Continuous Molecular Evolution of Protein-Domain Structures by Single Amino Acid Changes

Sebastian Meier,1,5,* Perrnille R. Jensen,1,6 Charles N. David,2 Jarrod Chapman,3 Thomas W. Holstein,4 Stephan Grzesiek,1 and Suat Özbek4,*

1 Department of Structural Biology
Biozentrum, University of Basel
Klingelbergstrasse 70
CH-4056 Basel
Switzerland
2 Department of Biologie II
Ludwig-Maximilians-Universität
Grosshadenerstrasse 2
D-82152 Planegg-Martinsried
Germany
3 Department of Energy
Joint Genome Institute
Walnut Creek, California 94598
4 Institute of Zoology
Department for Molecular Evolution and Genomics
University of Heidelberg
Im Neuenheimer Feld 230
D-69130 Heidelberg
Germany

*Correspondence: smeier@aki.ku.dk (S.M.), soezbek@zoo.uni-heidelberg.de (S.O.)

Abstract

Protein structures cluster into families of folds that can result from extremely different amino acid sequences [1]. Because the enormous amount of genetic information generates a limited number of protein folds [2], a particular domain structure often assumes numerous functions. How new protein structures and new functions evolve under these limitations remains elusive. Molecular evolution may be driven by the ability of biomacromolecules to adopt multiple conformations as a bridge between different folds [3–6]. This could allow proteins to explore new structures and new tasks while part of the structural ensemble retains the initial conformation and function as a safeguard [7]. Here we show that a global structural switch can arise from single amino acid changes in cysteine-rich domains (CRDs). The short CRDs of little more than 20 amino acids contain a conserved pattern of six cysteines (Figure 1) and form the N- and the C-terminal domain of various minicollagens in Hydra [18, 19]. In addition, they occur in the nematocyst outer-wall antigen (NOWA) as a C-terminal octad repeat [20]. NOWA forms globular aggregates, which probably function as positional organizers of minicollagen assembly [21]. Minicollagens are expressed as soluble precursors with three intramolecular disulfide bonds in their CRDs. During nematocyst maturation, they crosslink with NOWA in an intermolecular disulfide-reshuffling reaction to form the capsule suprastructure [21]. The N- and C-terminal domains of minicollagen 1 have been shown to form different structures with different disulfide bridges [8] and a different overall topology. The significance of cnidarians for early metazoan evolution and the structural variation in closely related domains make the CRDs from Hydra an attractive model for studying the evolution of new protein folds. The solution structure of the first CRD of NOWA (NW1) shares the disulfide pattern and overall structure of the N-terminal CRD of minicollagen 1 (Figure 1). The disulfide pattern (8–20,12–15,16–24) differs completely from the one (8–24,12–20,16–25) previously determined for the minicollagen 1 C-terminal CRD (Mcol1C) [22] despite
a sequence identity of 44% between the first and last cysteine residues of NW1 and Mcol1C. The two sequences need to turn several times to bring the cysteine side chains together. None of the turns, however, coincide between the two structures (Figures 1A and 1B). Furthermore, a conserved proline has cis conformation in the NW1 fold but a trans conformation in the Mcol1C fold. The NW1 structure has an overall left-handed topology, whereas the Mcol1C structure has a right-handed topology, resulting in an entirely different appearance of the two structures.

The small cysteine-rich structures exhibit characteristics that distinguish them from larger folds but turn out to be beneficial for a mutational study of protein-structure conversion. The NW1 and Mcol1C structures are both devoid of a hydrophobic core or salt bridges. Because the cysteine pattern is conserved, different turn propensities of the noncysteine residues must account for the differences in structure between the cysteine-rich domains. Consequently, noncysteine residues in the NW1 and Mcol1C sequences were tested for turn propensities [23] that favor their respective domain structure over the competing structure (Figure 1). This permitted a rational approach to identifying local features, which favor the different domain structures. The strongest effect was predicted for mutations introducing Pro21 (K21P) and Val11 (G11V) into the NW1 structure (Figure 1; also Figure S1 in the Supplemental Data available online).

Wild-type and mutant forms of the NW1 sequence were recombinantly expressed as $^{15}$N-labeled proteins. Purification yielded one major species for the NW1 wild-type (Figure 2A). Amide resonances of the $^{15}$N HSQC were assigned to give a “fingerprint” of the wild-type structure. In contrast to the wild-type, the K21P mutant form yielded two prominent separable species I and II in a ratio of 1:3.5 (Figure 2). Mass spectrometry confirmed that both species have the predicted molecular weight with three intact disulfide bonds (see Supplemental Experimental Procedures). NMR spectroscopy of both species showed that they are folded and structurally stable. Resonance assignments proved that the two K21P species differ significantly, and the chemical shifts in the cysteine-rich core indicate a global structural change (Figure 2B).

The HSQC spectrum of species I closely resembled the spectrum of the wild-type. Accordingly, independent structure determination showed that species I retains the structure of the wild-type NW1 domain (Figure 2C; also Figure S2). As a result of the geometric constraints imposed on the sequence by disulfide bonding, the $\gamma$ turn topology around Cys20 is retained in the K21P mutant, albeit not stabilized by a hydrogen bond. Species II, on the other hand, assumes a fold that closely resembles the Mcol1C structure, with disulfide bonds switched from (8–20,12–25,16–24) to (8–24,12–20,16–25), a right-handed topology instead of a left-handed topology, and a proline switched from cis to trans (Figure 2C; also Figure S2).

Both purified structures of the K21P mutant are stable in the absence of catalysts; that is, they do not reshuffle their intramolecular disulfide bonds. This is most likely due to the enormous activation energy required for unfolding, proline isomerization, and the rearrangement of disulfide bonds. However, the interconversion of domain structures I and II does occur in redox buffer containing reduced and oxidized glutathione (Figure 3). This indicates that conversion occurs via partially or fully reduced states and demonstrates that the different structures are not populated as kinetic traps. Rather, the two domain structures are populated under equilibrium conditions; the ground-state structure is not unique. Despite differences in high-energy structural
features, both structures have less than 0.8 kcal/mol energy difference in the K21P mutant, which thus forms a bridge state between two fundamentally different tertiary structures. The absence of one hydrogen bond in the K21P mutant destabilizes structure I relative to II by $\Delta \Delta G > 2.5$ kcal/mol. This is in full agreement with the energetic contribution of few kcal/mol expected from a partially buried hydrogen bond in proteins [24]. As a consequence, the equilibrium between domain structures I and II, estimated by peak volumes in NMR spectroscopy and analytical HPLC, shifts from $>95\%$ structure I to only $22\%$ structure I in the K21P mutant. Structure II in the K21P mutant can be further stabilized relative to structure I by the introduction of a G11V mutation. This shifts the equilibrium proportion of domain structure II from approximately 78% to approximately 95% (Figure 2A), according to $\Delta \Delta G = 1.0$ kcal/mol to the stabilization of the new structure relative to the previous one. A single G11V mutation in NW1, on the other hand, yields domain structures I and II at a ratio of 70% to 30%, as determined by NMR $^{15}$N HSQC peak intensities (not shown). In conclusion, the main energetic contribution for the structural switch arises from the removal of a single hydrogen bond, supplemented by a smaller contribution from a single intrinsic amino acid positional potential (Figure S1).

A depiction of the structural conversion from wild-type NW1 via the bridge state K21P to the G11V K21P mutant is shown in Figure 4. The two point mutations induce a nearly complete conversion of structure I in the wild-type sequence into structure II. The bridge state structures very closely resemble the natively and artificially evolved domains of the NW1 wild-type and double mutant. The artificially induced structure II is very similar to the naturally evolved Mcol1 C-terminal domain (Figure 4A). An alignment of known minicollagen sequences demonstrates the striking conservation of the residues we have identified as conformational switches between the different domain structures of Hydra CRDs (Figure 4B; Figure S3), indicating that the structural polarity between N- and C-terminal domains of Mcol1 is conserved in minicollagens. This suggests an evolutionary scenario in which a bridge state similar to the K21P mutant has diversified upon gene duplication to form two different domains fixed by their respective disulfide patterns (Figure 4).

A possible evolutionary bridge carrying only one of the two mutations identified here is found in the C-terminal CRD of Hydra minicollagen 7 (Figure 4B; Figure S3). Interestingly, in minicollagen sequences of the more primordial sea anemone Nemastostella vectensis, both mutations can be found separately (Figure S3). Clearly,
these evolved folds retain metastability in accordance with the need to undergo structural changes upon nematocyst wall maturation. The fact that the NOWA octarepeat domain, which spontaneously undergoes disulfide-dependent self assembly [21], has uniform CRD folds (S.M. et al., unpublished data) strongly points to a homophilic polymerization mechanism. Reoxidation of a solution containing two differently folded CRDs (NW1 wt and Mcol1C wt) and subsequent PAGE and mass spectrometric analysis supports the notion of a spontaneous homodimerization between CRDs of only the NW1 wild-type fold. The invention of a second fold to inhibit a premature elongation of the collagen polymer might therefore have facilitated a controlled step during final nematocyst maturation (Figure S4). In addition, a homophilic propagation process is presumed to result in a more extended and flexible network that adapts to the microtubule cage harboring the growing nematocyst vesicle and thus can accommodate the unique mechanical properties of the nematocyst.

Discussion

We demonstrate that two different naturally occurring tertiary structures of cysteine-rich domains from Hydra minicollagens are linked in sequence space by single amino acid changes via a bridge state sequence adopting both native structures. The determinants of the two domain structures are strikingly conserved among cnidarian minicollagen sequences. This argues for an important gene-duplication and -differentiation step in the genetic history of cnidarians for the invention of novel molecular phenotypes, thus pointing to the relevance of the continuous mutational paths retraced here. Although previous experimental proof has been sparse, local conformational variations and secondary-structure fluctuations in a stable tertiary context [7, 12] have pointed to the possibility of global structural switches in proteins. The only known example of a global tertiary-structure switch in a particular protein sequence without change in solvent conditions has been the interconversion of soluble protein and aggregating β-sheet-rich conformation in misfolding diseases. As a result of the dynamic instability of the folds involved, bridge states between different structures in solution have, however, been inherently harder to detect than structural populations that self-replicate and get trapped by aggregation in protein-misfolding diseases. Notably, though, this capability to self-replicate may play a developmental and evolutionary role because the structural diversity of prion proteins has been implicated in molecular memory formation [14].

The use of disulfide-linked domains in our study allowed the purification and characterization of different structural states that would freely interconvert for polypeptide chains without disulfide linkage. In addition, the domains used in this study retain marginal stability to fulfil their biological function. The structural switch in a small disulfide-rich sequence from Hydra is achieved by exclusive tailoring of dihedral angle preferences and hydrogen-bonding properties of the mutated residues, as indicated by the fact that sidechain interactions apart from the disulfide bridges are essentially absent in these small domains. Presumably, the structural evolution of larger protein folds will further depend on a co-evolution of amino acid pairs involved in long-range sidechain interactions [25]. In addition, a larger conformational space may become accessible to marginally stable proteins, which are not densely packed with disulfide bonds. Previously, different protein folds have been predicted to be separated by only a few amino acids or even to overlap in sequence space [15, 26]. Theoretical models of protein structure and evolution thus have questioned the uniqueness of the ground-state structure and have pointed to the relevance of avoiding overly stable “mutational traps” in order to maintain the evolvability of the sequence [15], which is in agreement with our findings. We conclude that theoretical modeling in conjunction with the experiments presented here provides strong support that smooth transitions may be a widespread feature in the evolution of naturally occurring protein folds.

Supplemental Data

Supplemental Data include the Experimental Procedures, four figures, and three tables and can be found online at http://www.current-biology.com/cgi/content/full/17/2/173/DC1/.

Acknowledgments

We thank M. Rogowski for the recording of mass spectra. We thank J. Engel, J. Stetefeld, D. Häussinger, and M. Allan for carefully reading the manuscript and for helpful discussions. This research was

Figure 3. Conversion of Domain Structure I into Equilibrium Quantities of I and II
(A) The structural conversion is catalyzed by refolding buffer containing 5 mM oxidized and reduced glutathione at pH 7.0 and 288 K and is monitored by 1H-15N HSQC spectra.
(B) Kinetics of structural conversion. 1H-15N HSQC peak intensities are shown for resonances of the spectral section displayed in (A); the averages of these intensities are fit to monoeXponentials. Apparent rate constants for the unfolding of domain I and the formation of domain II are 0.125 h⁻¹ and 0.121 h⁻¹, respectively, under the given conditions.
funded by the Swiss National Fund. Sequencing of Hydra ESTs was carried out by the Genome Sequencing Center, Washington University, St Louis with support from the National Science Foundation. The sequencing, assembly and annotation of the Nematostella vectensis genome was performed under the auspices of the U.S. Department of Energy’s Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, the Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, and the Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

Figure 4. Structural and Disulfide Switch upon Mutation of G11V and K21P in NW1

(A) The lowest-energy NMR conformers out of 100 calculated (see Supplemental Experimental Procedures) are shown as ribbon representations with cystines depicted in yellow and mutated residues 11 and 21 shown in red. Single point mutations of the bridge-state sequences can induce an abundance of more than 95% of either fold. Experimental details on NMR structure determination as well as PDB identifiers are given in the Supplemental Data; NMR structure bundles are shown in Figure S2. The comparison of the double mutant structure with the Mcol1C structure illustrates the identity of the synthetic with the natural fold (right). The only remaining difference in the mutated NW1 sequence is a glycine (G23) that induces a βII turn in position i + 3 of the synthetic NW1 sequence [23]; in comparison, there is a βI turn in Mcol1C.

(B) Sequence alignment of N- and C-terminal cysteine-rich domains of minicollagens from Hydra. Mutations shown to switch the domain structure are highlighted in light blue. The alignment indicates that the polarity between N- and C-terminal domains is conserved in minicollagens because identified determinants of the different structures are conserved.

References


Accession Numbers

Coordinates have been deposited in the Protein Data Bank, www.pdb.org, under PDB ID codes 2HM3 (NW1), 2HM4 (NW1 K21P, species I), 2HM5 (NW1 K21P, species II), and 2HM6 (NW1 G11V, K21P).