The minicollagens found in the nematocysts of Hydra constitute a family of invertebrate collagens with unusual properties. They share a common modular architecture with a central collagen sequence ranging from 14 to 16 Gly-X-Y repeats flanked by polyproline/hydroxyproline stretches and short terminal domains that show a conserved cysteine pattern (CXXXCCXXX-CXXX). The minicollagen cysteine-rich domains are believed to function in a switch of the disulfide connectivity from intra- to intermolecular bonds during maturation of the capsule wall. The solution structure of the C-terminal fragment including a minicollagen cysteine-rich domain of minicollagen-1 was determined in two independent groups by 1H NMR. The corresponding peptide comprising the last 24 residues of the molecule was produced synthetically and refolded by oxidation under low protein concentrations. Both presented structures are identical in their fold and disulfide connectivity revealing a robust structural motif that is supposed to serve as the polymerization module of the nematocyst capsule.

Minicollagens of nematocysts in Hydra, corals, and other cnidaria are very unusual proteins with structural domains not shared by other invertebrate or vertebrate collagens. From both the N and C terminus of the collagen triple helix emerge three polyproline-II-type helices, which consist of 5–23 proline or hydroxyproline residues (1–3). Each of the polyproline-II-type helices is terminated by a small Cys-rich domain termed minicollagen Cys-rich domain (MCRD). The N- and C-terminal MCRDs are homologous and share the cysteine pattern CXXXCCXXX-CXXX. A small propeptide region preceding the N-terminal MCRD is cleaved off during expression, and mature minicollagen has a rather symmetrical appearance with closely similar structural elements at both sides. This bipolar nature of minicollagen is unique among all other known collagens (4–7) and suggests a special function.

A unique function of minicollagen is also suggested by its restricted appearance in the capsule wall of nematocysts. Nematocysts are complex explosive organelles, which basically consist of a capsule, an inverted tubule armed with spines, and an operculum. The tubule is connected to the capsule wall and is twisted in many turns inside the osmotically charged capsule matrix. Following stimulation, the internal tube is expelled, the osmotic pressure is released, and the capsule contents including toxins are released at the tubule end. This specialized form of exocytosis proceeds with ultrafast rates and accelerations comparable to those of a fired bullet (8).

The capsule wall resists more than 150 atmospheres of osmotic pressure in the charged state. For Hydra nematocysts it consists mainly of two proteins, minicollagen and nematocyst outer wall antigen (NOWA) (9–11). The two proteins can be dissolved from capsule preparations only under reducing conditions. Already at a time when the amino acid sequence of minicollagens was unknown it was found that these collagens formed disulfide cross-linked polymers that were insoluble in SDS but easily soluble in the presence of a reducing agent (12, 13). Following the discovery of a family of minicollagens (3, 4) and recombinant expression of minicollagen-1 (9) it was found that the proteins are expressed in a soluble precursor form present in the endoplasmic reticulum and post-Golgi vacuoles in Hydra. They are converted to the disulfide-linked assembly form of the nematocyst wall upon wall compaction, during which a dense and well defined capsule wall is formed (9). The morphological changes, a loss of accessibility to antibodies against minicollagen-1, and a parallel loss of solu-
bility under non-reducing conditions suggested a close link between disulfide polymerization and the condensation of wall proteins. Both processes provide an explanation for the unusually high tensile strength of the mature nematocyst wall.

Minicollagen-1 of *Hydra* recombinantly expressed in mammalian cells contains internal disulfide bonds in its MCRDs but no interchain disulfide cross-links between chains (9). The trimeric collagen molecules dissociated into single chains when heated to 45 °C under non-reducing conditions (10). Reduction did not influence this transition temperature, indicating that only the collagen domain is responsible for trimerization. Recombinant minicollagen-1 was found to form aggregates in electron micrographs but no disulfide bridges were formed spontaneously under *in vitro* conditions (9). This material behaved like the precursor form *in vivo* and could be solubilized by SDS or other denaturants under non-reducing conditions. It was therefore concluded that disulfide isomerases or other external parameters are required for a disulfide reshuffling process by which internal disulfide bridges are converted to intermolecular links (14).

More recent results suggest that the structural function of minicollagens in wall hardening is complemented by NOWA (11). Altogether 10 domains were identified in this protein, namely a SCP-domain (see Smart data base, smart.embl-heidelberg.de, smart00198), a C-type lectin domain (CTLD, smart00034), and eight C-terminal domains with homology to the MCRDs. In particular, the pattern of six cysteines is shared by the corresponding domains in NOWA and minicollagen. The presence of homologous Cys-rich domains in both proteins suggested a joint function in disulfide-mediated polymerization. Supportive evidence was obtained by the analysis of breakdown products of native nematocyst capsules after limited sonication without reduction.2 Minicollagen and NOWA formed disulfide cross-linked units that, like the entire capsule, were readily dissolved under even mild reducing conditions.

Disulfide reshuffling processes and the formation of disulfide-linked complexes are common mechanisms in the extracellular space, and some recently explored systems may be referenced (15–19). The minicollagen/NOWA system may stand as a promising or stable but only as part of a larger structure. The fmcoc/tBu strategy, used by Barth, A. G. Milbradt, L. Moroder, J. Engel, and T. W. Holstein, manuscript in preparation.

---

**EXPERIMENTAL PROCEDURES**

**Peptide Synthesis and Oxidation**

Preparation 1—The 6xStBu protected peptide Ac-PCPVFCVA-QCVPCTPFCYCCPARRK-NH2 was synthesized on Fmoc-Rink-Amide polystyrene glycol-dimethylacrylamide coupling resin by the Fmoc/tBu method using double couplings with Fmoc-Xaa-OH/HBTU/HOBt/DIEA (4:4:4:8), intermediate Fmoc cleavage with 20% piperidine in N,N-dimethylformamide, and acetic anhydride/DIEA (4:8) for N-terminal acetylation. After resin cleavage/deprotection with trifluoroacetic acid/phenol/H2O/thioanisole/1,2-ethanedithiol (82.5:5:5:5:2.5) the product was isolated by reverse phase-HPLC: yield 6%; HPLC: tR = 12.5 min (>98%); electrospray ionization-mass spectrometry: m/z = 1055.0 [M+3H]+; 1581.8 [M+2H]+; M+ = 2632.31 calculated for C136H145N23O10S6.

Preparation 2—The linear peptide was synthesized with an ABI 433A synthesizer. Couplings were carried out on a Fmoc-Lys(Boc)-polyethylene glycol-polyamide resin (Perspective Biosystems, 0.21 mmol/g) using Fmoc amino acids (Fmoc-Arg(Pbf), Fmoc-Lys(Boc), Fmoc-Tyr(tBu), Fmoc-Cys(trityl), and Fmoc-Gln(trityl)) (Anaspec). After resin cleavage/deprotection with trifluoroacetic acid containing thioanisole (5% v/v), ethanedithiol (1% v/v), and anisole (2% v/v). After precipitation in diethylether, the crude peptide was purified by preparative HPLC (Vydac® C18, 10–15 μm), 300 Å, 250 × 50 mm, W.R. Grace, Columbia, MD). The peptide was characterized by electrospray ionization/quadrapole/time-of-flight mass spectrometry and amino acid analysis. A mass of 2590.9 Da was determined, and the expected amino acid composition was found.

The lyophilized purified peptide was dissolved in 25 mM Tris/HCl, pH 6.0, containing 100 mM NaCl at a concentration of 0.02 mM. Folding was allowed to proceed under N2 for 72–96 h at 4 °C. Oxidation was induced by changing the pH to 8.0–8.1 with saturated Tris and the addition of reduced and oxidized glutathione (10:1 molar ratio) to a final concentration of 1 mM. The solution was exposed to air at 4 °C for 4–10 days. Oxidation was stopped by the addition of trifluoroacetic acid to give a pH of 1.3. The oxidized peptide was purified by preparative HPLC. The yield of oxidized peptide was 103 mg (16% of theoretical yield).

Mass spectroscopy of the oxidized peptide showed a molecular mass of 2585.1 Da.

**NMR Data Collection**

Structure 1—NMR experiments for conformational analysis were carried out on 283 K on Bruker DRX500, DMX750, and IX900 spectrometers using a 3 mM sample of the peptide dissolved in a H2O/D2O (9:1) mixture at pH 3.5. Resonance assignments were performed according to the method of Wuthrich (38). 173 experimental interproton distance constraints were extracted from two-dimensional-nuclear Overhauser effect spectroscopy (39) experiments with mixing times between 75 and 200 ms. Five hydrogen bonds were identified from temperature shifts and 2H/H exchange. Acceptor carbonyl groups were identified in initial structure calculations. The nuclear Overhauser effect intensities were converted into interproton distance constraints using the classifications of very strong, 1.7–2.3 Å; strong, 2.2–2.8 Å; medium, 2.6–3.4 Å; weak, 3.0–4.0 Å; very weak, 3.2–4.8 Å; and the distances of pseudo atoms were corrected as described by Wuthrich (38). Distance geometry and molecular dynamics-simulated annealing calculations were performed with the INSIGHTII 98.0 software package (Accelrys, San Diego, CA) on Silicon Graphics O2, R5600 computers (SOF, Mountain View, CA) as described recently (20). In brief, one hundred structures were generated by distance geometry and refined with molecular dynamics-simulated annealing steps. The experimental constraints were applied at every stage of the calculations. The coordinates and structural restraints have been deposited in the Brookhaven Protein Data Bank under accession number 1SP7.

**Structure 2**—The chemically synthesized C-terminal MCRD of minicollagen-1 was purified after oxidation according to purification procedure 2. The determination of the NMR structure (PDB accession number 1SP7) was carried out in 5 mM sodium phosphate buffer, pH 6.5, at 15 °C by using homonuclear and heteronuclear techniques and inform.
Fig. 1A, minicollagen-1 amino acid sequence and domain structure. The propeptide sequence is blue, the MCRDs are red, proline-rich sequences are light green, the collagen repeat is dark green. B, alignment of MCRDs in minicollagen molecules from different cnidarians and in NOWA. MCol1h, minicollagen-1 Hydra; MCol2h, minicollagen-2 Hydra (3); MColad, minicollagen Acropora donei (2); MColac, minicollagen Acropora cervicornis (1); MColap, minicollagen Acropora palmarae (1); N- and C- terminal. The sequence of the Cys-rich region of NOWA in Hydra (NWa) starts with repeat 1 and terminates with repeat 8. Residues in MCol1hC are numbered starting at the proline preceding the first cysteine, and the same numbering was used in the NMR structures. The highly conserved cysteine residues are marked in red. Proline in position 12, which is conserved with two exceptions, is marked in purple. The sequence of the synthesized and investigated peptide is underlined.

**RESULTS**

The MCRD Constitutes a Conserved Sequence Module in Nematocyst Minicollagens and in NOWA—The sequence of minicollagen-1 with domain indications and the alignment of cysteine-rich domains from different minicollagens and NOWA are represented in Fig. 1. Minicollagen-1 consists of an N-terminal MCRD, N-terminal proline region, a central collagen sequence, and a shorter C-terminal proline polyproline stretch followed by a C-terminal MCRD (Fig. 1A). The preceding propeptide is cleaved off during the recombinant expression of minicollagen-1 and probably also in Hydra (9). As already mentioned in the Introduction the overall sequence homology of the Cys-rich domains is not very high, but the cysteine pattern is identical for all minicollagens and for the eight C-terminal domains of NOWA with the only variation being in the number of residues spacing the first two Cys residues. The sequence of the C-terminal minicollagen-1 MCRD, which has been investigated in the present study, is underlined and the numbering of residues corresponds to the synthetic peptides used in this study (Fig. 1B). Beside the cysteines there is only one conserved residue, which is Pro12 (Fig. 1B, shown in purple).

Peptide Synthesis and Oxidation—As the MCRD occurs in different molecular contexts, at the N- and C-terminal extensions of minicollagen molecules as well as eight times repeated at the C terminus of NOWA, we speculated that it might constitute an isolated domain with the capacity of independent folding. Peptide synthesis was carried out for the C-terminal MCRD of minicollagen-1 starting with the last proline residue of the C-terminal polyproline stretch and including the charged C terminus of the full-length molecule. The formation of disulfide bonds occurs in the presence of a redox buffer at 100 μM peptide concentrations to avoid aggregation by intermolecular disulfide bonds (see “Experimental Procedures”). The final product showed a single peak in mass spectroscopic analysis with the reduced and oxidized MCRDs having a different in molecular weight of 6 Da, thereby strongly indicating the formation of three intramolecular disulfide bonds (Table I). Disulfide bonds can be shown to be all intramolecular in mass spectroscopic analysis. Analytical ultracentrifugation confirmed the absence of significant aggregation or multimerization in solutions from 0.2 to 0.8 mM total peptide concentration (Table I), which was further supported by 1H NMR T2 relaxation times of more than 100 ms at 25 °C indicative of the prevalence of a monomeric state in solutions of 1.6 mM MCRD.

The NMR Structure of the MCRD—The solution structure of the MCRD was determined independently from two separate peptide preparations (see “Experimental Procedures”). Both structures show an identical tightly packed globular fold (Figs. 2 and 3), which consists of a short N-terminal α-helix between Val5 and Gln9 followed by an inverse γ-turn (Gln9-Val11), a type I β-turn (Val11-Cys14), and a type III β-turn (Pro15-Cys18). Thus cysteines 6, 10, 14, and 18 are directly located in turns, whereas cysteine 2 is located in a proline-rich N-terminal sequence, and cysteine 19 is oriented presumably by the turn 19 and a C-terminal proline. All proline residues in the MCRD are in trans conformation as evidenced by specific nuclear Overhauser effects.

The only conserved residue Pro12 (Fig. 1B) imposes a β-I turn topology on residues 11–14 because of its fixed φ angle of −60°. Hydrogen bonds are established in the turns between Val5 (O) and Gln9 (H5), Gln9 (O) and Val13 (H5), Val13 (O) and Cys14 (H9251) and Gln9 (O) and Val13 (H5), Val13 (O) and Cys14 (H9251).
TABLE I
Molar masses of the C-terminal MCRD peptide of minicollagen-1

| Molar mass | UC, analytical ultracentrifugation; MS, mass spectroscopy. |
|------------|----------------------------------------------------------|
|            | Molar mass | Molar mass |
|            | g/mol      | g/mol      | g/mol      |
| MColl1hC reduced | 2591    | 2592      |            |
| MColl1hC oxidized | 2700 ± 300* | 2585      | 2586      |

* Average of five determinations in the concentration range of 0.2–0.8 mg/ml; S.D. is indicated.

**Structure of Minicollagen Cys-rich Domain**

The backbone is displayed in blue and side chains in red except for the cysteine side chains in yellow. Structure 1 (left, PDB accession number 1SOH), the 10 lowest energy structures are superimposed using the backbone atoms of residues 2–20 with a root mean square distance of 0.6 Å. The 20 lowest energy conformers of the minicollagen-1 C-terminal MCRD of 100 structures calculated in CNS (42), with best fit for backbone heavy atoms of residues 2–21 in gray and the C-terminal sequence K22R23K24 (red arrows) is disordered at pH 6.5 in the absence of additional salt.

Fig. 2. NMR structures of the C-terminal MCRD of minicollagen-1. The backbone is displayed in blue and side chains in red except for the cysteine side chains in yellow. Structure 1 (left, PDB accession number 1SOH), the 10 lowest energy structures are superimposed using the backbone atoms of residues 2–20 with a root mean square distance of 0.6 Å. Disordered residues 21–24 are depicted in gray. Structure 2 (right, PDB accession number 1SP7), all heavy atom representation of the 20 lowest energy conformers of the minicollagen-1 C-terminal MCRD of 100 structures calculated in CNS (42), with best fit for backbone heavy atoms of residues 2–21. The charged C-terminal sequence K22R23K24 (gray) is disordered at pH 6.5 in the absence of additional salt.

The first disulfide bond, Cys2-Cys18, clamps the N and C termini of the domain, whereas the Cys6-Cys14 bond connects the N-terminal domain. The connections between the polyproline-II helices and the MCRDs might alter the accessibility of particular disulfide bridges in the MCRD.

DISCUSSION

Small Cys-rich domains are widely distributed building blocks of extracellular proteins in essentially all phyla including plants and bacteria. The most abundant domain type is the epidermal growth factor domain (EGF, smat00181). Including its variants EGFca (smart00179) and EGF-like (smart0001) several thousands of different EGF-domains are known in proteins of different functions. In most cases EGF-domains are arranged in arrays with other domains. Laminin (20) and fibrillin (21) are two of very many examples. Many Cys-rich domains, however, also exist as single autonomous proteins, and the epidermal growth factor domain is a well known example. This fact provides a bridge to the numerous low molar mass Cys-rich proteins, which are also products of larger precursor forms but express their function as toxins, antimicrobial peptides, or other small bioactive agents. Variations of sequences, cysteine patterns, and three-dimensional structures are, however, very large for this diverse class of small proteins, and clear homologies exist for small groups only (22). Small Cys-rich peptides with antimicrobial or toxic functions are part of the innate immunity and defense system of invertebrate animals. They are often stored in secretory granules and released in response to parasites via exocytosis. Nematocyst discharge in cnidarians represents a specialized form of exocytosis from a giant post-Golgi vesicle. The appearance of a structural motif related to Cys-rich peptide tox-
ins in proteins involved in the formation of a wall polymer, which is associated with the nematocyst membrane, might hint at a phylogenetic link between this group of defensive molecules and the evolution of the nematocyst.

The common feature of small Cys-rich domains is the prevalence of disulfide bridges in structure stabilization. This is already suggested by the large fraction of cysteines, which is 4/22 for gomesin (23), about 6/40 for EGF domains, 8/42 for crambin (24) and hellethionin (25) as compared with 6/20 for the Cys-rich domain of minicollagen-1. The later domain has one of the highest ratios of cysteines to total residues known so far. Comparably dense cysteine patterns are only described for small conotoxin peptides like PnIVA and PnIVB (26). The peptide whose structure was elucidated in the present work is four residues longer than the essential core. The structure shows that the first proline and the last three residues are randomly oriented and probably not required for correct oxidative folding. The remaining structure is well defined, and data demonstrate formation of a single topoisomer under optimized oxidative conditions.

In the reduced state no conformational preferences were found in the MCRD by NMR. Similar observations were reported for many other Cys-rich domains underlining the importance of disulfide bonds for stabilization. As shown for tachyplesin mutational replacements of Cys residues by alanine lead to a loss of structure (27). Interestingly the global fold of tachyplesin was rescued by hydrophobic interactions between pairs of tyrosines or phenylalanines when cysteines were replaced by these residues (27).

Although the equilibrium structure of MCRD is predomi-
nantely determined by the covalent interactions between Cys residues, kinetic intermediates must be responsible for correct disulfide pairings. A folded-precursor mechanism (28, 29, 36) and a quasi-stochastic mechanism (30, 31) were proposed for other monomeric proteins. True pathways are probably between these two extremes. For the Cys-rich domain of minicollagen formation of a type I β-turn by Pro15 and hydrogen bonding between Cys14 (Hβ) and Val11 (O) is most likely a very important intermediate step. As can be seen from the structure (Figs. 2 and 3) this bend and the γ-turn bring the cysteines into close vicinity for the correct intramolecular disulfide bridging.

It is amazing that few non-covalent interactions in the MCRD of minicollagen and in many other short peptides lead to correct connectivities between several Cys residues, which statistically could also interact in very many different intra- and intermolecular modes. To establish the disulfide connections of MCRD, pairwise bridges between adjacent Cys-residues have to be prevented. The three intervening amino acid residues found in MCRDs render pairwise bridges unfavorable (32), and this feature is most likely essential for proper folding.

In small disulfide-rich proteins identical folds are found in homologous domains in contrast to large variations in non-cysteine residues. Striking examples are the -200 solved structures of different EGF domains, which exhibit the same fold, although sequence identities of non-cysteine residues are difficult to detect (Smart data base, key words EGF and structure).

This high conservation of structure leads us to assume that the related Cys-rich repeats in the minicollagen of corals and in NOWA proteins have the same global fold as the MCRD of minicollagen-1 of Hydra, which was investigated in this study. Conservation of residues is low (Fig. 1) and would not allow the prediction of structure identities for conventional globular proteins (33).

Earlier data on the interaction of minicollagens and NOWA in the nematocyst wall conclusively demonstrated a switch from a soluble state of both proteins with all disulfides internally linked to an insoluble dense polymeric state with intermolecular disulfide bonds (see Introduction). Analogous assembly processes involving disulfide reshuffling were reported for virus capsid proteins (19), collagen IV (34), and other systems (17, 18). The only cysteines in minicollagen-1 are located in the MCRDs. Consequently a disulfide reshuffling interaction was proposed between these domains. The data also suggested corresponding interactions between the Cys-rich domains of minicollagen and NOWA. In vivo the situation may be rather complex because of possible interactions between N- and C-terminal domains of different minicollagens in the same organism. At each end of a minicollagen three MCRDs are present (Fig. 4). This opens the possibility of a simultaneous interaction between minicollagens and NOWA.

Extensive searches of the data base failed to reveal MCRD-like domains in non-nematocyst proteins. The domain type seems to be highly specialized for disulfide-mediated cross-linking of the nematocyst wall. For several organisms with nematocyst organelles minicollagen and NOWA genes have been found, but the sequencing of cnidaria genes is still too incomplete to allow a general conclusion. Interestingly in the minicollagen of corals (AF507373.1) Cys-residues frequently occur after glycines in the collaganeous part of the sequence. Cysteine residues, which interlink collagen chains of the same molecule are located at the ends of a triple helix to form a stable disulfide knot (35, 36) or located in interruptions of the regular Gly-X-Y repeat (37). Cys-residues in the X position of collagen are highly unusual and are most likely used to link between collagen molecules. It may be speculated that the nematocyst walls of corals are cross-linked by disulfide bridges between collagen triple helices in addition to those between MCRDs thus providing a higher tensile strength.

Future work will focus on the mechanism of the disulfide reshuffling process using isolated Cys-rich domains of minicollagen and NOWA. A likely possibility is the interaction between N- and C-terminal domains of minicollagens. In vitro, the disulfide exchange reaction probably is catalyzed by disulfide isomerases or other helper proteins, which have to be defined prior to in vitro studies. Our working hypothesis is that only one or two of the three disulfide bonds will be reshuffled from intra- to intermolecular. Likely candidates are the bonds Cys24-Cys18 and Cys11-Cys19, which are both surface-exposed.

Acknowledgment—We thank Eric Steel for expert technical assistance.
40. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J. Biomol. NMR 6, 277–293
41. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) J. Magn. Res. 95, 214–220
42. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, I. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
43. Koradi, R., Billeter, M., and Wüthrich, K. (1996) J. Mol. Graph 14, 29–32, 51–55
44. Meier, S., Häussinger, D., Pukidysheva, E., Bachinger, H. P., and Grzesiek, S. (2004) FEBS Lett., in press
The Structure of the Cys-rich Terminal Domain of *Hydra* Minicollagen, Which Is Involved in Disulfide Networks of the Nematocyst Wall

Elena Pokidysheva, Alexander G. Milbradt, Sebastian Meier, Christian Renner, Daniel Häussinger, Hans Peter Bächinger, Luis Moroder, Stephan Grzesiek, Thomas W. Holstein, Suat Özbek and Jürgen Engel

*J. Biol. Chem.* 2004, 279:30395-30401.
doi: 10.1074/jbc.M403734200 originally published online May 3, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M403734200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 39 references, 14 of which can be accessed free at http://www.jbc.org/content/279/29/30395.full.html#ref-list-1