. In both cases, they are long elongated molecules that are extended and may act as a scaffold for binding and coordinating ligand partners. B, the homology models generated of the individual domains of the CCN proteins modeled into the surface envelope of the small angle x-ray scattering model. The domains are colored with the Trx tag in black, the IGFBP domain in red, the VWC domain in blue, the TSP domain in yellow, and the CT domain in green. CCN5 lacks the CT domain. The figures were generated using PyMOL (62).
the monodisperse nature of the peaks seen in the dynamic light scattering experiments, which were both >99% pure by dynamic light scattering analysis.

**Ab Initio Modeling**—Given the complete absence of structural data available for the CCN molecules, *ab initio* modeling was performed using the program DAMMIN (45) with 20 runs of DAMMIN (45) being averaged using the program DAMAVER (46). The resultant surface envelope models suggested that both proteins exist as extended, not globular, molecules in solution (Fig. 4A), differing from the globular hypothesis proposed by Perbal (9). This long extended flexible structure, although not immediately obvious as a stable protein, is in keeping with the known structural information about many of the other ECM multidomain proteins that exhibit similar long extended highly flexible structures (25). Fibronecetin forms a long extended formation of small repeats (49), as does fibrillin (50), and all the current data on the related ECM modulator thrombospondin (51) also support the idea of a long extended molecule to allow for easy docking of ligands or multiple domains to interact within the same ligand simultaneously. The long linker region separating the N- and C-terminal halves of the molecule allows for a great deal of flexibility and could accommodate large domain movements for multiple domains to act in synergy with the binding ligands, as seen for domains 3 and 4 of CCN2, to simultaneously bind to VEGF$_{165}$ as reported by Inoki et al. (20) or by Kubota et al. (24), using the four independent domains of CCN2. In addition to allowing flexible binding of multiple domains to ligands, the extended molecule allows access to the protease-susceptible sites that link the domains and may be why biologically active truncates are so prevalent among the CCN family (52, 53). These proteases include several matrix metalloproteases (MMP-1, MMP-2, MMP-3, MMP-7, MMP-9, and MMP-13) that target the large fibrillin (50), and all the current data on the related ECM model thrombospondin (56), and the structure of the TSP domain from malaria TRAP protein (PDB 2BBX) was used for both CCN3 and CCN5 (57). Finally, the CT domain has a cysteine knot that is similar to that found in the growth factors such as VEGF, TGF, and BMPs (58). The CT domain in CCNs may have some structural differences outside of the cysteine knot as only a partial model of the CCN3 domain could be built using the structure of TGFβ-1 (PDB 1KLA) (59) as a template (1, 2).

With only partial homology models and an absence of structural data, we were unable to employ rigid body modeling programs such as SASREF (60); however, using molecular graphics programs, we were able to manually dock the modeled domains (1, 2) into the surface envelope produced by DAMMIN (45). In both cases, the *ab initio* surface envelope was able to comfortably accommodate the Trx tag and the functional domains for CCN3 and CCN5, aligning each domain so that the N and C termini were in the approximately correct orientation to link to the next domain. The available space between the domains also gave a large enough degree of freedom to comfortably accommodate the unstructured linker regions between each domain.

Given the high degree of homology (6) between CCN1–4 and CCN6, it is likely that this long extended flexible structure with each domain readily accessible for binding will hold true for every member of the CCN family. Although CCN5 lacks the CT domain that is responsible for many of the biological functions in CCN1–3 (9, 10, 12), CCN5 is still biologically active (13) and seems to share the same extended structure, suggesting that it is the interaction with other domains that is responsible for its biological activity and that subtle changes in the structure could be responsible for the observed activity.

**Conclusion**—Small angle x-ray scattering data have now given us our first insight into the structure of the full-length CCN proteins. It can now be hypothesized that they exist in the ECM as long extended scaffolds able to move and flex to allow multiple domains to bind a ligand, these domains working in concert to pass on messages throughout the cell. This first glimpse of the CCN structure when combined with higher resolution structural techniques may soon begin to answer some of the important questions about the molecular recognition of CCN proteins through their ligands. In addition, with further scattering experiments on CCN-ligand complexes, we may be able to observe the flexible nature of CCN binding.

**REFERENCES**

1. Holbourn, K. P., Acharya, K. R., and Perbal, B. (2008) *Trends Biochem. Sci.* 33, 461–473
2. Holbourn, K. P., Perbal, B., and Acharya, K. R. (2009) *J. Cell Commun. Signal.* 3, 25–41
3. O’Brien, T. P., Yang, G. P., Sanders, L., and Lau, L. F. (1990) *Mol. Cell. Biol.* 10, 3569–3577
4. Bradham, D. M., Igarashi, A., Potter, R. L., and Grotendorst, G. R. (1991) *J. Cell Biol.* 114, 1285–1294
5. Joliot, V., Martinerie, C., Dambrine, G., Plassiart, G., Briscac, M., Crochet, J., and Perbal, B. (1992) *Mol. Cell. Biol.* 12, 10–21
6. Bork, P. (1993) *FEBS Lett.* 327, 125–130
7. Pennica, D., Swanson, T. A., Welsh, J. W., Roy, M. A., Lawrence, D. A., Lee, J., Brush, J., Taneyhill, L. A., Deuel, B., Lew, M., Watanabe, C., Cohen, R. L., Melhem, M. F., Finley, G. G., Quirke, P., Goddard, A. D., Hillan, K. J., Gurney, A. L., Botstein, D., and Levine, A. J. (1998) *Proc. Natl. Acad. Sci.*
Solution Structure of CCN3 and CCN5 Proteins

Shimo, T., Nakanishi, T., Nishida, T., Asano, M., Kanyama, M., Kuboki, T., Tamatani, T., Tsuchiya, K., Takemura, M., Matsumura, T., and Takigawa, M. (1999) J. Biochem. 126, 137–145

Gygi, D., Zumstein, P., Grossenbacher, D., Altwegg, L., Lüscher, T. F., and Gehring, H. (2003) Biochem. Biophys. Res. Commun. 311, 685–690

Thibaut, H., Martineau, C., Crémignon, C., Godeau, F., Boudou, P., Le Bouc, Y., and Laurent, M. (2003) J. Clin. Endocrinol. Metab. 88, 327–336

Ball, D. K., Rachfal, A. W., Kemper, S. A., and Brigstock, D. R. (2003) J. Endocrinol. 176, R1–R7

Tong, Z. Y., and Brigstock, D. R. (2006) J. Endocrinol. 188, R1–R8

Dobson, C. M. (1990) Nature 348, 198–199

Arai, R., Wrigners, W., Nishikawa, Y., Nagamune, T., and Fujisawa, T. (2004) Proteins 57, 829–838

Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (1997) J. Pept. Res. 50, 73–75

Huang, T. C., Toraya, H., Blandon, T. N., and Wu, Y. (1993) J. Appl. Crystallogr. 26, 180–184

Konarev, P. V., Volkov, V. V., Koch, M. H. J., and Svergun, D. I. (2003) J. Appl. Crystallogr. 36, 1277–1282

Svergun, D. I. (1991) J. Appl. Crystallogr. 24, 485–492

Svergun, D. I. (1999) Biophys. J. 76, 2879–2886

Volkov, V. V., and Svergun, D. I. (2003) J. Appl. Crystallogr. 36, 860–864

Holbourn, K. P., and Acharya, K. R. (2011) Biochem. Biophys. Res. Commun. 407, 837–841

Rudresh, Jain, R., Dani, V., Mitra, A., Srivastava, S., Sarma, S. P., Varadarajan, R., and Ramakumar, S. (2002) Protein Eng. 15, 627–633

Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) Cell 84, 155–164

Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (1996) Cell 85, 597–605

Carlson, C. B., Lawler, J., and Mosher, D. F. (2008) Cell Mol. Life Sci. 65, 672–686

Ayler-Lelievre, C., Brigstock, D., Lau, L., Pennica, D., Perbal, B., and Yeger, H. (2001) Mol. Pathol. 54, 105–120

Perbal, B. (2004) Lanced 363, 62–64

Sit, T., Popowicz, G. M., Siwanowicz, I., Huber, R., and Holak, T. A. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 13028–13033

O’Leary, J. M., Hamilton, J. M., Deane, C. M., Valeley, N. V., Sandell, L. J., and Downing, A. K. (2004) J. Biol. Chem. 279, 53857–53866

Tan, K., Duquette, M., Liu, J. H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J., and Wang, J. H. (2002) J. Cell Biol. 159, 373–382

Tossavainen, H., Pihlajamaa, T., Huttunen, T. K., Ruol, E., Rauvala, H., Permi, P., and Kilpelaïinen, I. (2006) Protein Sci. 15, 1760–1768

McDonald, N. Q., and Hendrickson, W. A. (1993) Cell 73, 421–424

Hinck, A. P., Archer, S. J., Qian, S. W., Roberts, A. B., Sporn, M. B., Weatherson, J. A., and Downing, A. K. (1999) J. Clin. Endocrinol. Metab. 84, 1237–1250

Petoukhov, M. V., and Svergun, D. I. (2005) Biophys. J. 89, 1237–1250

Guinier, A. (1939) Ann. Phys. 12, 161–237

Delano, W. L. (2010) The PyMOL Molecular Graphics System, version 1.3r1, Schrödinger, LLC, New York