Theses and Dissertations

8-2017

Structural Characterization of Bacterial Collagenases

Ryan Bauer
University of Arkansas, Fayetteville

Structural Characterization of Bacterial Collagenases

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemistry

by

Ryan Bauer
University of Arkansas
Bachelor of Science in Chemistry, 2006

August 2017
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

_____
Dr. Joshua Sakon
Dissertation Director

_____   _____
Dr. Colin Heyes   Dr. Wesley Stites
Committee Member   Committee Member

_____
Dr. Suresh Thallapuranam
Committee Member

Abstract

Clostridium histolyticum secretes collagenases ColG and ColH to cause extensive tissue destruction during myonecrosis. The collagenases are multi-domain enzymes consisting of a N-terminal collagenase module, s1, polycystic kidney disease (PKD)-like domains (s2 in ColG; s2a and s2b in ColH) and collagen-binding domains (CBD) (s3a and s3b in ColG; s3 in ColH). The individual CBD and PKD-like domains chelate calcium to modulate stability and domain rearrangement. Though used by bacteria to break down collagen in the extracellular matrix during infection, therapeutic use of the enzyme to break down excess connective tissue has been approved by the FDA. Meanwhile, the use of the targeting segment to anchor therapeutics at the lesion site is currently underway. To better understand both the mechanism of collagenase during infection and facilitate its use in drug delivery, the high-resolution structures of the PKD-like domains and CBD of collagenases ColG and ColH were solved using X-ray crystallography. The structures of Ca^{2+}-absent (*apo*)-s2 and s2a, as well as Ca^{2+}-bound (*holo*)-s2a, s2b, s3a-s3b, and s3 are new structures, while the structure of *holo*-s3b was re-refined at higher resolution. Individually, the structures of the CBD are similar and share conserved Ca^{2+} and collagen-binding pockets. In the tandem CBD, s3a-s3b, the domains are related by a pseudo two-fold symmetry that may allow tandem CBD to bind to separate collagen molecules. Furthermore, three aromatic residues that were identified to be crucial to collagen-binding are fully conserved in s3, but only partially conserved in s3a. Evolutionary pressure likely decreased collagen affinity of s3a to prevent s3a-s3b from binding too tightly on the surface of collagen fibril. Meanwhile, the structures of the PKD-like domains are similar, though surface exposed aromatic residues are found only on the ColH PKD-like domains. Furthermore, the change in C_α B-factors upon Ca^{2+}-binding further suggests a divergence of function for ColH and ColG PKD-like

domains. In each case, Ca^{2+}-binding further stabilized the robust domains against denaturants heat, urea, and guanidium chloride. For s3a-s3b, the influence of Ca^{2+} was further explored using small angle X-ray scattering, which indicates the linker between domains gradually contracts in the presence of increasing Ca^{2+} concentrations.

Acknowledgments

First, I would like to thank my family for their unwavering support for me. I would next like to thank my research advisor, Dr. Sakon, for his support throughout this journey. Thank you for welcoming me into your lab and for always pushing me to be a better scientist. On a minor note, thank you for organizing the molecular gastronomy course and bringing by the extra food from the class. To my committee, Drs. Colin Heyes, Suresh Kumar, and Wes Stites, thank you for your patience when the time to organize committee meetings arose and for the advice you provided along the way.

I would also like to thank Dr. Srinivas Jayanthi for his efforts to debug the software associated with the Hitachi fluorescence spectrophotometer so that no longer crashed throughout the experiment. A special thank you goes to Kz Shein for his efforts to maintain the equipment I and so many other labs use. It did not matter if it was a water purifier, a centrifuge, an incubator, or the X-ray machine, if a problem arose, he responded without hesitation. Within the lab, I would like to thank Katarzyna Janowska and Keisuke Tanaka for being awesome lab mates and for being available whenever I needed help.

In 2012, I had the immense honor to travel to Okayama University in Japan and work with Drs. Osamu Matsushita and Takehiko Mima. Thank you for your mentorship and for your efforts to assist me with any problems I had.

Lastly, I would like to thank the National Institutes of Health and Arkansas Biosciences Institute, for research funding, the Chemistry Department for travel funding, and the Graduate School for conference travel funding and for the doctoral fellowship.

List of Publications

Chapter 2: Bauer, R., et al. (2013). "Structural comparison of ColH and ColG collagen-binding domains from Clostridium histolyticum." *J Bacteriol* 195, 318-327.

Chapter 3: Bauer, R., et al. (2015). "Crystal structures of three polycystic kidney disease-like domains from Clostridium histolyticum collagenase ColG and ColH." *Acta Cryst D* D71, 565-577.

Table of Contents

Chapter 1: Introduction

Collagen

In animals, procollagen chains containing Xaa-Yaa-Gly repeats, where Xaa and Yaa are frequently proline and hydroxyproline respectively, can assemble into triple helical tropocollagen. Mammalian heat shock protein 47 (HSP47) is a collagen specific chaperone that associates with procollagen and assists with folding into tropocollagen (*1*). In the fibrillar collagens, this tropocollagen further assembles into microfibrils and fibrils that provide the structural support to tissues such as skin and bone. Overall, the different types of collagen can be grouped into subfamilies based on their supramolecular structure: fibril forming types I, II, III, V, and XI collagens; network forming types IV, VIII, and X collagens; anchoring fibril type VII collagen, transmembrane domain type XIII collagen, and FACIT type IX, XII, XIV, and XVI collagens that interact with beaded filament forming type VI collagen (*2*). As part of the processing into mature type I or type III collagen, non-helical flanking segments called propeptides are proteolytically removed, though a small, non-helical segment called the telopeptide is retained. The telopeptides are the location of intramolecular cross-links that help stabilize the fibril and improve tensile strength (*3*). Three dimensional structural characterization of collagen fibril and fiber is hindered by the size and heterogeneity of the molecule, though X-ray fiber diffraction of type I collagen does suggest that the tropocollagen packs into a quasi-hexagonal arrangement that exposes the C-terminus of the molecules in the fibril (*4, 5*). The tight packing of collagen in the fibril, as well as the triple helical confirmation of tropocollagen shields peptide bonds from common proteases, such as trypsin. In response, specialized collagenolytic proteases are have evolved to degrade collagen, either for remodeling of aged or damaged collagen, or as part of bacterial infection. In mammals, a variety of matrix

1

metalloproteinases (MMPs) are used to break down collagen as part of the tissue remodeling process. Bacteria utilize collagenases, which can break down a wider variety of collagens and digest the collagen into shorter fragments compared to the MMPs.

Comparison of MMPs and bacterial collagenases

Mammalian MMPs and the bacterial collagenases from *Clostridium histolyticum*, which are discussed in this dissertation, are multidomain enzymes that utilize contrasting mechanisms to degrade collagen. Likewise, the motif responsible for zinc binding is different. Bacterial collagenases utilize the HEXXH motif, while MMPs utilize an elongated HEXXHXXGXXH motif (*6-9*). The MMPs comprise a network of enzymes that each fill a specific role in dismantling of collagen (*6*). While MMP-2 acts on the collagen triple helix, it can also hydrolyze gelatin (denatured collagen) and therefore, is also considered a gelatinase (*10*). For type I collagen, which is composed of two α1 chains and one α2 chain, the MMP cleaves the Gly775-Ile776 bond in the αI chain and the Gly775-Leu776 bond of the α2 chain (*6*). Hydrolysis at this site results in a ¾ length, N-terminal fragment called the TC_A fragment and a ¼ length, C-terminal fragment called the TC_B fragment. Conformational instability in these segments helps further unwind the collagen helix. The resulting gelatin can then be processed by MMP2. MMPs are multidomain enzymes that feature a catalytic domain that is connected to a four-blade β-propeller hemopexin domain. The zinc coordinating motif is found in the catalytic domain. MMP is expressed as a zymogen called pro-MMP. Structures of pro-MMP-1 (PDB ID: 1SU3) and active MMP-1 (2CLT) are available. Comparison of the two reveals the inhibitory mechanism. In the pro-MMP-1 structure, the N-terminal inhibitory peptide not only blocks the catalytic site using a cysteine-switch interaction with zinc (*7*), but also draws the domains together by interacting with the C-D loop of blade 1 in the hemopexin domain (*11, 12*).

Comparison with active MMP-1 also indicates that cleaving the inhibitory peptide creates a groove that forms the collagen binding site (*11, 12*). MMPs also feature "exosite" regions that are believed to be responsible for not only substrate recognition (*13*), but also unwinding collagen to allow only one chain in the catalytic site at a time (*14*). Analysis of the E219A mutant of MMP-1 and collagenous peptide using hydrogen deuterium mass spectrometry suggested that exosites were located on blades one and four of the hemopexin domain, while mutagenesis identified Ile290 and Arg291 specifically as exosites (*15*). Further mutagenesis studies identified Phe301 as the most vital residue in collagen binding (*10*). The crystal structure, solved in the absence of collagen, suggest that this residue is buried within the interface between the domains. Small angle X-ray scattering (SAXS) data, however, suggests that transient movement of the domains exposes this residue (*10*).

Unlike the MMPs, each bacterial collagenase can rapidly hydrolyze multiple sites on collagen (*16, 17*), and are utilized to destroy the host extracellular matrix during infection. However, controlled applications of collagenases can be tailored to target therapeutics to their active site, remove excess connective tissue, isolate cell cultures from the extracellular matrix, and reduce environmental pollution during leather production (*18*). These collagenases comprise peptidase families M9, S1, S8, and S53, and U32 of the MEROPS database (*19*). The M9 collagenases are zinc protease family and can be further divided into subfamilies M9A (secreted by *Vibrio*) and M9B (secreted by *Clostridium*). All enzymes in this family share an N-terminal catalytic unit. However, the remaining domain structure varies. In M9B, the C-terminus consists of one to three domains that are called the collagen-binding domain (CBD), and carry out the namesake non-covalent attachment of the enzyme to collagen. In between these domains are the polycystic kidney disease (PKD)-like domain. The role of the domain is not well defined, though

a potential division of roles in relationship to the number of CBD present is presented in chapter 4. Except for ColT, the collagenase secreted by *C. tetani*, which has no PKD-like domains, the M9B collagenases contain one or two PKD-like domains. In M9A, the collagenases can be further divided into class II and class III proteases (class I proteases belong to the M4 family have no collagenolytic activity). Only the collagensase module is found in class II collagenases (*20, 21*), while class III collagenases have a PKD-like domain and C-terminal domains, which are classified as prepeptidase C-terminal domains, as it is currently not known if the domain is a collagen-binder (*22-24*). The collagenase from *Grimontia hollisae* is closely associated with M9A class III collagenase, though it does not have a PKD-like domain. The *Grimontia* genus was originally classified as a member of the *Vibrio* genus (*25*), though in 2003, the bacterium was reclassified as a member of *Grimontia* (*26*). To date, no structural knowledge for the domains of M9A collagenase exists. The enzyme has commercial interest as its specific activity is four times greater than that of *C. histolyticum* (*27*). Like the class III collagenases, the enzyme from *Grimontia* can degrade soluble and insoluble collagen, as well as the and Pz-PLGPR peptide (*28*). Unlike these class III enzymes, the enzyme cannot degrade casein (*27*). To date, no structure of the M9A collagenase domains is available.

The current structural knowledge of bacterial collagenases comes from ColG and ColH, which are secreted by *C. histolyticum*. As noted above, each contains a collagenase module, called s1, though ColG has one PKD-like domain, s2, and two CBDs, s3a and s3b, while ColH has two PKD-like domains, s2a and s2b, and only one CBD, s3. The first domain to be solved was ColG-s3b (*29*), which was solved both in the presence of calcium ions (PDB ID 1NQD), and in the absence of calcium ions (PDB ID: 1NQJ). In each condition, the overall structure has a β-jelly roll conformation. The CBD each bind to two Ca^{2+}, while the PKD-like domains bind to

one Ca^{2+}. Ca^{2+}-binding stabilizes s3b against denaturants (*30*) and is shown to stabilize s3 and

the PKD-like domains in chapters two and three respectively. In chapter four, the role of Ca^{2+} in

triggering domain rearrangement in s3a-s3b is shown and the potential implication for collagen-

binding by the segment is discussed. Though these domains do not possess collagenolytic

activity, they make up the targeting segment of the collagenase and are necessary for the enzyme

to dismantle insoluble collagen. Beyond applications of collagenase to desired removal of excess

collagen, the targeting segments themselves are being evaluated as drug delivery vehicles. The

structure-based contribution of either individual CBD, s2b-s3, or s3a-s3b is discussed in chapters

two through four. Significant to the goal of targeted drug delivery, both collagen-binding and

therapeutic activity are retained in fusion proteins of the drug and collagen-targeting segment.

In 2011, the first structure of a collagenase module, from ColG, was solved (*31*). Three

structures were solved: *apo* (calcium and zinc absent) Se-Met substituted structure (PDB ID:

2Y3U), the zinc containing structure (2Y50), and the structure inhibited by isoamyl-phosphonyl

Gly-Pro-Ala (2Y6I). The s1 structure resembles a saddle composed of an N-terminal activator

domain and C-terminal catalytic domain connected by a nine residue linker (*31*). Within the

activator domain ten HEAT motifs are found. These motifs may facilitate in substrate

recognition, and may also help with catalytic turnover after hydrolysis. The overall domain is

necessary to hydrolyze soluble collagen, though the catalytic domain itself is the minimal

gelatinolytic segment (*31*).

Given that the bacterial collagenases can each hydrolyze collagen fibril into small

peptides, it would initially seem odd that *C. histolyticum* would invest in carrying two

collagenases in its genome. Utilizing two collagenases, however, facilitates the bacterium's goal

to expose new nutrient sources in the host before the host immune system can respond. ColG and

ColH have been shown to work synergistically to degrade collagen (*32*), in part by initially acting on separate regions (*17*). Though the technique has not been applied to ColH, high-speed atomic force microscopy has been used to observe collagen fibril digestion by ColG in real time (*33*). ColG was initially observed anchoring onto disordered regions of the fibril and then processively moving towards the N-terminus of the fibril. Curiously, trimmed collagen molecules, during this movement were observed migrating onto and thickening neighboring fibrils. The structure of s3a-s3b presented in chapter four may help to explain these phenomena.

Observation of rare cis-peptide bonds in proteins

The structures of s3b and s2a contain a unique bond rarely found in protein, the non-prolyl *cis*-peptide bond. The bond is extremely rare and energetically unfavorable, and thus, tends to be conserved in highly restrained environments where it fulfills key structural and functional roles (*34*). In s3b, the bond between N-terminal linker residues Glu901 and Asn902 isomerizes into the *cis* confirmation upon Ca^{2+} binding (*29*). While stabilized by the ions, isomerization serves critical function to position s3a into the unique pseudo-two-fold arrangement described in chapter four. The peptide bond, unlike the phi and psi bonds that also make up a protein's main chain, is rigid due to resonance-induced delocalization of electrons in the carbonyl π bond and nitrogen lone pair. Consequently, the peptide bond torsion angle can be characterized as either *cis*, where the Cα of the connected amino acids lie on the same face of the peptide bond (Cα-N-C-Cα torsion angle close to 0°), or *trans*, where the Cα atoms lie on opposite faces (torsion angle close to180°).

The isomerization barrier from *trans* to *cis* varies. Prolyl peptide bonds connect any amino acid to proline (Xaa-Pro), while non-prolyl peptide bonds, the nitrogen comes from any amino acid that is not proline. For non-prolyl peptide bonds, the *trans* confirmation is

significantly favored as steric hindrance between Cα atoms in the *cis* confirmation is absent in the *trans* confirmation. Due to this steric strain, non-prolyl *cis*-peptides are extremely rare. A survey of 571 proteins revealed that only 0.03% of non-prolyl peptide bonds were in the *cis* confirmation (*35*). For prolyl peptide bonds, steric strain is also found in the *trans* confirmation since the Cδ essentially destabilizes the confirmation by taking the place of the Cα atom after isomerization. The *trans* confirmation is still favored, however. In the above survey, 94.8% of prolyl peptide bonds were found in the *trans* configuration (*35*). It is significant to note that in the above survey, the resolution of the crystal structures studied was 2.0 Å and above. A recent survey of non-prolyl *cis*-peptide bonds indicated that identified *cis*-peptide bonds were likely erroneously placed in lower resolution structures (*36*). The lowest resolution structure of s3b is 1.65 Å, while the lowest resolution structure of s2a is 1.9 Å. Each structure is solved at sufficiently high resolution to justify the observation of non-prolyl *cis*-peptides. Due to their rareness and their implications for the structure and function of proteins, non-prolyl *cis*-peptides have emerged as a topic for theoretical and modeling studies. The role of Ca^{2+} in catalyzing the isomerization and stabilizing the subsequent non-prolyl *cis*-peptide bond in s3b has been studied using a combination of quantum mechanics and molecular mechanics (*37*). The results showed that the presence of two Ca^{2+} lowers the free energy barrier for isomerization by 11 kcal/mol and helps to initiate a hydrogen bonding network that stabilizes the region. The non-prolyl *cis*-peptide bond is not induced by Ca^{2+} binding in s2a, but its role in replacing proline to form a β-bulge could facilitate modeling studies on how the bond influences protein folding.

Chapter 2: Structural comparison of ColH and ColG collagen-binding domains from *Clostridium histolyticum*

Abstract

Clostridium histolyticum secretes collagenases, ColG and ColH that cause extensive tissue destruction in myonecrosis. The C-terminal collagen-binding domain (CBD) of collagenase is required for insoluble collagen fibril binding and subsequent collagenolysis. The high resolution crystal structures of ColG-CBD (s3b) and ColH-CBD (s3) are reported in this paper. The new X-ray structure of s3 was solved at 2.0 Å resolution (R=17.4%, R_{free}=23.3%), while the resolution of the previously determined s3b was extended to 1.4 Å (R=17.9%, R_{free}=21.0%). Despite sharing only 30% sequence identity, the molecules resemble one another closely (r.m.s.d. C_α = 1.5 Å). All but one residue whose sidechain chelates with Ca^{2+} are conserved. The dual Ca^{2+} binding site in s3 is completed by an unconserved aspartate. Differential scanning calorimetric measurements showed that s3 gains thermal stability, comparable to s3b, by binding to Ca^{2+} (*holo* T_M=94.1 °C, *apo* T_M=70.2 °C). *Holo* s3 is also stabilized against chemical denaturants, urea and guanidine HCl. The three most critical residues for collagen interaction in s3b are conserved in s3. The general shape of the binding pocket is retained by altered loop structures and side chain positions. Small angle X-ray scattering data revealed that s3 also binds asymmetrically to mini-collagen. Besides the calcium-binding sites and the collagen-binding pocket, architecturally important hydrophobic residues and hydrogen-bonding network around the *cis*-peptide bond are well-conserved within metallopeptidase subfamily M9B. CBDs were previously shown to bind to extracellular matrix of various tissues. Compactness and extreme stability in physiological Ca^{2+} concentration possibly make both CBDs suitable for targeted growth factor delivery.

Introduction

Clostridium histolyticum is one of the causative agents for clostridial myonecrosis. The organism produces collagenases responsible for extensive tissue destruction. Although collagenases are harmful during infection, their ability to break down a wide variety of collagen types make them beneficial as a treatment for excessive connective tissue build up. The enzymes are approved by the Food and Drug Administration (FDA) to break down the tough cords in Dupuytren's contracture (*38*). The two classes of collagenase, ColG and ColH, differ in domain structures (s1, s2, s3a, s3b for ColG; and s1, s2a, s2b, s3 for ColH). The s1 is the collagenase module that belongs to a metallopeptidase M9 subfamily B (M9B) of bacterial collagenases. The structure of ColG-s1 was recently solved, and revealed an overall saddle shape built up by an activator domain linked to a peptidase domain (*31*). The amino acid sequences of s2, s2a and s2b resemble a polycystic kidney disease domain. The role of this domain is speculative. It was shown that the PKD-like domain of the serine protease from *Pseudoalteromonas* sp. SM9913 could bind to and swell collagen microfibrils (*39*). The PKD-like domain has also been shown to enhance collagen binding (*40*). The C-terminal domains, s3a, s3b and s3 (swissprot: Q46085, pubmed: BAA34542) are homologues with approximately 120 amino acid residues, and are classified as bacterial pre-peptidase C-terminal domains (PPC super-family). The PPC domain is found in some members of metalloprotease families M4, M9 and M28 as well as in a serine protease family S8. *In silico* work on the PPC domain suggests a distant relationship with the PKD-like domain (*41*). Moreover, the role of the PPC domain could be diverse. Some of the domains are removed after secretion for certain protease activation (*41*), but others are retained in the mature enzymes (*42*). Both ColG-derived s3a-s3b and ColH-derived s3 bind to either insoluble collagen *in vitro* or to collagen fibrils in the extracellular matrix of various tissues (*43,*

9

44). Therefore, each of these PPC domains is sub-classified as a collagen-binding domain (CBD).

CBD derived from ColG (s3b) has been extensively characterized. It is a β-sandwich fold, and a pocket formed on a β-sheet was identified to interact with the substrates by alanine-scan mutagenesis and mini-collagen binding assays (*29*). The s3b binds to collagenous peptides with triple helical conformation but not to the similar peptides lacking triple helical conformation or to gelatin (denatured collagen), suggesting that the CBD-collagen interaction is conformation-specific (*40, 43*). The most recent solution studies of s3b in complex with collagenous peptides by small angle X-ray scattering (SAXS), and heteronuclear single quantum coherence (HSQC) NMR titration of various spin-labeled collagenous peptides to ^{15}N-labeled s3b showed that s3b binds unidirectionally to the C-terminus of collagenous peptides (*45*).

In the presence of Ca^{2+}, s3b shows shortened hydrodynamic radius, better stability and more efficient substrate binding (*29*). The X-ray crystal structures of s3b were solved in the presence of Ca^{2+} (*holo*) as well as, in the absence of Ca^{2+} (*apo*) to show a secondary structure transformation of the linker at its N-terminus (*29*). The linker is coiled as an α-helix in the *apo* form, but is unwound to form a parallel β-strand in the *holo* form, which possibly resulting in domain rearrangement. Two Ca^{2+} bind cooperatively with macroscopic association constants of $K_1=5.01x10^5$ M^{-1} and $K_2=2.28x10^5$ M^{-1} to trigger the structural transformation (*30*). Two Ca^{2+} ions chelate to s3b in close proximity to each other (3.7Å). The N-terminal linker of s3b adopts an energetically unfavorable non-prolyl-*cis*-peptide in the presence of Ca^{2+}. Simulations provided mechanical insights of calcium ions catalyzing the *trans-cis* isomerization of the non-prolyl peptide bond (*37*). Previously, it was shown that ColG-s1 does not require the PKD-like

10

domain or CBDs to degrade gelatin (*40*) or soluble collagen (*31*). These domains, however, are necessary to degrade collagen fibrils.

The s3 from ColH has been shown to anchor growth factors to extracellular matrix, and extend their half-lives. A fusion protein of basic fibroblast growth factor and s3 domain stayed at the site of injection for more than 10 days to promote cell proliferation (*46*). Although various preclinical applications are currently being examined (*46-50*), s3 has not been characterized in detail. The s3 and s3b share 30% sequence identity. Only three aromatic residues within nine residues in s3b involved in the substrate binding are conserved in s3. Based on the sequence alignment, one of the Ca^{2+}-chelating aspartates (Asp927) in s3b is replaced by serine (Ser896) in s3 (Note that in this manuscript, the sequence number for each CBD refers to the mature protein: the s3b sequence numbering reflects a 110-residue long secretion tag being cleaved, while s3 numbering reflects a 40-residue long tag being cleaved). Each of the two sidechain oxygen atoms of Asp927 chelates with a different Ca^{2+} ion. It was uncertain as to how Ser896 in s3 could construct the dual calcium-binding site in place of the bidentate Asp. To address its structure and function relationship applicable for various drug delivery applications (*46-50*), structural work on s3 was initiated. The structures of dissimilar CBD molecules in metallopeptidase family M9B enabled us to compare and contrast their Ca^{2+} sites and collagen-binding mechanism.

Methods

Expression and purification of collagen-binding domain

Expression and purification of collagen-binding domain as a glutathione S-transferase (GST) fusion protein was achieved using methods described by Matsushita (*43*).

Crystallization and Structure Determination of s3 from ColH

Recombinant s3 (7.0 mg/mL in 10 mM Tris-HCl pH 7.5) crystals were grown by hanging drop vapor diffusion from 28% (w/v) PEG3350, 0.1 M Li_2SO_4 0.1 M Tris-HCl (pH 6.5), and 2 mM $CaCl_2$. The crystals were subsequently flash frozen in liquid nitrogen without the need of additional cyroprotectant and stored in liquid nitrogen until data collection. In house X-ray diffraction facility (Rigaku 007, Osmic Blue confocal mirrors, Saturn CCD detector) was used to characterize the unit cell parameters for s3. The calcium bound s3 crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell size a=62.0, b=66.2, and c=96.4 Å. The data were collected at -164°C and diffracted to 2.6 Å resolution. Single crystal diffraction data were then collected at the 19-ID beam line at the Advanced Photon Source at Argonne National Laboratories. The crystals diffracted to 2.0 Å resolution. Data were processed using HKL3000. Molecular replacement technique using PHASER on a data set truncated to 3 Å resolution located three molecules in an asymmetric unit (*51*). The F_o-F_c difference map clearly showed two calcium atoms that were omitted from the search model, and increased confidence in the phase information. Where possible, amino acid residues of molecule A were corrected to those of s3. Using the partially built molecule A as a search model, molecular replacement techniques were repeated yielding phases that allowed for building the complete s3. The models were refined using REFMAC to 2.0 Å resolution (*52*). In the strong difference map peaks (F_o-F_c with σ>3.5 and 2F_o-F_c with σ>1.0), 326 solvent molecules were built. NCS restraints were not used during the refinement. PROCHECK was used to ensure the stereo-chemical quality of the protein. Data collection and refinement statistics are summarized in Table 1.

Crystallization and structure refinement of s3b from ColG

Recombinant s3b (5.5 mg/mL in 50 mM Tris-HCl pH 7.5) crystals were grown by hanging drop vapor diffusion from 25-27% PEG 3350, 100 mM sodium acetate pH 4.6 and 600

mM LiCl at 4°C. As it is known to exist in two structural forms, the s3b was supplemented with $Ca(NO_3)_2$ to a concentration of 50 mM in order to push the equilibrium towards the ion-bound form (*29*). Crystals were dragged through 100% paraffin oil to remove excess solvent, and then frozen in liquid nitrogen. The initial conditions suitable to grow s3b crystals were identified by the sitting drop method using high throughput screen (Hampton Research Crystal Screen HT). Cryo-protection procedures resulted in a low mosaic spread, 0.46°, and extended our data to 1.35 Å (unique reflections = 44,856) versus 1.65 Å resolution (unique reflections = 27,154) for our previously deposited structure (*29*). The data were collected at BioCARS 14ID-B at Advanced Photon Source. All the data were indexed and scaled using Denzo and Scalepack from HKL2000 (*53*). SHELX97(*54*) and REFMAC (*52*) was used for refinement (Table 1). The *holo* structure (PDB code 1NQD) was used as a starting model. The manual adjustments were aided by the use of MIFit (*55*). Seven alternate confirmations for surface residues (919, 959, 968, and 976 of molecule A, 895, 968, and 976 of molecule B) were built. The positional and anisotropic B-factor refinements were carried out using initially Shelxl97(*54*) then REFMAC (*52*). Hydrogen atom positions were constrained to the ideal position during the refinement. However neither non-crystallographic symmetry (NCS) restraints nor Ca…O bond restraints were applied. Positional refinement, manual adjustments, and anisotropic B factor refinement resulted in an R=17.9%, R_{free}=21.0% (Table 1). The anisotropic temperature refinement is justified because R_{free} dropped nearly 5%. Seventy percent of all atoms including water molecules are very well ordered (B-factor below 30 $Å^2$), and their positional errors are less than 0.05Å.

Small Angle X-ray Scattering Experiments

The SAXS data were collected on solutions of s3, various collagenous peptides, and the CBD-collagenous peptide complexes (between 4 and 6 mg/mL) in 10 mM Tris-HCl (pH 7.5)

containing 100 mM NaCl and 20 mM CaCl$_2$ at the XOR beamline of sector 12-ID at the Advanced Photon Source (APS) in the Argonne National Laboratory. The main advantage of X-ray scattering is that it can be carried out in solution in nearly physiological conditions ([13]). The monochromatic radiation source (10keV, λ=1.2398 Å) was the APS Undulator A insertion device using a Si-111 monochromator, with 1:1 horizontal focusing and higher harmonic rejection from a Rh coated mirror, and beam defining slits set at 0.3 mm vertical and 0.25 mm horizontal. A 1.6 mm diameter capillary flow-cell with a flow rate of 4 μL/second was used to capture four frames with exposure time of 10 seconds. The SAXS detector used was a Mar165 scintillator fiber-optic coupled CCD detector and covered the momentum transfer range $0.005 < q < 0.198$ Å^{-1}, where $q = 4\pi \sin\theta/\lambda$ (2θ is the scattering angle). The WAXS detector was a custom made Roper scintillator fiber-optic coupled CCD detector and covered $0.191 < q < 1.8$ Å^{-1} S ([56]).

All scattering data were acquired at 10ºC. The four scattering patterns from each detector were averaged and merged with the rejection of outlying scans. For further analysis the program IGOR Pro 5.5A (WaveMetrics) was used. The scattering profiles of the protein, peptide and their complexes were obtained after subtracting the buffer profiles. The reduced scattering data were plotted as scattering intensity $I(Q)$ vs. Q (Supplementary Fig. 1A). The radius of gyration, R_g, was obtained from the Guinier approximation by linear least squares fitting in the $QR_g < 1$ region, where the forward scattering intensity $I(0)$ is proportional to the molecular weight of the protein complex. An indirect Fourier transformation of the $I(Q)$ data, using GNOM ([57]), provided the pair-distance distribution function $P(r)$ in the real space. Where $P(r)$ intersects with x-axis represents the maximum diameter, D_{max} (Supplementary Fig. 1B and table 2). The molecular envelopes were constructed from the SAXS data after *ab initio* calculation with the

program GASBOR (*58*). The *P(r)* ranges were varied to generate various (>10) molecular envelopes using GASBOR (*58*). No symmetry restraints were applied to any of the shape reconstructions. The ten highest scoring *ab initio* models were chosen and averaged using DAMAVER (*59*). Atomic models were docked into *ab initio* envelopes with the program SUBCOMB (*60*). The theoretical scattering curves from the atomic models were calculated and compared with the experimental curves with the program CRYSOL (*61*). Rigid-body modeling to the experimental scattering data was performed using SASREF (*62*).

Docking Model

The mini-collagen:CBD complex was generated from the CBD and 1K6F for the collagenous peptide. To obtain the complex, the soft docking algorithm BiGGER was used (*63*). The manual adjustments were aided by the use of MIFit (*55*).

Fluorescence Spectroscopy Measured Equilibrium Denaturation of s3

For s3, λ_{max} for the native protein occurs at 310 nm while λ_{max} for the denatured protein occurs at 350 nm. The ratio of intensity at 350 nm versus the intensity at 310 nm was used to track the unfolding process. Fluorescence data were collected on a Hitachi F-2500 fluorimeter with excitation and emission bandwidths at 2.5 nm and 10 nm, respectively. The excitation wavelength used was 280 nm, and fluorescence emissions were monitored between 300 nm and 450 nm. During thermal denaturation trials, temperature of the protein solution was maintained with a Neslab RTE-110 circulating water bath (Thermo Scientific, Newington, NH, USA). In the thermal denaturation trials, the protein concentration was 3µM. While in the chemical denaturation trials, the protein concentration was 1.5 µM. *Holo*-s3 was supplemented with 1 mM $CaCl_2$, while *apo*-s3 was supplemented with 1 mM EDTA. In all cases, the protein was diluted in 10 mM Tris-HCl and 100 mM NaCl, and the pH was maintained at 7.5. When urea was used

as the denaturant, the native s3 was exposed to concentrations of denaturant that increased linearly by 0.2 M intervals from 0.0 to 9.8 M. When guanidine hydrochloride (GuHCl) was used as the denaturant, the s3 was exposed to concentrations of denaturant that increased linearly by 0.2 M intervals from 0.0 to 5.8 M. During thermal denaturation, the s3 was exposed to temperatures that linearly elevated by 2.5°C intervals from 7.5 to 100°C. ΔG_{H2O} and m values were calculated as described (*30*).

Differential Scanning Calorimetry

DSC was conducted using a Nano-Differential Scanning Calorimeter (NDSC) model CSC 6300 (Calorimetry Sciences Corporation). The s3 was dissolved in 10 mM Tris-HCl (pH 7.5), and 1 mM $CaCl_2$ for the *holo* form to a final concentration of 0.5 mg/mL. The *apo*-s3 was dissolved in 10 mM Tris-HCl (pH 7.5) to a final concentration of 0.5 mg/mL. Following a scan of buffer vs. buffer to establish the baseline, samples were run as follows. Degassed samples were heated from 25°C to 130°C (1°C/min) and then cooled back to the initial temperature at the same rate. A second heating scan was then performed to test the reversibility of the process. Unfolding was reversible if a difference in heat capacity was again observed during the second heating scan. DSC analysis software (CpCacl1) provided by the instrument manufacturer was used to analyze the transition temperature after subtracting the baseline.

Results and Discussion

Structure Description of s3 at 2.0 Å Resolution

The structure of s3 from ColH is similar to the structure of s3b from ColG. The s3 also adopts a β-sandwich 'jelly-roll' conformation composed of nine β-strands. Two calcium atoms per molecule are found between the linker (861-878) and the β-strand C. There are three molecules in the asymmetric unit. Molecule A consists of 117 amino acids (863-979), molecule

B consists of 120 amino acids (862-981) and molecule C consists of 121 amino acids (861-981). The backbone structures of the three molecules (A-C) are virtually identical. The root mean square deviations of C_α positions between molecules A and B, determined using Superpose from the CCP4 suite of programs, was 0.29 Å, that between B and C was 0.45 Å, and that between A and C was 0.43 Å. Three molecules were independently refined without NCS restraints. Inherent flexibility was observed at both the N-terminus (861-863) and C-terminus (977-981). Another deviation occurs only in molecule C at the loop between β-strands H and I (966-969) caused by a peptide flip between Ser967 and Tyr968. Both calcium atoms are ordered (B-factor < 18 Å^2).

Structural Comparison of holo s3b and holo s3

Here we present the first structure of a domain from ColH collagenase. The tertiary structures of the CBD molecules are very similar to one another (C_α r.m.s.d = 1.4 Å), despite being of low sequence identity. β-Sheet regions are the most similar. Even the linker residues 898-910 of s3b and 867-879 of s3 adopt very similar structure. The Ser874–Thr877 segment of s3 adopts a short 3_{10} helix as also observed in the Ser905-Lys908 segment of s3b. In both structures two calcium ions bind between the N-terminal linker and the β-strand C. The s3 does not adopt β-strand A found in s3b, and the region adopts a random coil. Also, the solution structure of *apo*-s3b does not form β-strand A(*64*). The crystal structures of s3 and s3b are most divergent at the N-terminus and loops connecting the β-strands. For example a loop of 8 residues (Glu961-Ile968) between β-strands E and F in s3b are line broadened and non-assignable on NMR and the flexible loop is unobserved in the crystal structure (*29, 45*). In contrast, the equivalent loop in s3 (Glu930-Val935) is well ordered. Another significant deviation in loop structures occurred at a loop between strands B and C (924-928 in s3b; 894-899 in s3). The N-termini of s3b (891-897) and s3 (861-868) adopt very different paths. It is possible that this

deviation can be explained by crystal packing restraining the position of the termini. Termini deviation as a result of crystal packing for example, was seen in the N-terminal linker of *apo*-s3b (*29, 64*).

Coordinations around the two calcium atoms in s3 are virtually identical to those described in detail for s3b later. Seven oxygen atoms of five side chains and one main-chain form pentagonal-bipyramid coordination (Fig. 1). Eight oxygen atoms of four side chains and a main-chain and one water form square antiprismatic coordination. The average bond distances observed in s3 are similar to those in s3b (Supplementary Tables 2 and 3). There are some differences between the two CBD molecules. Asp904 in s3b is replaced by Asn873 in s3; and therefore the net charge around the calcium sites in s3 is -4. Earlier sequence alignment of CBD molecules showed that Asp927 in s3b is replaced by Ser896 in s3 (*29*). However, serine is rarely found to chelate with Ca^{2+} (*65*). The crystal structure of s3 revealed that Asp897 is found in place to chelate with Ca^{2+} atoms and Ser896 is pointed away from the ions. In order for Asp897 to chelate equivalently to Asp927 of s3b, a loop (894-898) of s3 meanders differently to position OD1 and OD2 atoms of Asp897. Two energetically unfavorable torsional strains were found in s3b (*29*). One of them is the peptide bond between residues Glu901 and Asn902 in s3b that adopts an uncommon and energetically unfavorable non-prolyl-*cis*-peptide conformation. OD1 of Asn902 makes a hydrogen-bond with main-chain N of Asp904. The interaction is one of the important hydrogen-bonds in the stabilization of the transition state for the peptide isomerization (*37*). In contrast, the structurally equivalent peptide bond between Glu870 and Pro871 in s3 adopts a common prolyl-*cis*-peptide conformation (Fig. 2). The second torsional strain was found in the side chain χ_1 angle of residue Tyr931 of s3b. Tyr931 adopts an energetically unfavorable angle of 64°, which is stabilized by hydrogen- bonding with NE2 of His959 of s3b

18

(distance = 2.6 Å). The equivalent residue Tyr900 of s3 adopts a favorable χ_1 angle of -68° (Fig. 3). The s3b is extremely thermostable (T_M=93°C) despite the torsional strains (*30*). The s3 is only slightly more stable (T_M=94°C). The stability of s3 is described in more detail in the following section.

The collagen-binding surface in s3b was identified by binding assay of 32 surface mutants (*29*) and by HSQC-NMR titration (*45*). Three aromatic side chains critically involved in collagen-binding are positioned similarly in s3. Two tyrosine residues (Tyr970 and Tyr994) in the s3b are conserved in s3 (Tyr937 and Tyr962), but Tyr996 is replaced by phenylalanine (Phe964) in the s3. Seven other residues identified to interact with the collagenous peptides are not conserved in s3. SAXS and NMR studies of s3b in complex with collagenous peptides showed that Ser928 and Arg929 interact with the C-terminus of collagenous peptide (*45*). Neither residue is conserved in s3. In s3, main-chain carbonyl Thr895 occupies where Ser928 side chain is found, and the three-dimensional space occupied by Arg929 in s3b is occupied by Gln898 and Tyr900 in s3 (Supplementary Fig. 2). Despite substitutions, the resulting collagen-binding pocket in s3 closely resembles that of the s3b. Using SAXS of s3 in complex with [G(POG)₇PRG]₃ showed that s3 also binds asymmetrically to the collagenous peptide (Fig. 4). Our previous work demonstrated that both (POG)₁₀ (hydroxylproline is represented as O) and G(POG)₇PRG adopt a triple helical structure in solution. They are thus represented as [(POG)₁₀]₃ and [G(POG)₇PRG]₃, respectively. The resulting atomic models for s3 and s3b fit very well into the *ab initio* SAXS envelopes. The theoretical R_g values calculated from the complex models of s3 and s3b are in excellent agreement with the experimental values (Table-2). The theoretical scattering curves calculated from the models using CRYSOL (*61*) are superimposable with the experimental curves with minimal discrepancy (Table-2). Rigid body modeling to the

experimental scattering data was performed using SASREF (*62*). Collagenous complexes with either s3 or s3b were fitted independently to the X-ray scattering data. The simulations were repeated 10 times. The model with the lowest value is shown (Fig. 3 and Supplementary Fig 3). S3b binds asymmetrically to the C-terminal $(POG)_3$ repeats or C-terminal $(POG)_2PRG$ (*45*). Since critical aromatic residues are oriented similarly in both structures, it is likely that s3 also targets the C-terminal region. The respective collagenase could target common structural features in collagen fibril. C-terminal region of collagenous peptide is slightly untwisted (under-twisted). Our studies on s3b and s3 imply that the cleft-like shape of the binding pocket scans the collagen for under-twisted regions and conserved aromatic residues of the pocket intercalate to the triple-helical collagen (*66*).

Ca^{2+} Coordination of s3b at 1.4 Å Resolution

The 1.4 Å resolution structure of the Ca^{2+} bound s3b showed excellent electron density for nearly all protein atoms and water molecules. The protein is crystallized as a non-crystallographic symmetry (NCS) related dimer. Since the structural features have already been described for s3b, the discussion will focus on calcium binding sites. Calcium associates with s3b in two coordination geometries, pentagonal bi-pyramidal and square antiprismatic (Fig. 1). Both calcium atoms were highly ordered (B-factor <14 $Å^2$). The pentagonal bi-pyramidal coordination is commonly observed in calcium chelating proteins, of which calmodulin's EF hand motif is the classic example. Five ligating atoms encircle the calcium ion as a near-planar pentagon and two remaining ligands sit at axial positions above and below. As in the ligation pattern of EF hands, both carbonyl oxygen atoms of s3b's Glu901 provide two of the five equatorial ligation positions. The atoms describing the equatorial plane have an average deviation from planarity of 0.16 Å (Supplementary Table 2), which compares favorably to the

0.18 Å average deviations across calcium sites in calmodulin (PDB code 1OSA) (*67*). While calmodulin has a water molecule occupying one axial position, allowing for greater flexibility in ligand position, both axial positions in s3b are supplied by amino acid side chains (Asn903, Asp927).

The s3b's second calcium site with square antiprismatic coordination sits 3.7Å from the pentagonal bi-pyramidal site. The planar faces of calcium-bound s3b's square antiprism deviate only slightly from flat geometry (Supplementary Table 3). One face shows an average out of plane deviation by a mere 0.03 Å for coordinated atoms, while the other face shows only 0.13 Å deviations. These values vary from few other available protein structures with square antiprismatic bound ions (Supplementary Table 4). Square antiprismatic coordination is relatively uncommon among macromolecules (*65*), and few are currently found in the Protein Databank. Ding *et al*., described a Zn^{2+} bound ribonuclease T_1, at 1.8 Å resolution, in which six water molecules and a sole aspartate coordinate zinc (*68*). When defined as square antiprismatic, the carbonyl oxygen atoms of aspartate provide a ligand for each square face, creating significant distortions from planar geometry for these faces and a very short ion-to-plane distance. On further examination, the ion coordination in ribonuclease T_1 is best described as trigonal dodecahedral, rather than square antiprismatic, which explains the high deviations from planarity. Potassium ions are found in square antiprismatic coordination in two structures, the KcsA potassium channel and an *Oxytricha nova* G-quadruplex (*69, 70*). Main chain carbonyl oxygens coordinate the KcsA potassium ions within the channel, and due to the four-fold crystallographic symmetry, the square faces are perfectly flat. In the G-quadruplex structure, an unusual DNA structure found in telomeric DNA, potassium ions sit similarly in the center of a channel created this time by nucleotides. Thymine base O6 atoms coordinate the potassium and

provide near-planar square faces, with two cubic coordinations sandwiching a square antiprismatic coordination (Supplementary Table 4).

In addition to the CBDs, two calcium-binding proteins that utilize square antiprismatic geometry are found in the literature. Thermitase, an extracellular serine proteinase, when crystallized in 100 mM calcium revealed a square antiprismatic calcium site in addition to two pentagonal bipyramidal sites. At lower calcium concentrations the site exhibits pentagonal bipyramidal coordination (*71*). Secondly, alpha-amylase binds two calcium ions with CN=8 (*72*). One of these sites follows a square antiprismatic geometry with three water molecules acting as ligands, and the deviations from planarity are near that of the thermitase coordination (Supplementary Table 4). However, not all binding sites with CN=8 have square antiprismatic geometry. The second site in alpha-amylase involves five waters and is best described as cubic. Proteinase K binds to one calcium ion with CN=8 (*73*), but coordination is dodecahedral in geometry. Four water molecules act as ligands.

The bidentate ligating nature of aspartates and glutamates, often found in ion coordination, distort geometries from ideal, inorganic values. Therefore, square faces are rarely square, with the exception of proteins taking advantage of multimerization as in KscA. Previous publications showed that Ca^{2+} is involved in activation but not K^+. However, Li^+ was found in the Ca^{2+} binding site (*29*). Barring occupancy differences, Ca^{2+} and K^+ would have identical electron density in an X-ray diffraction experiment. For the proteins shown in Supplementary Table 4, the distance of calcium from the square faces formed in antiprisms averages 1.37 Å, while potassium distances are a longer 1.58 Å. Additionally, the average ion-to-ligand distance for calcium and potassium are 2.54 Å and 2.81 Å, respectively. The larger coordination radius observed for potassium is not surprising since its ionic radius is greater than that of calcium.

Based on the coordinate errors for oxygen atoms, the errors associated with the bond distances in this study (Supplementary Table 4) are estimated to be less than 0.03 Å even accounting for 10% underestimation (9). Based on the analyses of bond distances, it is clear that two Ca^{2+} bind to s3b.

Another novel aspect of the dual calcium site is that it contains the rare tridentate acidic side chains (Glu901 & Asp930). As far as we are aware, tridentate acidic side chain has never been observed. Previous analysis of 1,605 Ca^{2+} binding sites showed that bidentate oxygen atoms chelate with longer Ca^{2+}-O distances than those found in monodentate (2.6 Å vs. 2.4 Å) (65). Indeed, Ca^{2+}-O in bidentate residues (Glu899 & Asp927) is longer than that in monodentate (Asn903 & Asp904). We also note that the two bidentate residues chelate to Ca^{2+} differently. Two carbonyl oxygen atoms of Glu899 side chain chelate with one Ca^{2+} atom where as two carbonyl oxygen atoms of Asp927 chelate with two different Ca^{2+} atoms. Ca-O-C angles in Glu899 (89°, 94°) are similar to the values observed for bidentate glutamates in other Ca^{2+} binding proteins (93.6±11.3°) (65), but the angles observed for Asp927 (131° and 144°) are closer to angles observed in monodentate (140.4±15.2°) (65). In both tridentate residues (Glu901 & Asp930) one of the three bond distances is significantly shorter (<2.36 Å) than other two. Ca-O-C angles (87°, 96°, and 159°) are virtually identical for both tridentate residues. Electrostatic repulsion between the cations is likely neutralized by the acidic residues. The coordination geometry reported here could be useful in predicting proteins that utilize acidic residues to chelate a cluster of divalent cations such as Ca^{2+}, Mn^{2+}, and Mg^{2+}.

Stability Contributions of Calcium Binding

Both s3b and s3 are compact, and extremely stable in the presence of physiological Ca^{2+}. This may prolong enzyme activity against insoluble collagen in the extracellular matrix. Calcium

bound *holo*-s3b was previously shown to be more stable than *apo*-s3b based on DSC and fluorescence spectroscopy monitored equilibrium denaturation data (*29, 30*). Similar to the observation in s3b, denaturation pathways obtained using either DSC (Fig. 5) or fluorescence (Fig. 6) showed that *holo*-s3 is more stable than the *apo*-s3 (Supplementary Table 5). Stability data were compared with those of s3b. DSC data showed that *holo*-s3 (T_M= 94°C) was slightly more stable than *holo*-s3b (T_M= 93°C), though *apo*-s3 (T_M= 70°C) was less stable than *apo*-s3b (T_M=74°C) (Supplementary Table 5). Whether, heat, urea, or guanidine hydrochloride (GuHCl) was used to denature the protein, the same trends were observed. As a side note, the heat denaturation was reversible for *apo*-s3 but not for *holo*-s3. Based on fluorescence monitored thermal denaturation, neither *holo*-CBD fully unfolded at 100°C. Unfolding of s3 was monitored using the fluorescence of intrinsic Trp925 that is sandwiched between β-sheets. Whereas denaturation using urea or GuHCl on *holo*-s3 occurred as a cooperative transition from the native to the denatured protein, the unfolding transition of *apo*-s3 is abrupt (Fig. 6). *Apo*-s3 begins to unfold even at initial levels of chemical denaturant or heat. Meanwhile, the T_M value for *apo*-s3 based on fluorescence data is significantly lower than that determined by DSC (Supplementary Table 5). The discrepancy in T_M was also observed in s3b (*30*). Analysis of *holo*-s3 structure using NACCESS (*74*) demonstrated that Trp925 is solvent inaccessible. Also, crystal structures of *holo*-s3b exhibited tighter packing around the Trp than *apo*-s3b (*30*). The folding studies suggest that Trp in *apo*-s3 is much less tightly packed than *apo*-s3b.

As observed for s3b, a locally stable, solvent inaccessible core around Trp925 in *holo*-s3 could explain the higher T_M value observed using fluorescence data, while a locally unstable, solvent accessible core around Trp925 in *apo*-s3 could explain the lower T_M value observed using fluorescence data.

Evolutionary Related CBD

Both s3b and s3 molecules reported in this paper belong to the PPC super-family and are collagen-binding domains (CBD). Common structural features described in the previous sections enabled us to update the sequence alignment of the CBD in the M9B subfamily (Fig. 7). Conserved residues are important for one of four reasons: calcium chelation (red), *cis-trans* isomerization of the linker (yellow), collagen-binding (blue) or protein folding (green). Conserved residues in unsolved CBD structures will likely fulfill roles found in s3 and s3b. The dual calcium-binding site is formed by four chelating residues (Glu899, Glu901, Asn903, and Asp904) within the N-terminal linker, two chelating residues (Asp927 and Asp930) from the β-strand C and invariant Tyr1002 hydrogen-bonds and orients Asp930. Residue numbers used in this paragraph are of s3b. Likewise other supporting cast such as Gly921 is conserved in the middle of β-strand strategically placed to make room for Glu899. The dual calcium chelation site is fashioned sometimes by a neighboring residue. As mentioned, Asp897 of s3 acts equivalently to Asp927 of s3b. Asp897 equivalents are tentatively identified in *B. brevis* s3a and s3b, *C. botulinum* A3 s3a and *C. histolyticum* ColG s3a. Tridentate and divalent Asp and Glu residues are conserved with only *C. sordellii* s3a as the exception. The monodentate Asp904 residue is sometimes substituted by Asn. For those substituted, the net charge of the dual calcium site is neutral rather than -1.

A *cis*-peptide bond forms between residues 901-902 of *holo*-s3b and 870-871 of *holo*-s3 (Fig 2). The residue equivalent to Asn902 in other CBD molecules is either Pro, Asp or Asn. Pro frequently succeeds the peptide bond to ease *trans-cis* isomerization. In s3b, Asn902_OD1 and Asp904_N form a critical hydrogen-bond for the peptide isomerization (*37*). For CBD molecules where Asp succeeds the *cis* bond, side chain oxygen could play the same role as Asn902_OD1.

Other hydrogen-bonds identified by simulation studies important in stabilizing the transition states (*37*) are well conserved. The simulation study did not account for hydrogen-bonding via water mediated interactions involving Lys900_O, Asn902_O, Lys908_O, and Thr910_OG1 that are also conserved (Supplementary Table 6). Calcium ions could catalyze the isomerization in all the CBD molecules and their transition states and catalytic mechanism may look very similar.

Conserved residues are important in ensuring either proper folding or architectural stability. Hydrophobic residues packed between the β-sheets are better conserved if they are located in the vicinity of functionally critical residues. For example, invariant Trp956 of strand E is packed between the β-sheets. The residues flanking (Thr955 & Thr957) interact with mini-collagen. Tyr932 is packed between the sheets and helps positioning Tyr1002. Residues at tight turns are conserved as well. Gly975 is well conserved to allow a type II' turn in s3b (*29*). Gly942 (Gly975 equivalent) in s3 allows Asp941 side chain to stabilize the reverse turn. A highly conserved six-residue stretch, 986-991 (PGKYYL), adopts a tight turn and precedes the functionally important strand H. The region is well ordered in the crystal structures with low B-factors, and is the least dynamic based on NMR (*45*) and limited proteolysis MALDI-TOF MS (*64*). The main-chain carbonyl and amino groups of Arg985 hydrogen-bond with OH of Tyr989 to stabilize the turn. Only Gly987 can make room for the bulky Tyr989 side chain. Tyr990 packs against the invariant Ala909 and conserved 3_{10} helix. Ala909 is at the base of the linker that undergoes α-helix→β-strand transformation. The tight turn may ensure that collagen interacting Leu992, Tyr994, and Tyr996 would be correctly positioned. Tyr994 is the most critical residue in interacting with collagenous peptides (*29*). The strands adjacent to strand H, *i.e.* strands C and E, are very well conserved. The three antiparallel strands mold the collagen-binding pocket. Strand F staples the β-sheets by interacting with both sheets. The β-strand first interacts in an

antiparallel orientation with strand E then breaks its direction at Gly971 to interact with strand G. In place of Gly971, Ala or Pro is found at the location where the strand switches its allegiance. The dual interaction of the strand helps positioning Tyr970 to interact with mini-collagen.

Three residues (Tyr970, Tyr994, and Tyr996) shown to interact strongly with mini-collagen (*29, 45*) are conserved. Tyr996 of s3b is a critical residue in binding mini-collagen, and is partially conserved. This residue is replaced with Phe in s3, though both side chains have identical orientation. In other CBD molecules, an aromatic residue, such as Phe or His, is sometimes found at the site. Meanwhile, either a β-branched residue or Leu is found at the position equivalent to Thr957 in most of the CBDs. Six other residues identified by ^{15}N-HSQC-NMR titration to interact with mini-collagen are not well conserved. Since divergent CBDs (s3 and s3b) adopted a similar cleft-shaped binding pocket, other CBDs may also adopt a similar collagen-binding structure.

Divergent CBD could target different collagen sequences and could possibly target different collagen types; however, this structural study suggests otherwise. Rather, all the CBD domains may bind similarly to an under-twisted region such as the C-terminus of a collagen fibril. The C-terminus of type I collagen is exposed in the fibril surface based on X-ray fiber diffraction experiments (*5*), and it is the most accessible site for the bacterial collagenase to initiate assaults. However, CBD binding only at the C-terminal region of tropocollagen is unfounded. Gold particle-labeled tandem ColG-CBD (s3a-s3b) labeled with gold particle bound to type I collagen fibrils exhibited no periodicity (*44*). In the collagen fibrils, the molecules are staggered from each other by about 67nm. Therefore CBD could target partially under-twisted regions in the middle of a tropocollagen that are also vulnerable for assaults (*66*). Both type I and II tropocollagen molecules are suggested to consist of triple helix with some relaxations (*75*).

Energetic requirement to unwind these regions could be much less than tightly wound helical regions. Both s3 and s3b bind similarly to a mini-collagen, thus M9B collagenase molecules could initiate collagenolysis from analogous structural features in various collagen fibrils.

Some bacterial collagenases consist of up to three CBD molecules. The linker that induced structural transformation is a common feature found in M9B collagenase. It could act as Ca^{2+} sensor to trigger domain rearrangement as means of enzyme activation against the insoluble fibril. Physiological Ca^{2+} will likely trigger a domain rearrangement in these collagenases. Properly positioned tandem and triple CBDs should ensure collagenase to be anchored to the weakest region in fibril.

The function of the PPC domain in M9A (PPC-M9A) is unknown. Characterized M9A are collagenases and the PPC domain from *V. vulnificus* is thought to interact with type IV collagen (*76*). No structure of PPC-M9A has been reported. The crystal structure of PPC from serine protease S8 superfamily with a sequence identity of 15% (PDB accession code 1WME (*77*)) cannot offer additional insights either. The function of the PPC is not known, either, but PPC-M9A may adopt a similar tertiary fold to that of CBD. Like s3, strand A in PPC-M9A may not exist. The capability to monitor the concentration change in Ca^{2+} by CBD is critical for the activation of M9B enzymes against insoluble collagen. Along with Zn^{2+} at the catalytic center, *V. parahaemolyticus* M9A enzyme uses Ca^{2+} as a cofactor (*78*). However, neither Ca^{2+} binding residues, nor collagen binding residues identified in CBD are found in PPC-M9A. The structures of s3 and s3b could only serve as a template to the general fold of PPC-M9A.

As mentioned, clinical applications using fusion proteins of CBD and various growth factors are underway. Fusion proteins of any CBD derived from M9B collagense and a growth factor should result in comparable clinical outcome.

Accession Codes

The full atomic coordinates of the CBDs and their corresponding structure factor amplitudes have been deposited in the Protein Data Bank (PDB accession codes 4HPK, 3JQW, and 3JQX assigned to Ca^{2+}-bound s3b, Ca^{2+}-bound s3, and Ca^{2+}/Cd^{2+}-bound s3).

Tables

Table 1. Data collection and refinement statistics

	Holo–s3b	*Holo*-s3
Data collection statistics (a)		
Wavelength (Å)	0.98045	0.97937
Space group	P2$_1$	P2$_1$2$_1$2$_1$
a, b, c (Å)	40.9, 59.2, 48.8	62.0, 64.2, 95.4
beta(°)	100.4	90.0
Unique Reflections	44,856	26,207
Resolution (Å) (b)	16.61-1.35	28.50-2.00
Completeness (%) (c)	99.8 (100)	99.4(99.9)
I/σI (b)	26.3 (2.6)	16.7 (3.1)
R$_{meas}$ (%) (d)	6.4 (65.9)	9.9 (56.6)
Redundancies	4.6 (3.2)	4.0 (4.1)

Table 1 (Cont.): Data collection and refinement statistics

	Holo–s3b	*Holo*-s3
Refinement Statistics		
Resolution (Å)	16.61-1.35	28.50-2.00
R-factor (%)	17.9 (30.9)	17.4 (19.5)
R_{free} (%) 10% of data	21.0 (38.0)	23.3 (28.4)
Average B-factor main chain in A (A^2)	11.6	25.2
Average B-factor side chain in A (A^2)	14.2	27.5
Average B-factor main chain B (A^2)	15.8	21.4
Average B-factor side chain B (A^2)	17.8	23.8
Average B-factor main chain C (A^2)	NA	23.4
Average B-factor side chain C (A^2)	NA	26.2
RMS Deviations from restraint target values		
Bond Lengths (Å)	0.01	0.02
Angle (°)	1.2	1.7
Distances form Restraint Planes	0.004	0.007
Ramachandran Statistics		
Most favorable (%)	92.1	90.2
Additionally allowed (%)	7.9	9.8
Disallowed (%)	0.0	0.0

(a) For more complete data collection statistics, see Supplementary Tables 7 and 8.

(b) Highest resolution shell for *holo*-s3b: 1.40-1.35 Å. Highest resolution shell for *holo*-s3: 2.03-2.00 Å

(c) Data for the highest resolution shell are given in parenthesis

(d) $R_{meas} = \Sigma_h \sqrt{\frac{n_h}{n_{h-1}}} \Sigma_l |I_{hl} - \langle I_h \rangle| / \Sigma_h \Sigma_l \langle I_h \rangle$

Table 2: SAXS derived molecular parameters

Mini-collagen:CBD Complexes	Small Angle X-ray Scattering (SAXS)			Discrepancy
	D_{max} (Å)	R_g (Å) (expt)	R_g (Å) (calc)	χ^2
[(POG)$_{10}$]$_3$:s3	87	23.11 ± 0.09	24.43	0.47
[(POG)$_{10}$]$_3$:s3b	93	22.62 ± 0.04	24.79	0.75

Fig. 1: Ca^{2+} coordination in s3 and s3b. Comparison of Ca^{2+} coordination of s3 (A) and s3b (B): For both CBDs, the pentagonal base around Ca1 is indicated by gray dashes, while the axial positions are indicated by yellow dashes. Seven oxygen atoms from six residues interact with this Ca^{2+}. For Ca2 of both CBDs, one square face is indicated by blue dashes, while the other face is indicated by green dashes. Seven oxygen atoms from five residues and a water molecule interact with this Ca^{2+}. Each oxygen atom is labeled with residue type and residue number. D897 of s3 fulfills the role of D927 of s3b.

Fig. 2: Linker of CBD. (A) Close-up view of linker in s3 (ColH). The peptide bond between Glu870 and Pro871 adopts prolyl *cis* conformation. Glu868, Glu870, Asn872 and Asn873 in the linker region chelate with Ca^{2+} ions (brown spheres). (B) Close-up view of the linker in s3b (ColG). The peptide bond between Glu901 and Asn902 adopts an energetically unfavorable non-prolyl *cis* conformation. Glu899, Glu901, Asn903 and Asp904 in the linker region chelate with Ca^{2+} atoms (brown spheres).
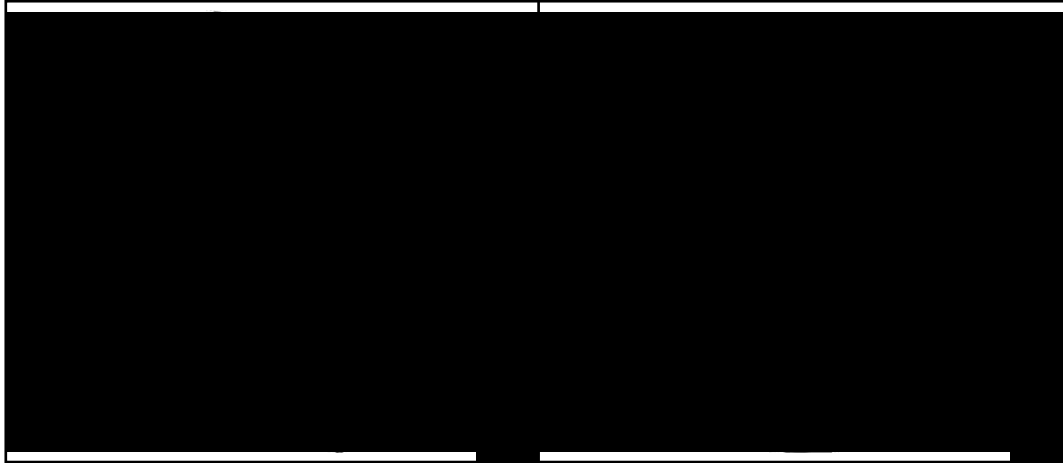
Fig. 3: Collagen-binding cleft of CBD. (A) Collagen-binding cleft in s3 (ColH). Conserved residues Thr924, Val926, Tyr937, Tyr962 and Phe964 may shape the binding cleft. Tyr900 adopts a favorable χ_1 and it is facing the binding cleft. (B) Collagen-binding cleft in s3b. Residues Ser928, Arg929, Thr957, Tyr970, Tyr994 and Tyr996 are the key residues shaping the binding cleft. The collagen-binding cleft matches the width of collagen (~10Å). Tyr931 adopts an atypical χ_1 angle.
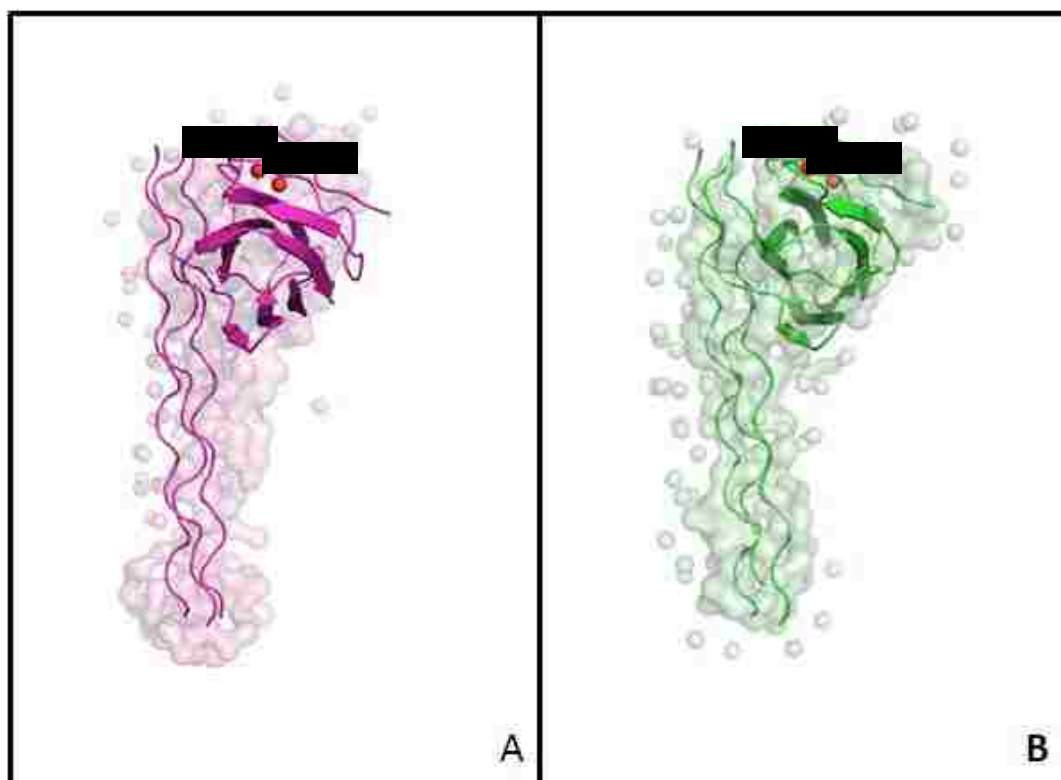
Fig. 4: Structures derived from SAXS scattering profiles using *ab initio* simulated annealing calculations are shown as gray surface for s3-[G(POG)$_8$]$_3$ complex (A) and s3b-[G(POG)$_7$PRG]$_3$ complex (B) Calcium ions are shown in red.
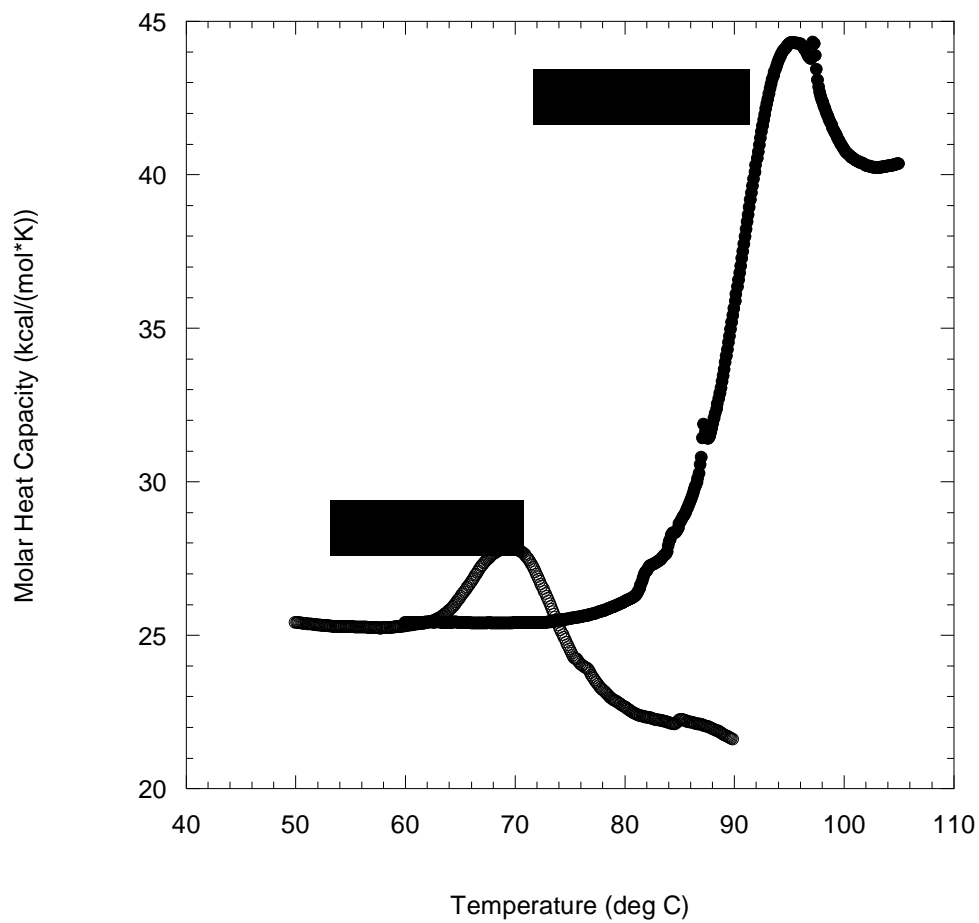
Fig. 5: DSC profile for both *apo* and *holo* s3
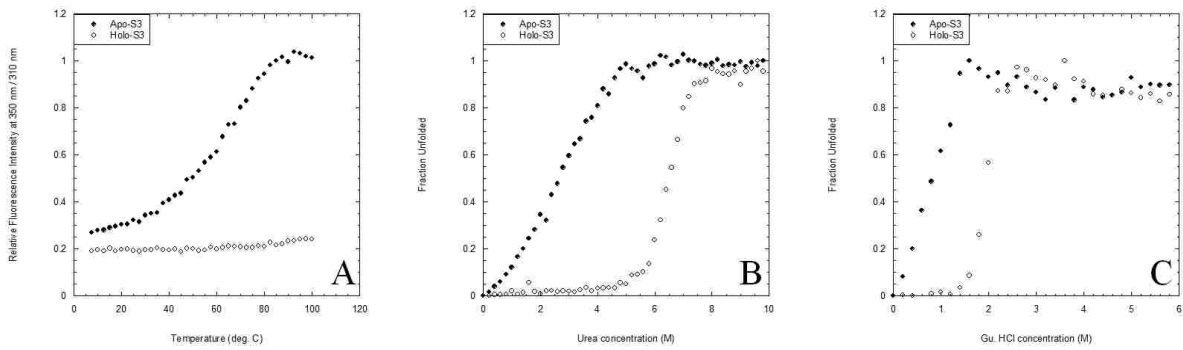
Fig. 6: Results of fluorescence measured equilibrium denaturation of s3 in its *apo* form (closed circles, and in its *holo* form (open circles). (A): Heat denaturation pathway of s3. (B): Urea denaturation pathway of s3. (C): GuHCl denaturation pathway of s3.
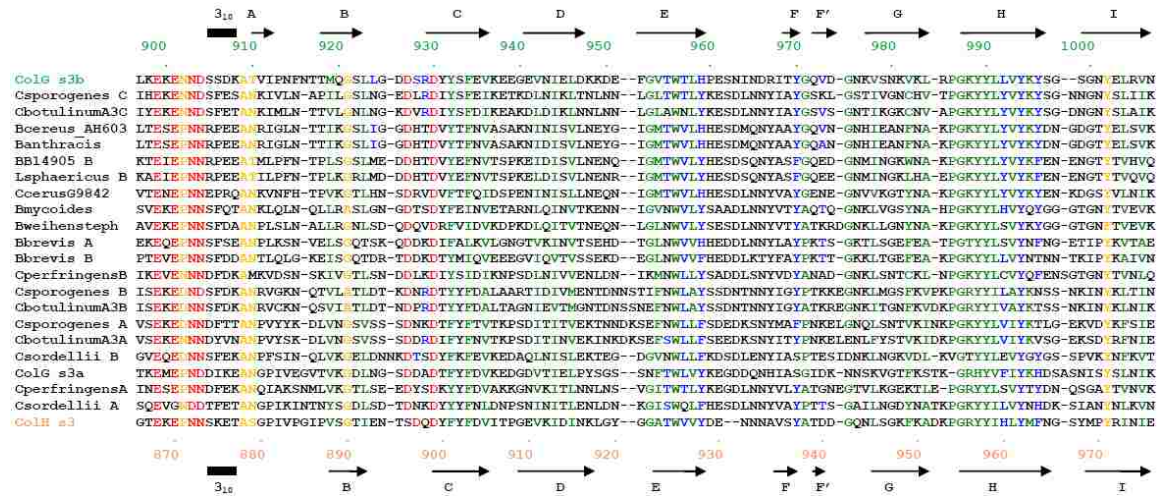
Fig. 7: Structure based sequence alignment of collagen-binding domains from M9B family. Calcium chelating residues, those critical for *cis-trans* linker isomerization, collagen-binding residues, and architecturally important residues are highlighted in red, yellow, blue and green, respectively. Sequences alignment was aided by the use of ClustalW2 (*79*).

Supplemental Material

Crystallization and Structure Determination of s3-Form II from ColH

The second form of s3 crystals were obtained by hanging drop vapor diffusion from 1.1-1.4 M sodium acetate, 0.1 M HEPES (pH 7.5), 1 mM $CaCl_2$ and 0.05 M $CdSO_4$. The crystals were flash frozen in liquid nitrogen without the need of additional cryoprotectant. This crystal form of s3 diffracted to 2.2Å resolution and was isomorphous to the first form. We were planning to solve s3 structure by single isomorphous replacement method. Ca^{2+} ions and solvent molecules were deleted from the ColH model described above before refinement using REFMAC (*52*). The size of F_o-F_c difference peaks indicated that 12 of them are of cations. Two cations per molecule were geometrically identical to Ca^{2+} bound to s3b and those of s3 in the form I. Six peaks were therefore assigned as Ca^{2+}. Later it became clear that the B-factor of one of the two Ca^{2+} was too low (<5 $Å^2$). It was therefore reassigned as Cd^{2+}. The rest of the strong F_o-F_c difference peaks were assigned as Cd^{2+} ions. Two Cd^{2+} ions are chelated with surface glutamic acids (Glu876 and Glu930), respectively in each of the three molecules. The main-chain (C_α) root mean square deviation between Form I and Form II was only 0.35Å.

Tables

Supplementary Table 1. Data collection and refinement statistics

	Holo-s3 Form II
Data collection statistics	
Wavelength (Å)	0.97937
Space group	$P2_12_12_1$
a (Å), b (Å), c (Å)	61.6,64.6, 96.0
beta(°)	90.0
Unique Reflections	19,933
Resolution (Å) (a)	28.7-2.2
Completeness (%) (b)	98.8(99.7)
I/σI (b)	9.6 (2.4)
R_{meas} (%) (c)	15.2 (54.0)
Redundancies	13.9 (3.4)

Supplementary Table 1 (Cont.). Data collection and refinement statistics

	Holo-s3 Form II
Refinement Statistics	
Resolution (Å)	20-2.2
R-factor (%)	17.6 (20.4)
R_{free} (%) 10% of data	25.0 (27.2)
Average B-factor main chain in A (A^2)	19.6
Average B-factor side chain in A (A^2)	21.7
Average B-factor main chain B (A^2)	18.3
Average B-factor side chain B (A^2)	20.5
Average B-factor main chain C (A^2)	19.4
Average B-factor side chain C (A^2)	22.0
RMS Deviations from restraint target values	
Bond Lengths (Å)	0.02
Angle (°)	1.9
Distances form Restraint Planes	0.008
Ramachandran Statistics	
Most favorable (%)	91.0
Additionally allowed (%)	9.0
Disallowed (%)	0.0

(a) Highest resolution shell for holo-s3 form II: 2.24-2.20
(b) Data for the highest resolution shell are given in parenthesis

(c) $R_{meas} = R_{meas} = \Sigma_h \sqrt{\frac{n_h}{n_{h-1}}} \Sigma_l |I_{hl} - \langle I_h \rangle| / \Sigma_h \Sigma_l \langle I_h \rangle$

Supplementary Table 2. Calcium-oxygen atom distances and deviation from planarity in pentagonal bipyramid coordination.

s3b	Residue	E901	E901	D904	R929	D930	N903, axial	D927, axial	
	Atom	OE1	OE2	OD2	O	OD1	OD1	OD1	Ca^{2+}
O-Ca^{2+} distances (Å) (a)	Average of Molecules A and B	2.56 (2)	2.41 (1)	2.34 (2)	2.36 (1)	2.35 (2)	2.35 (1)	2.43 (2)	3.75(1)
Distance out of equatorial LSQ plane (Å) (ab)	Average of Molecules A and B	0.29	0.32	0.21	0.19	0.15	NA	NA	0.05
s3-Form I	Residue	E870	E870	N873	Q898	D899	N872	D897	
	Atom	OE1	OE2	OD1	O	OD1	OD1	OD1	Ca^{2+}
O-Ca^{2+} distances (Å)	Average of Molecules A, B and C	2.56 (6)	2.45 (3)	2.34 (5)	2.39 (1)	2.38 (4)	2.34 (5)	2.28 (2)	3.77 (2)
Distance out of equatorial LSQ plane (Å) (ab)	Average of Molecules A, B and C	0.34 (5)	0.28 (9)	0.27 (12)	0.24 (8)	0.12 (6)	NA	NA	0.06 (4)
s3-Form II	Residue	E870	E870	N873	Q898	D899	N872	D897	
	Atom	OE1	OE2	OD1	O	OD1	OD1	OD1	Ca^{2+}
O-Ca^{2+} distances (Å)	Average of Molecules A, B and C	2.62 (4)	2.55(10)	2.37 (5)	2.35 (6)	2.40 (4)	2.34 (5)	2.30 (1)	3.66 (3)

A number within parenthesis represents an estimated standard deviation obtained after least square refinement (2.54 (2) means 2.54±0.02 Å). Measurements are for the distance of specific atom out of the plane created by the five members and do not include the axial members. Additionally, the distance of calcium from each plane was also calculated.

Supplementary Table 3: Calcium- oxygen atom distances and deviation from planarity in square antiprismatic coordination.

s3b	Residue Atom	E899 OE1	E899 OE2	E901 OE2	H$_2$O O	Ca^{2+}	S922 O	D927 OD2	D930 OD1	D930 OD2
O-Ca^{2+} distances (Å) (a) Distance out of square LSQ plane (Å) (ab)	Average of Molecules A and B	2.51 (2)	2.46 (1)	2.36 (2)	2.44	3.75 (1)	2.32 (2)	2.41 (2)	2.47 (1)	2.60 (2)
	Average of Molecules A and B	0.10	0.16	0.08	0.05	NA	0.02	0.02	0.03	0.03
s3-Form I	Residue Atom	E868 OE1	E868 OE2	E870 OE2	H$_2$O O	Ca^{2+}	T891 O	D897 OD2	D899 OD1	D899 OD2
O-Ca^{2+} distances (Å) Distance out of square LSQ plane (Å) (ab)	Average of Molecules A, B and C	2.64 (10)	2.46 (7)	2.26 (4)	2.41 (5)	3.77 (2)	2.29 (6)	2.50 (1)	2.54 (5)	2.52 (3)
	Average of Molecules A, B and C	0.05 (3)	0.05 (4)	0.03 (2)	0.04 (3)	NA	0.21 (11)	0.17 (7)	0.25 (10)	0.23 (11)
s3-Form II	Residue Atom	E868 OE1	E868 OE2	E870 OE2	H$_2$O O		T891 O	D897	D899 OD1	D899 OD2
O-Ca^{2+} distances (Å)	Average of Molecules A, B and C	2.57 (13)	2.49 (6)	2.36 (8)	2.78	3.66 (15)	2.25 (11)	2.34 (3)	2.41 (7)	2.41 (1)

The two planar faces are listed separately. Measurements are for the distance of specific atom out of the plane created the four members of each face. Again, a number within parenthesis represents the standard deviation obtained after least square refinement.

Supplementary Table 4. Deviation from planarity in square antiprisms found in proteins.

Protein	s3b	s3	2AAA	3TEC	2PRK	1JPQ	1K4C
Resolution (Å)	1.4	2.0	2.1	2.0	1.5	1.6	2.0
Ion	Ca^{2+}	Ca^{2+}	Ca^{2+}	Ca^{2+}	Ca^{2+}	K^+	K^+
Average planar deviation (Å)	0.08	0.01	0.14	0.12	0.01	0.02	0.0
Average ion-to-ligand distance (Å)	2.58	2.45	2.59	2.45	2.48	2.78	2.83

PDB access codes are for proteinase K (2PRK), alpha amylase (2AAA), thermitase (3TEC), G-quadruplex (1JPQ), and KcsA potassium channel (1K4C). The planes for 1K4C are defined by crystallographic symmetry.

Supplementary Table 5: Stability information of *apo* and *holo* s3

Parameters	Apo-s3			Holo-s3		
	Heat	Urea	GuHCl	Heat	Urea	GuHCl
ΔG_{HOH} (kcal*mol^{-1})	5.2	1.7	1.8	NA	9.6	8.1
m (kcal*mol-1*M^{-1})	0.1	0.6	2.3	NA	1.5	4.2
C_M (M)	NA	2.7	0.8	NA	6.5	1.9
T_M (°C)	70.2[a] 64.6[b]	NA	NA	94.1[a] >100[b]	NA	NA

(a) Value represents T_M for DSC data shown in Fig-5.

(b) Value represents T_M for denaturation curve shown in Fig-6A

Supplementary Table 6: Conserved water mediated hydrogen-bonds in s3 and s3b.

HOH mediated H-bonds in s3b	HOH mediated H bonds in s3
K908_O...HOH...T910_OG1	T877_O...HOH...S87_OG1
K900_O...HOH...T910_OG1	K869_O...HOH...S87_OG1
K908_O...HOH...N902_O	T877_O...HOH...P87_O
A909_O…HOH…K900_O	A878_O…HOH...K86_O

K908_O…HOH…T910_OG1 indicates that a water molecule is found between carbonyl oxygen of K908 and side-chain alcohol group of T910.

Figures
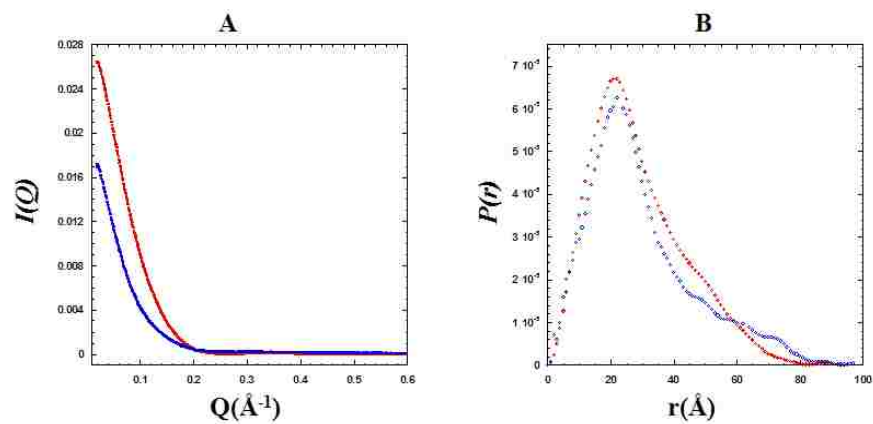
Supplementary Figure 1. SAXS data of s3-[G(POG)$_8$]$_3$ complex (blue) and s3b-[G(POG)$_7$PRG]$_3$ complex (red). (A) The scattering profile where the intensity, *I(Q)* is plotted against scattering vector Q. (B) Pair-distance distribution function *P(r)* in the real space.
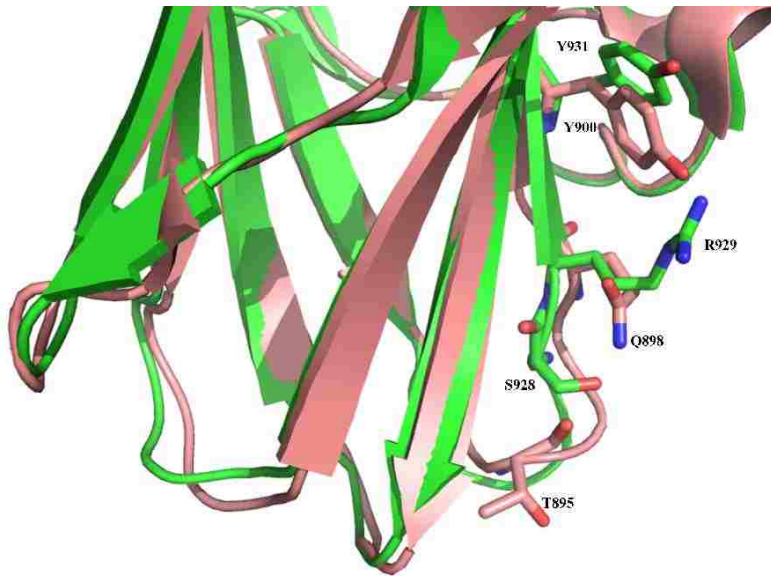
Supplementary Figure 2. Superimposed structure of s3 (salmon) and s3b (green), showing that Y900 of s3 occupies the three-dimensional space of R929 of s3b and that T895 of s3 occupies the three-dimensional space of S928 of s3b.
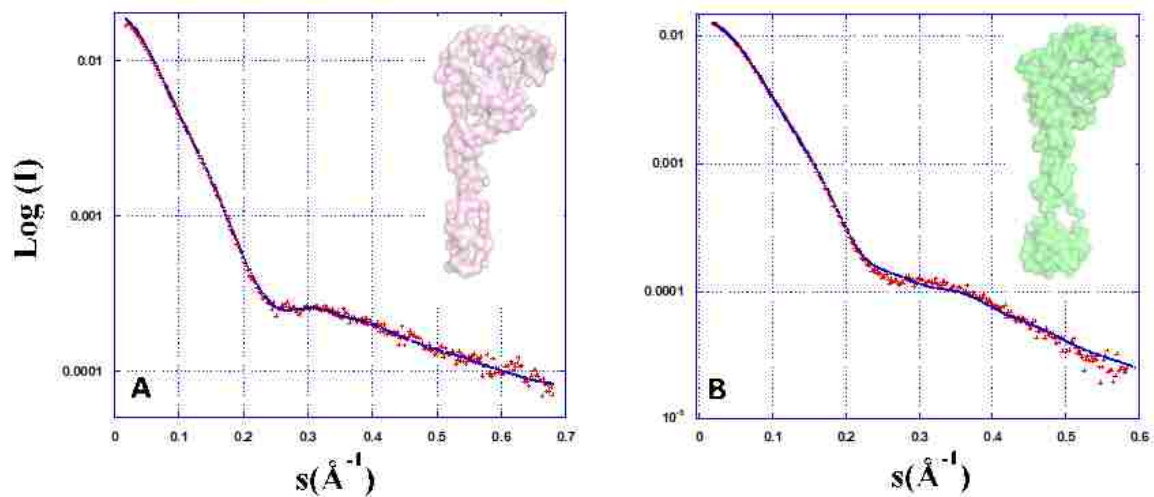
Supplementary Figure 3. SAXS curves of s3-[G(POG)$_8$]$_3$ complex (A) and s3b-[G(POG)$_7$PRG]$_3$ complex (B) are compared to the theoretical scattering durve. The theoretical Rg values calculated from the models by using CRYSOL (*61*)are in excellent agreement with the experiemental curves (Table-2). Rigid body model with lowest χ value for s3-[G(POG)$_8$]$_3$ complex (A) and s3b-[G(POG)$_7$PRG]$_3$ complex (B) are also shown.

Chapter 3: Structures of three polycystic kidney disease-like domains from *Clostridium histolyticum* collagenase ColG and ColH

Abstract

Clostridium histolyticum collagenases ColG and ColH are segmental enzymes that are thought to be activated by Ca^{2+}-triggered domain reorientation to cause extensive tissue destruction. The collagenases consist of a collagenase module (s1), a variable number of polycystic kidney disease-like (PKD-like) domains (s2a and s2b in ColH; s2 in ColG), and a variable number of collagen-binding domains (s3 in ColH; s3a and s3b in ColG). The X-ray crystal structures of Ca^{2+} bound *holo*-s2b (1.4 Å resolution, R = 15.0%, R-free = 19.1%) and *holo*-s2a (1.9 Å resolution, R = 16.3%, R-free = 20.7%), as well as Ca^{2+} absent *apo*-s2a (1.8 Å resolution, R = 20.7%, R-free = 27.2%) and two new forms of N-terminal truncated *apo*-s2 (1.4 Å resolution, R=16.9, R-free=21.2%; 1.6 Å resolution, R=16.2, R-free=19.2%) are reported. The structurally similar PKD-like domains resemble the V-set Ig fold. In addition to a conserved β-bulge, the PKD-like domains feature a second bulge that also changes the allegiance of the subsequent β-strand. This β-bulge and the genesis of a Ca^{2+} pocket in the archaea PKD-like domain suggest a close kinship between bacterial and archaeal PKD-like domains. Different surface properties and indications of different dynamics suggest unique roles for the PKD-like domains in ColG and in ColH. Surface aromatic residues found on ColH-s2a-s2b, but not on ColG-s2, may provide the weak interaction in the biphasic collagen-binding mode previously found in s2b-s3. B-factor analyses suggest that in the presence of Ca^{2+}, the midsection of s2 becomes more flexible, but the midsections of s2a and s2b stay rigid. The different surface properties and dynamics of the domains suggest that PKD-like domains of M9B bacterial collagenase can be grouped into either a ColG-subset or ColH-subset. Conserved properties of PKD-like domains in ColG and in ColH

concern Ca^{2+} binding. Conserved residues not only interact with Ca^{2+}, but also position the Ca^{2+} interacting water molecule. Ca^{2+} aligns the N-terminal linker approximately parallel to the domain's major axis. Ca^{2+} binding also increases stability against heat and guanidine hydrochloride, and may improve their longevity in ECM. The results of this study will further assist in developing collagen-targeting vehicles for various signal molecules.

Introduction

Clostridium histolyticum collagenases are causative agents for gas gangrene. The two classes of collagenase, ColG and ColH, differ in domain structures (s1, s2, s3a, s3b for ColG; s1, s2a, s2b, s3 for ColH) (*43*) (Fig. 1), and work synergistically to degrade collagen fibers (*32*). S1 is the collagenase module that belongs to the metallopeptidase subfamily M9B. The amino acid sequences of s2, s2a and s2b resemble the polycystic kidney disease domain (PKD) that was first identified in the polycystic kidney disease protein PKD1 (The International PKD Consortium, 1995). The C-terminal domains s3a, s3b and s3 are collagen-binding homologues that are a subclass of bacterial pre-peptidase C-terminal domains (PPC super-family) (*29, 45, 66, 80*). The collagen-binding segment composed of the PKD-like domain and CBD is not necessary to degrade gelatin (denatured, non-triple helical collagen) and acid solubilized collagen. The segment, however, is necessary to degrade insoluble collagen fibers.

Understanding the interaction of Ca^{2+} is significant due to its role in regulating both stability and enzyme activity in the ECM (*30, 80, 81*). Full-length ColH has been shown to undergo Ca^{2+} dependent structural changes demonstrated using SAXS and limited proteolysis (*81*). In ColG, Ca^{2+} triggers the linker between s3a and s3b to undergo secondary structure transformation from an α-helix to a β-strand to increase collagen affinity (*29, 64*). Similar to the N-terminal linker structure of s3b (*30*), the N-terminal linker structure of the PKD-like domain is

52

also thought to be Ca^{2+} dependent, and thus, high-resolution structures of both *apo* and *holo* states for the PKD-like domains are needed in order to elucidate their activation mechanism. Thus far crystallographic methods have been used to describe *apo*-s1 from ColG (*31*), the *holo*-peptidase sub-domains of ColH (*82*), *apo-* and *holo*-s2 (*31, 82*), *apo-* and *holo*-s3b of ColG (*29*), and *holo*-s3 of ColH (*80*). In the *apo*-s2 structure, however, the conserved Pro688 near the Ca^{2+} binding site was mutated to Thr. As a side note, we use amino acid sequence numbering for the mature enzymes. Numbering for s2b and s2a accounts for cleavage of a 40 amino acid prepro-peptide present in ColH, while numbering for s2 accounts for cleavage of a 110 amino acid prepro-peptide present in ColG. In this paper, we describe crystal structures of ColG, and for the first time, ColH PKD-like domains. Thermal and chemical stability differences upon Ca^{2+}-binding for the PKD-like domains are also reported.

The collagenolytic mechanism between mammalian matrix metalloproteinases (MMP) and bacterial collagenase differ (*18, 83*). Unlike bacterial collagenases, MMPs are sequence specific and proposed to actively unwind the triple-helix (*84-87*). Meanwhile, each domain in bacterial collagenase is believed to play a unique role in collagenolysis (*88*). The C-terminal CBD unidirectionally binds to under-twisted sites in the triple-helical collagenous peptide (*66, 80*). The CBD does not unwind mini-collagen, and hence, targeting under-twisted regions of tropocollagen may circumvent the energy barrier required for unwinding the helix. Various roles for the PKD-like domains have been proposed. The PKD-like domain has been shown to swell, but not unwind collagen fibrils (*39*). Clostridial PKD-like domains do not bind tightly to collagen fibril (*40, 43*), though s2b has been shown to enhance s3's ability to bind to the collagen fiber. S2b-s3 binding is biphasic; the initial low affinity ($K_d = 2.11 \times 10^{-6}$ M) leads into higher affinity ($K_d = 3.39 \times 10^{-7}$ M) (*40*). The N-terminal collagenase module, s1, has a two-domain

architecture that disbands the collagen microfibril into monomeric triple-helices, and then cleaves the exposed peptide bond preceding the Gly residue (*31, 82*). Crystallographic packing analysis of s2 suggested a side-by-side assembly of s1, s2, s3a and s3b that matched the width of the collagen microfibril (*31, 89*). The proposed *holo*-ColG structure is compact; s2 helps align the active site of s1 with the binding clefts of s3a and s3b. In contrast, the solution envelope of ColH resembled a tadpole (*81*), and thus, the role of its PKD-like domains of ColH could differ from that of ColG. The work presented here provides structural framework to better decipher the role of the PKD-like domain.

Despite their detrimental role in bacterial infection, bacterial collagenases and their collagen-binding segments are investigated for therapeutic applications. A cocktail of *C. histolyticum* ColG and ColH is used in both the nonsurgical treatment of Dupuytren's contracture (*38*), and the isolation of pancreatic islets (*90, 91*). Other applications are in preclinical stages (*18*). Moreover, fusion proteins consisting of growth factors, cytokines, or hormones and the collagen-binding segment s2b-s3 are non-diffusing and long lasting at wound sites (*46, 92-94*), and hence, the binding segments are being developed as drug-delivery vehicles. Therapeutic strategies based on these results are proposed to enhance efficacy by minimizing the quantity of signal molecules necessary and reducing side effects. In contrast, bone distribution of the fusion protein of parathyroid hormone with s3 only (PTH-s3) was efficacious in bone mineral density increase in osteoporotic models, though fusion proteins of PTH-s2b-s3 demonstrated little efficacy (*48-50, 95*). When applied to skin, PTH-s3 was efficacious in hair follicle regeneration in alopecia models (*47, 96*). This study of PKD-like domains is necessary for commercialization and optimization efforts for various drug candidates.

Methods

Expression and purification of s2a, s2b and s2

Expression and purification of each PKD-like domain as a glutathione *S*-transferase (GST) fusion protein was achieved using previously described methods (*43*).

^{15}N-HSQC NMR characterization of apo-s2

Though stably folded and monodispersed in solution, s2 with its N-terminal linker did not crystallize. ^{15}N enriched *apo*-s2 was made to measure the dynamics of the protein using NMR. ^{15}N enriched s2 was prepared as described (*97*). NMR experiments were performed at 298 ± 0.5 K on a Bruker 700 MHz spectrometer equipped with cryoprobe^{TM}. The concentration of the protein used was 0.1 mM in 50 mM Tris-HCl (pH 7.5). In the HSQC spectra, thirteen residues could not be identified due to band broadening (Supplemental Fig. 1). Using the homology modeled s2 (based on PDB ID 2C4X (*98*)), we reasoned that unobserved HSQC peaks corresponded to a highly dynamic N-terminus that hindered crystallization. Guided by the solution data, thirteen residues were truncated from the N-terminus. The truncated s2 crystallized readily.

Crystal structure determination of PKD-like domains

Initial conditions suitable to grow *apo*-s2a, *holo*-s2b, and the two forms of *apo*-s2 crystals were identified by the sitting drop method using a high throughput screen (Hampton Research Crystal Screen HT). Subsequent crystallization trials using the initial conditions were carried out using the hanging drop method. *Apo*-s2a (at a concentration of 30.5 mg/mL) was crystallized from 3 M $(NH_4)_2SO_4$, 0.1 M MES (pH 4.5), and 15% (w/v) PEG 4000 at 290 K, whereas *holo*-s2a (4.8 mg/mL) was crystallized from 30% (w/v) PEG 4000, 0.2 M $MgCl_2$, and 0.1 M Tris-HCl (pH 8.5) at 17°C. These crystals were subsequently soaked in 15% (w/v) PEG

4000, 0.2 M MgCl$_2$, 50 mM CaCl$_2$ and 0.1 M Tris-HCl (pH 8.5) before initial unit cell characterization and data collection. Meanwhile, *holo*-s2b (13.7 mg/mL) was crystallized from 35% (w/v) PEG 5000, 0.2 M ammonium sulfate, and 0.1 M MES (pH 6.5) at 277 K. Both crystal forms of s2 (12.0 mg/mL) were grown from 41-49% 2-methyl-2, 4-pentanediol (MPD), 100 mM Bis-Tris (pH 5.5) and 0.1-0.3M ammonium acetate at 277 K. The in-house X-ray diffraction facility (Rigaku 007, Osmic Blue confocal mirrors, Saturn CCD detector) was used for initial characterization of each PKD-like domain crystal, and in the case of the s2a crystals, was also used for data collection.at 113 K. s2b and s2 crystals were cryo-cooled and subsequently stored in liquid nitrogen until data collection. Diffraction data were collected at 109 K, in the 19-ID beamline of the Advanced Photon Source at Argonne National Laboratory.

Each s2a data set was indexed and scaled using d*trek (*99*) whereas each s2b and s2 data set was indexed and scaled using HKL-3000 (*53*). In each case, a data set truncated to 3 Å was used for molecular replacement using PHASER (*51*). The PKD-like domain from the carbohydrate-binding module (PDB code 2C4X) was used as the search model during structure determination of s2. s2 Form I was subsequently used as the search model during structure determination of s2a and s2b. Four molecules were found in the asymmetric unit (ASU) of *apo*-s2a crystals, while eight molecules were found in the ASU of *holo*-s2a crystals. Meanwhile, two molecules were found in the ASU of *holo*-s2b and each form of *apo*-s2 crystals.

The subsequent structure determination for each model was accomplished using an iterative process of manual adjustments aided by the use of MIFit (*55*) and refinement using REFMAC(*52*). During manual adjustments, ARP/wARP (*100*) was used to place water molecules. R$_{free}$ was lowered for s2 Form I, *apo*-s2a, and *holo*-s2a models by utilizing Babinet's principle for bulk solvent scaling. In each s2a model, R$_{free}$ was also lowered by applying TLS

and tight NCS restraints. The s2a models were refined with isotropic B-factors, whereas s2-Form I and s2b models were refined with anisotropic B-factors. PARVATI (*101*) calculated the anisotropy of s2b and s2 to be 0.5 ± 0.1 and 0.4 ± 0.1 respectively. Isotropic B-factors would result in anisotropy of 1.0. Each model exhibited excellent geometry as analyzed by MolProbity (*102*). Full data collection and refinement statistics are summarized in Table 1 for s2a, s2b, and s2 Form I and in Supplemental Table 1 for s2-Form II. Alternate conformations are detailed in Supplemental Table 2.

Fluorescence spectroscopy measured equilibrium denaturation of PKD-like domains

PKD-like domains share similar topology, and unfolding of each domain was monitored by the change in intrinsic fluorescence of a conserved Trp residue. All experiments were carried out at room temperature on a Hitachi F-2500 fluorimeter with excitation and emission bandwidths at 2.5 nm and 10 nm, respectively. The excitation wavelength used was 280 nm, and fluorescence emissions were monitored between 300 nm and 450 nm. For s2b, s2a, and s2, λ_{max} for the folded state occurred at 325, 328, and 327 nm, respectively. For each domain, λ_{max} for the denatured states occurred at 350 nm. The ratio of intensity at 350 nm verses the intensity at the native state λ_{max} was used to track the unfolding process. During thermal denaturation trials, the temperature of the protein solution was maintained with a Neslab RTE-110 circulating water bath (Thermo Scientific, Newington, NH). In the thermal denaturation trials for s2b and s2, the protein concentration was 3 μM. In the chemical denaturation trials, the protein concentration was 1.5 μM. In case of s2a, protein concentration for both thermal and chemical denaturation was 5μM. Each *holo*-PKD-like domain was supplemented with 2 mM $CaCl_2$, while each *apo*-PKD-like domains was supplemented with 2 mM EDTA. In all cases, the protein was diluted in 10 mM Tris-HCl (pH 7.5) and 100 mM NaCl. When heat was used as the denaturant, each

domain was exposed to temperatures that linearly elevated from 280.5 to 363 K in 2.5 K increments. When guanidine hydrochloride (Gu.HCl) was used as the denaturant, each domain was exposed to concentrations of denaturant that increased linearly from 0.0 to 5.8 M in 0.2 M increments. ΔG_{HOH}, C_M, and m values were calculated as described previously (*30, 80*).

Results and Discussion

The X-ray crystal structures of Ca^{2+} bound *holo*-s2a (C2), Ca^{2+} absent *apo*-s2a (P6$_1$), Ca^{2+} bound *holo*-s2b (P2$_1$), and two new forms of N-terminal truncated wild-type *apo*-s2 (P2$_1$, P2$_1$2$_1$2$_1$) are reported for the first time. Between the novel s2a and s2b structures, s2b was solved at higher resolution, and correspondingly, is described in the most detail. New insights into s2 are subsequently reported.

Overall structure descriptions of apo and holo-s2a

In the following discussion, *holo*-s2a will be described first (Fig. 2A). The eight *holo*-s2a molecules are virtually identical (average RMSD = 0.2 Å ± 0.1 Å). Here, the molecules spiral along the crystallographic (1, 0, 1) axis. Along this axis, molecule pairs A and G, B and E, C and F, and D and H are related by NCS translation that results in an off-origin peak that is 63.9% of the origin peak in the Patterson map.

Similar to the molecules in the *holo*-s2a crystal, the four *apo*-s2a molecules are similar (average RMSD = 0.5 Å ± 0.2 Å) with molecules C and D being the most similar (RMSD = 0.2 Å), and molecules A and D, being the most divergent (RMSD = 0.8 Å). Molecules A and B, as well as molecules C and D, are non-crystallographic two-fold related. Temperature factors for each structure are relatively high (Table 1), possibly as a consequence of the high solvent content in the crystal (61.8%).

The *holo* and *apo* structures resemble each other (RMSD = 0.6 Å ± 0.2 Å). As expected, the most notable difference between the structures occurs near the N-terminus, where Ca^{2+} reorients interacting residues Asn685 and Ser686. While neither structure could be refined using anisotropic B-factors, comparison of B-factors revealed that with the exception of N-terminal residues, no significant change in B-factor upon Ca^{2+} binding occurred.

Unlike s2b or s2, s2a truncates β-strand A through an approximately 126° rotation of the ψ bond of Ile692. To accommodate the change, Tyr696 packs with Phe706 and is involved in a Δ4 Tyr corner, in which the side chain hydroxyl group hydrogen bonds with the backbone four residues prior to it (*103*). Interestingly, the Tyr corner also stabilizes the non-prolyl *cis*-peptide bond between Gly694 and Thr695 of s2a that forms the bulge that realigns β-strand A' to interact with β-strand G (Fig. 2A and 2B). A second bulge between β-strands B and B' is stabilized by a hydrogen bonding network that features a monodentate interaction between Asn735 and Ser708.

Overall structure description of holo-s2b

Similar to *holo*-s2a, the two NCS related *holo*-s2b structures are virtually identical (RMSD = 0.4 Å). In the structures, one Ca^{2+} was found near the N-terminal linker (Fig. 2C). Each *holo*-s2b structure begins from residue 766 and the last residue is 860. The PKD-like domain resembles a V-set Ig fold that lacks strand D (β-strand C' of the Ig fold correspond to β-strand D of the PKD-like domain fold). β-Strands, B and F in the PKD-like domain structures are shorter than the corresponding β-strands of a prototypical V-set Ig fold (PDB ID 1BRE (*104*)), while strands F and G in the PKD-like domains are longer than the corresponding β-strands in the V-set Ig fold. In the PKD-like domain structures, β-strands A, A', B, B', and E form one sheet, while strands C, D, F, and G form the opposing sheet. β-strands A' and G form a parallel β-sheet, while the remaining strands form anti-parallel β-sheets. Meanwhile, β-strand B forms a

sheet with β-strand E, with the exception of Tyr796 which is aligned with β-strand A. Given the β-sheet sandwich fold, the PKD-like domain was predicted to resemble the CBD (*41*), though structural alignment of *holo*-s2b with *holo*-s3 (PDB ID 3JQW (*80*)) suggests little homology.

Two prominent features are the conserved bulges in the domain that interrupt β-strands A and B and help to form a ridge along the ABE face (Fig. 2). The first bulge occurs when Pro784 breaks β-strand A and pushes the subsequent Lys785 outward, which in turn leads to an approximately 127° angle between β-strands A and A' that also changes the allegiance of β-strand A' to β-strand G. The second conserved bulge is introduced by Lys798 and Gly799. This bulge removes the backbone hydrogen bonding partner of Tyr780 of β-strand A. To help stabilize the bulge, the side chain hydroxyl group of conserved Thr800 hydrogen bonds with the backbone amide of the Tyr780. To further stabilize the bulge and position the hydroxyl group of Thr800, the carbonyl oxygen of Gly797 hydrogen bonds with the side chain hydroxyl group and main chain amide of Thr800. The bulge is also stabilized by conserved Asn825 hydrogen bonding with the amides of Gly797 and Lys798 (Fig. 2C). The second bulge helps form a prominent ridge. Surface exposed aromatic residues are found at both sides of the ridge and are discussed later.

Temperature factors for both NCS related structures are low (average B-factor for molecule A = 11.7 Å^2; that of molecule B = 9.9 Å^2). Anisotropic B-factor analyses using the Anisotropic Network Model web server (ANM) (*105*) showed that the main-chain is mostly isotropic and potential correlated movement occurred exclusively at the N-terminal linker. The calcium coordination for both *holo*-s2b and *holo*-s2a is described in detail in a later section.

Structure descriptions of apo-s2

Despite being solved in two crystal forms, the the crystal structures of Forms I and II of *apo*-s2 from ColG are similar. Each ASU contains two non-crystallographic two-fold symmetry related molecules. All four molecules are virtually identical to each other (RMSD < 0.5Å). Each structure begins from residue 685, though the first two residues (Gly-Ser) are remnants from GST-tag cleavage. The last residue is either 770 or 771. The NCS related molecules are stabilized by anti-parallel type inter-molecular interaction between β-strand A. Comparison of our *apo*-s2 structures to the previously solved *apo*-s2 (PDB code 2Y72), in which conserved Pro688 is mutated showed that the N-terminal mutation pushes the N-terminus out by 3 Å at the C_α of Ala687. Furthermore, while residues that make up the previously described bulges are conserved, the interaction pattern does slightly differ from the pattern found in s2a and s2b. Here, the hydroxyl group of Ser707 mediates the interaction between conserved Asn735 and the backbone amides of Gly708 and Lys709 (Fig. 2D).

Structure based sequence comparisons of PKD-like domains

The sequence comparison of divergent PKD-like domains revealed conserved residues that are essential for the overall fold and Ca^{2+} chelation. Residues involved in Ca^{2+} chelation will be discussed in the subsequent section. Peptidase M9 family members are all thought to be collagenase and consists of subfamily A (Vibrio) and B (Bacillus and Clostridium). M9A enzymes lack CBD, and consequently may utilize different approaches to collagenolysis. Therefore, this section will discuss M9B derived PKD-like domains exclusively, and will utilize s2b numbering. PKD-like domain found in M9B enzymes share two conserved clusters of residues (Fig. 3). The first conserved stretch, [802]DxDGxIxxYxWDFGDG[817], contains β-strand C (Underlined residues adopt the β-strand). The stretch is conserved for its Ca^{2+}-binding and architectural importance. The first two Asp residues chelate with Ca^{2+}. Invariant Tyr810 is

accommodated by the second β-bulge. Asp813 is sometimes replaced with an acidic Glu residue. The side-chain of Glu may easily fulfill the role of the side chain of Asp813, which terminates β-strand C and stabilizes the subsequent sharp turn by hydrogen bonding to the amide of Gly815. Phe814 stacks against conserved His828 so that the imidazole ring can also form a salt-bridge with conserved Asp816. Gly817 allows Asp816 to also stabilize the turn by hydrogen bonding to the amide of Ser818. Though a sharp turn follows β-strand C in all PKD-like domains, the type of turn is different. In s2b, the insertion of Asp819 results in the region adopting an ω-loop (i→i+10). Both s2a and s2 lack Asp819 in this stretch, and subsequently, each is involved in an α-turn (i→i+4) that forms a β-hairpin. The second conserved stretch, ^{825}NPxHxYxxxGxYxVxLxVxDxxG^{847}, forms β-strands E and F. Tyr830 and Tyr836 stack against each other to stabilize the interactions between the sheets. Tyr836 further stabilizes the sheets by forming a Δ4 Tyr corner. Asp844 is responsible for one of the axial interactions with Ca^{2+}. The β-strands also wrap around the conserved turn and β-strand C to form the most unique feature of the PKD-like domain. Gly847 allows for β-strands F and G to be separated by a β-turn (i→i+3).

Ca^{2+} chelation in s2a and s2b

Ca^{2+} coordination in s2a, s2b, and s2 are virtually identical to each other. Since *holo*-s2 has been described, this section describes the binding sites on s2a and s2b in detail. One calcium-binding site, consisting of pentagonal bipyramidal geometry, was identified near the N-terminus in each domain (Fig. 4). The pentagonal base is composed of OD1 of Asn685, the main chain carbonyl of Ser686, OD1 and OD2 of Asp713, and a water molecule, while the axial positions are filled by OD1 of Asp715 and OD2 of Asp754. Ca-O bond distances and planar deviations amongst oxygen atoms responsible for the pentagonal base (Table 2) are similar to values found

in clostridial CBDs (*80*). The Ca^{2+} coordination geometry in PKD-like domains has been described as octahedral (*89, 98*), though our results demonstrate that water is involved in forming a pentagonal base. The coordinating water molecule is positioned by OD1 of Asp755 (Fig. 4A). Calcium coordination in s2 (4AQO) (*82*) resembles that of s2a. Both the water and calcium ion are ordered (B-factor < 8.1 Å^2). Based on the sequence alignment (Fig. 3) of PKD-like domains, residues contributing side chain interactions with calcium are strictly conserved. s2b chelates with a Ca^{2+} ion very similarly, except for the residues that position the water molecule (Fig. 4B). s2b utilizes OG from both Ser845 and Ser846 in lieu of Asp755 in s2a.

Ca^{2+} induced structural change

Ca^{2+} chelation appears to align the N-terminal linker approximately parallel to the domain's major axis (Supplemental Fig. 2). In s2b, Ca^{2+} chelation by Asn774 and Lys775 could stabilize a 3_{10} helix (residues 770-774) that aligns with the cylinder axis. In s2 and s2a, N-terminal residues are positioned so that the N-terminal linker could also be positioned parallel to the domain's major axis. The residues prior to Asn685 cannot be observed in the electron density, and consequently, the region's secondary structure remains ambiguous. Structural comparison of the Ca^{2+}-binding site between the *apo-* and *holo*-PKD-like domains revealed that Ca^{2+} has varied influence on the loop between β-strands B' and C. Proline, positioned between aspartates equivalent to Asp802 and Asp804, restricts the loop flexibility so that minimal change occurs between *apo* and *holo* states. However, in s2, which does not have a proline positioned between Asp713 and Asp715, the Ca^{2+} interaction moves Asp715 2 Å from the binding site (Supplemental Fig. 2C).

Overall, average B-factors of s2b and s2 are low, and average B-factors amongst *apo-* and *holo*-s2a are similar (Table 1). Despite different symmetry interactions, the B-factor per-

63

residue trend of 18 PKD-like domains (eight *holo*-s2a molecules, four *apo*-s2a molecules, two *holo*-s2b molecules, and four *apo*-s2 molecules) are very similar to each other (Supplemental Figure 3A-D). *Holo*-s2, however, does not follow this trend (Supplemental Figure 3E). Comparison of *holo* and *apo* structures revealed a stark contrast between s2a and s2 in the influence of Ca^{2+} (Fig. 5A and 5C). Drops in C_{α} B-factor in the *holo*-s2 are found in three stretches (Gly698-I704, Gly708-Tyr721, and His738-Thr761) that are immediately preceded by stretches where the B-factor is higher (Lys691-Thr697, Glu705-Ser707, and Gly733-Val737). In the *holo*-s2 structure, the mid-section became more flexible, though both terminal regions became more rigid. Differences in crystal packing could account for the dynamics reversal, though it is possible that the crystal packing is a consequence of dynamic changes. Both termini of *holo*-s2 are pinned down by symmetry related molecules that could suppress their dynamics, while the mid-section of the barrel lacks the inter-molecular interactions observed in the *apo*-state. In *apo*-s2, intermolecular anti-parallel β-sheet interactions involving β-strand A could suppress dynamics of the region.

In contrast, comparison of *holo*-s2a and *apo*-s2a structures revealed that Ca^{2+} did not increase the B-factor of mid-section (Fig. 5A). In the s2a structures, the termini are the most flexible. As mentioned, B-factor trends in all *holo*- and *apo*-s2a structures are similar despite the difference in crystal packing. C_{α} B-factors for the mid-section of *holo*-s2b are low (Fig. 5B), and suggest that s2b behaves similarly to s2a. Overall, the starkly contrasting dynamics between ColG and ColH derived PKD-like domains suggest diverging roles during collagenolysis.

Ca^{2+} induced stability gain of PKD-like domains

The *apo*-state of s2a, s2b, and s2 are thermally stable proteins (T_M ~373 K), but they gain further stability in the presence of Ca^{2+}. In s2b, the fluorescence of Trp812 was monitored, while

in s2a and s2, the fluorescence of Trp723 was monitored. During fluorescence monitored thermal denaturation, none of the PKD-like domains fully unfolded in the *holo* state (Fig. 6A-C). Such hyper-thermostability was also observed for *holo* states of s3 (*80*) and s3b (*30*). The stability of both PKD-like domains and CBD may allow prolonged collagenolytic activity in the ECM. Heat is thought to denature proteins by disrupting electrostatic interactions. As such, the conserved hydrogen bonding network around the bulges may play a strong role in overall stability of the domains, while the Ca^{2+}-O interactions may contribute to increased stability in the *holo*-state. PKD-like domains consist of a conserved (shown in green in Fig. 3), well-packed core, and are likewise stable against Gu.HCl denaturation. Here, denaturation occurs through a cooperative transition from the folded state to unfolded states (Fig. 6D-F). In contrast to heat, Gu.HCl is thought to denature protein by predominately disrupting hydrophobic interactions (*106*). Of the three domains, s2b is the most stable (Table 3). The difference in ΔG_{H2O} between *apo* and *holo* states ($\Delta\Delta G_{H2O}$) is approximately the same for all PKD-like domains. In addition to reorienting the N-terminal linker, the proposed Ca^{2+}-induced helical base of the N-terminal linker may have a partial capping effect that shields the core against Gu.HCl. It is also well documented that metalloproteins are more stable in the presence of their metal ligand (*107*).

In the clostridial collagen-binding domain, Ca^{2+}-induced stability could be partially accounted for by a reduction in void volume and an increase in hydrogen bonds (*30*). Analysis of void volume of PKD-like domains using CASTp (*108*) revealed a common cavity located near the N-terminus shrinks. The common cavity located near the C-terminus of the *holo* and *apo* pairs of both s2a and s2 curiously remains essentially unchanged upon Ca^{2+} binding. Furthermore, Ca^{2+} binding does not lead to a significant, change in hydrogen bond totals in any of the domains (Supplemental Table 3). For therapeutic applications, *in vitro* stability of s2b and

s3 may explain the prolonged activity of growth factors and signal molecules when fused to s2b-s3 *in vivo*.

Surface Characteristics of ColG and ColH PKD-like domains

s2a and s2b, unlike s2, respectively contain two and four surface aromatic residues that are located on the ABE face (Fig. 7A and 7B). Interestingly, these residues are also located along the previously mentioned ridge. These residues could be involved in collagen binding given that aromatic residues are found at the hot-spot of the collagen-binding pocket in CBD. In s3b, mutations of Tyr970, Ty994, and Tyr996 to Ala greatly reduced binding to collagenous peptide as monitored by surface plasmon resonance (*29*). NMR studies also showed that these aromatic residues are involved in collagen binding (*45*). The structure-based sequence alignment of PKD-like domains (Fig. 3), suggest that the PKD-like domain of collagenases consisting of only one CBD will likely contain surface aromatic residues. Conversely, the PKD-like domain of collagenases consisting of multiple CBDs, such as s2, appears to have no surface aromatic residues. Collagenases from *B. brevis*, *C. botulinum*, and *C perfringens* contain multiple CBDs. Their respective PKD-like domains lack aromatic residues, and hence, may not directly interact with collagen.

A putative structure of *holo*-ColH can be built from the homology modeled activator domain of s1 and helical linker (residues 7-301 based on PDB ID 2Y50), and the crystal structures of the peptidase domain (residues 302-681, PDB ID 4AR1), s2a (residues 685-770), s2b (residues 766-860), and s3 (residues 861-981, PDB ID 3JQW). The overall dimensions (length = 133 Å, height = 36 Å, width = 88 Å) match the tadpole shape observed in the SAXS envelope of *holo*-ColH (*81*). In the model, the five-residue overlap between s2a and s2b structures was superimposed (RMSD = 1.0 ± 0.1 Å) to assist with formation of the s2a-s2b

segment. The aromatic residues mentioned are found on one side of s2a-s2b (Fig. 7C). In this model, the surface aromatic residues on s2a-s2b may either span across multiple tropocollagen molecules on the surface of the fibril or bind along one tropocollagen molecule when the binding surface of s3 is docked onto the collagen fibril surface. The interactions may serve to prevent the collagen-binding segment from diffusing away after the s3-collagen interaction is transiently broken. Likewise, the domains may provide loose contacts with the collagen fibril that allow the enzyme to scan the fibril surface for optimal regions for tight CBD interaction. In these roles, the PKD-like domain strengthens collagen avidity so that only one CBD is required for collagen binding. The zinc ion involved in activation of a water molecule is approximately 115 Å away from Tyr962 found in the collagen-binding pocket of s3. In this model, the PKD-like domains may also be critical in positioning the catalytic domain with respect to CBD.

Potential role of PKD-like domains in synergistic collagenolysis

The apparent differences between the ColG derived PKD-like domain and the ColH derived PKD-like domains may aid synergistic collagenolysis. The putative *holo*-ColH structure and the structure-based insights on the PKD-like domains allow us to begin to speculate on how ColG and ColH work together to degrade collagen. Currently, it is not known whether any of the clostridial PKD-like domains swell collagen fibers. Both s3 and s3b share a common preference for under-twisted regions of collagen (*80*), though ColG and ColH initially cleave different sites in collagen (*17*). When digesting the insoluble fiber, ColH is the workhorse (*32*). The higher collagen affinity observed for s2b-s3 may increase further by the addition of s2a. The increased affinity could hold ColH close to the collagen fibril so that it can slide along the fibril and find vulnerable regions. Meanwhile, the proposed structure of *holo*-ColG adopts a compact structure in which domains of the collagen-bonding segment are aligned by intermolecular β-sheet type

hydrogen bond interactions (*31*). ColG's tandem CBDs may allow the enzyme to anchor itself to the most vulnerable region of the fibril. In this context, the spring-like dynamics of s2 may allow it to swell the fibril. The swelled fibril would then expose the interior of the fibril and expose new sites for ColH collagenolysis.

PKD evolution

Human PKD1 (1B4R (*109*)) and PKD-like domains from archaea and bacteria share a high degree of structural similarity that suggests the fold laterally transferred across the kingdoms. As expected, s2a, s2b, and s2 resemble the *C. thermocellum* endoglucanase PKD-like domain (2C4X) more closely than either the archaeal surface protein PKD-like domain (1L0Q (*110*)) or human PKD1. While the bulge between β-strands A and A' appears to be well conserved only in bacterial PKD-like domains, the bulge between β-strands B and B' is also conserved in the archaeal PKD-like domain, but is not conserved in PKD1. Correspondingly, residues Thr800 and Asn825, which are critical for stabilizing this bulge, are conserved in the archaeal PKD-like domain. Oddly, only Thr800 is conserved in the endoglucanase PKD-like domain. Normally, surface interactions are not well conserved, but surprisingly, the salt bridge formed between Asp816 and His828 in s2b is found in the archaeal PKD-like domain. Structurally equivalent residues in the endoglucanase PKD-like domain, Asp47 and Tyr60, utilize hydrogen bonding between OD1 and OD2 of Asp and OH of Tyr, in lieu of the salt bridge. In the domains, these interactions serve to stack the β-sheets together and stabilize the region where β-strand D of the Ig fold is deleted. The interaction is not found between the equivalent Asp and His residues in the NMR structure of human PKD1, though it should be noted that the NMR structure is derived from main-chain NOE restraints, and therefore the side-chain orientations are not experimentally obtained. Thus, these residues may also assist in the

interaction of β-strands C and E. Within the core, Trp812 and Phe795 are conserved throughout the three kingdoms. Our PKD-like domain structures suggest this Phe strengthen the interactions of β-strands B and C through packing with the strictly conserved Trp and Phe in strand D. The residue further supports the barrel architecture through hydrophobic packing with the C-terminal region of the barrel.

Comparison of bacterial *holo*-PKD-like domains with either archaea or mammalian PKD suggests that the Ca^{2+} binding site in bacteria evolved from archaea. In addition to the overall structural similarity, five out of the seven oxygen atoms that coordinate to Ca^{2+} are present in the archaeal PKD-like domain. The archaeal domain lacks the initial asparagine residue and one of the axial aspartate residues required for Ca^{2+} binding. Ca^{2+} interacting residues Asn774 and Lys775 of s2b are replaced by Pro302 and Val303 in archaea. In addition to removing an oxygen atom responsible for the pentagonal base, Pro appears to constrain Val303_O to the position occupied by Ca^{2+} (Fig. 8). The archeal PKD-like domain possesses the bidentate Asp802 equivalent. However, the loop is significantly shortened compared to the bacterial domains, which consequently removes the Asp804 equivalent. It should be noted that water positioning residues in s2b appear to be conserved in archaea (Ser845 is conserved, while Ser846 is replaced with asparagine). The mammalian PKD meanwhile, lacks all residues that interact with Ca^{2+}.

Comparison of clostridial PKD-like domain structures with V-set kappa light chain Bence-Jones protein (1BRE (*104*)) as well as with archaea PKD-like domain and human PKD suggests that the PKD1 domain fold in eukaryotes descended from the simpler Ig fold and then may have spread to archaea. From archaea, the fold laterally spread to bacteria. Characteristics of the V-set Ig fold that are shared with the PKD-like domain fold are: (1) The tertiary structure consists of a two-faced β-sheet architecture made up from a well-packed hydrophobic core. (2)

β-strand A is broken by a conserved bulge that changes the allegiance of the subsequent β-strands A'. (3) β-strand C contains the conserved tryptophan, and along with β-strand C' (β-strand D of the PKD-like domain fold), forms a β-hairpin connected by an approximately i→i+8 ω-loop. (4) The turn leading into β-strand F is stabilized by a Δ4 Tyr corner.

Conclusion

Comparison of the crystal structures of ColG s2 with the crystal structures of ColH s2a and s2b, suggests that despite common tertiary folds, PKD-like domains could be grouped into two subsets. The subset containing ColH derived domains exhibits exposed aromatic residues and is found in M9B collagenases with a single CBD. The surface aromatic residues could be involved in secondary interactions that allow weak collagen binding. In contrast, the subset containing s2 is likely to be different; the lack of surface aromatic residues on s2 suggests that the domain is less directly involved in interactions with collagen. Overall, this subset is found in M9B collagenases with multiple CBDs. The unique differences in dynamics and surface characteristics between s2a-s2b and s2 may aid in synergistic collagenolysis.

Meanwhile, the N-terminal linker structure of a PKD-like domain is described for the first time in the *holo*-s2b structure, and suggests that Ca^{2+} repositions the linker along the barrel axis. The helical structure of the linker upon Ca^{2+} binding may shorten the distance between s2b and s2a, and may help account for the previously described proteinase resistance (*81*). Lastly, our stability data show that the domains are extremely stable in the presence of physiological Ca^{2+}. Structural and stability data are critical for the development of PKD-like domains as part of the site-directed delivery of signal molecules such as growth factors and cytokines.

Accession Codes

The full atomic coordinates of the PKD-like domains and their corresponding structure

factor amplitudes have been deposited in the Protein Data Bank (PDB accession codes 4U6T, 4U7K, 4JGU, 4TN9, and 4JRW assigned to *apo*-s2a, *holo*-s2a, *holo*-s2b, and forms I and II of *apo*-s2 respectively).

Table 1: Data collection and refinement statistics.

	Holo-s2a	*Apo*-s2a	*Holo*-s2b	*Apo*-s2 Form I
Data Collection				
X-ray wavelength (Å)	1.54	1.54	0.97937	0.919
Space group	C2	P6$_1$	P2$_1$	P2$_1$
a (Å)	102.3	88.3	49.38	25.00
b (Å)	87.6	88.3	38.87	71.76
c (Å)	104.4	123.6	54.66	47.89
β (°)	116.0	90.0	98.43	95.56
γ (°)	90.0	120.0	90.00	90.00
Resolution (Å)	93.8-1.9	76.5-1.8	22.19-1.42	47.67-1.40
Highest resolution bin (Å)	1.98-1.91	1.81-1.76	1.44-1.42	1.44-1.40
Number of reflections	224,470	252,508	137,190	124,611
Redundancies (a)	3.6 (3.1)	4.7 (2.1)	3.5 (2.9)	3.3 (2.4)
Completeness (%) (a)	96.4 (93.8)	99.4 (95.7)	96.7 (85.8)	92.5 (83.7)
I/σI (a)	13.8 (3.0)	11.5 (2.3)	33.0 (3.1)	21.3 (2.0)
R$_{meas}$ (%) (a, b)	5.4 (31.0)	7.5 (37.4)	4.6 (38.5)	8.1 (46.9)

Table 1 (cont.): Data collection and refinement statistics.

	Holo-s2a	*Apo*-s2a	*Holo*-s2b	*Apo*-s2 Form I
Refinement				
Unique reflections	58,720	50,932	35,993	28,786
R_{cryst} (%) (a, c)	16.3 (25.5)	20.7 (26.8)	15.0 (25.4)	16.9 (28.3)
R_{free} (%) 5% of data (a, d)	20.7 (29.7)	27.2 (32.0)	19.1 (33.2)	21.2 (35.6)
Average B-factor: Main chain A (Å^2)	18.9	17.6	9.6	12.5
Average B-factor: Side chain A (Å^2)	23.3	21.1	13.7	16.5
Average B-factor: Main chain B (Å^2)	18.9	17.9	8.3	12.5
Average B-factor: Side chain B (Å^2)	23.5	21.4	12.2	16.8
Average B-factor: Main chain C (Å^2)	18.9	18.9	N/A	N/A
Average B-factor: Side chain C (Å^2)	23.6	22.0	N/A	N/A
Average B-factor: Main chain D (Å^2)	18.7	19.3	N/A	N/A
Average B-factor: Side chain D (Å^2)	23.1	22.3	N/A	N/A
Average B-factor: Main chain E (Å^2)	19.0	N/A	N/A	N/A
Average B-factor: Side chain E (Å^2)	23.5	N/A	N/A	N/A
Average B-factor: Main chain F (Å^2)	18.7	N/A	N/A	N/A
Average B-factor: Side chain F (Å^2)	23.0	N/A	N/A	N/A
Average B-factor: Main chain G (Å^2)	18.8	N/A	N/A	N/A
Average B-factor: Side chain G (Å^2)	23.1	N/A	N/A	N/A
Average B-factor: Main chain H (Å^2)	19.0	N/A	N/A	N/A
Average B-factor: Side chain H (Å^2)	23.5	N/A	N/A	N/A
Average B-factor: Solvent (Å^2)	46.1	42.2	25.6	31.1
Ramachandran statistics				
Favored (%)	98.5	99.4	98.9	99.4
Additionally allowed (%)	1.5	0.6	1.1	0.6
Outliers (%)	0	0	0	0

(a) Data for the highest resolution shell are given in parenthesis

(b) $R_{meas} = \sum_h \sum_i \sqrt{\frac{N_h}{N_h - 1}} |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$

(c) $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 95% of reflection data used for refinement.

(d) $R_{free} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 5% of reflection data excluded from refinement.

Table 2: Ca-O bond distances and planar deviation for *holo*-s2b and *holo*-s2a.

s2b	Residue	Atom	Distance (Å)	Planar Deviation (Å)
Molecule A	HOH	O	2.49	0.09
	N774	OD1	2.39	0.20
	K775	O	2.41	0.27
	D802	OD1	2.42	0.38
	D802	OD2	2.44	0.22
	D804: axial	OD1	2.34	NA
	D844: axial	OD2	2.42	NA
Molecule B	HOH	O	2.40	0.14
	N774	OD1	2.37	0.24
	K775	O	2.38	0.28
	D802	OD1	2.46	0.38
	D802	OD2	2.49	0.19
	D804: axial	OD1	2.34	NA
	D844: axial	OD2	2.39	NA
s2a	Residue	Atom	Distance (Å)	Planar Deviation (Å)
Average of Molecules A through H	HOH	O	2.53 (5)	0.03 (2)
	N685	OD1	2.52 (6)	0.10 (3)
	S686	O	2.43 (6)	0.14 (3)
	D713	OD1	2.51 (3)	0.15 (3)
	D713	OD2	2.50 (6)	0.09 (4)
	D715:axial	OD1	2.42 (3)	NA
	754: axial	OD2	2.40 (5)	NA

Table 3: Stability paramaters for Gu.HCl denaturation of PKD-like domains.

	holo-s2b	apo-s2b	holo-s2a	apo-s2a	holo-s2	apo-s2
ΔG (kcal*mol^{-1})	9.8	6.9	7.8	4.9	6.0	4.3
m (kcal*mol^{-1}*M^{-1})	3.1	2.6	3.6	3.1	2.4	2.7
C_M (M)	3.2	2.6	2.2	1.6	2.5	1.6

Figures
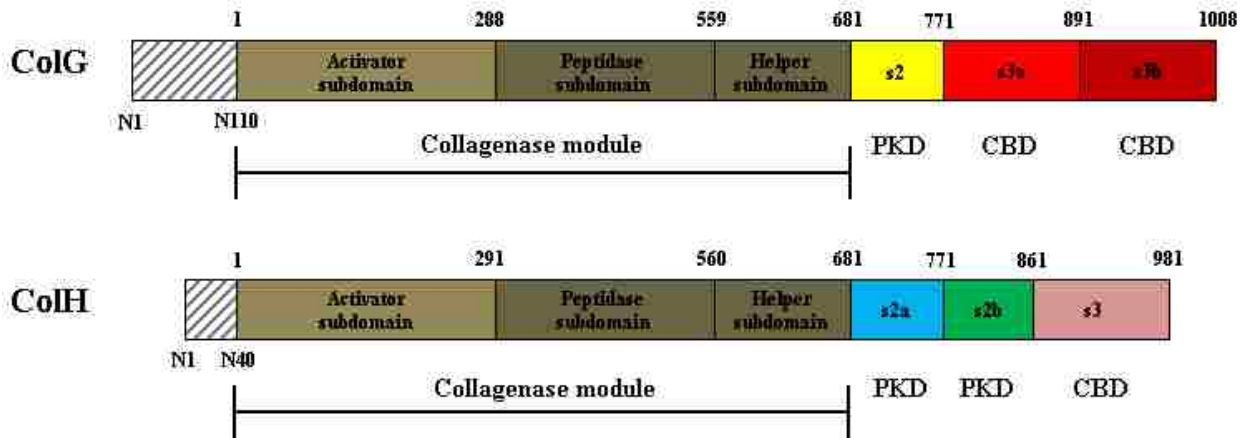
Figure 1: Domain map of collagenases ColG and ColH from *Clostridium histolyticum*. The prepro-peptide (gray dashes) is cleaved from the mature enzyme and indicated by sequence numbering N1-N110 (ColG) and N1-N40 (ColH). The collagenase module is composed of an activator subdomain (olive) and peptidase subdomain (dark olive) that is accompanied by a helper subdomain. The PKD-like domain(s) (yellow for ColG; cyan and green for ColH) connect the collagenase module to the C-terminal CBD(s) (red for ColG; salmon for ColH) that are responsible for collagen-binding.

Figure 2: Structural comparison of *holo*-s2a (A), *apo*-s2a (B), *holo*-s2b (C), and *apo*-s2 (D). Hydrogen bonds and sharp turns that stabilize β-bulges are highlighted.

```
                                    3₁₀      A         A'        B         B'        C          D         E         F          G
                                    ▬       →         →         →         →         →          →         →         →          →
                               770      780       790        800        810        820        830        840        850        860
Number of CBD                   *        *         *         *          *          *          *         *          *          *
    1        ColH s2b          AEIKDLSENKLPVIYMHVP-KSGALNQKVVFYGKGTYDPDGSIAGYQWDFGDGSDFSSEQNPSHVYTKKGEYTVTLRVMDSSGQMSEKTMKIKITD
    1        B.weihenstephanensis --------NKAPVAVINGP-YSGHTNEAISFKSDGSKDEDGKIASYNWEFGDGA-VSNEQNPTHVYTKEGTYTAKLTVTDDKGLTNTATTSVTVQK
    1        B.cereus          --------NKAPVAVINGP-YSGNVNEAISFTSNGSKDEDGKIISYKWEFGDGA-VSNEQNPTHVYTKEGAYTAKLTVTDDKGATNTATASVTVQK
    1        B.mycoides        --------NKAPVAVINGP-YSGNVNEAISFKSDGSKDEDGKIASYKWEFGDGT-ESSEQNPTHVYTKEGTYTAKLTVTDDKGLTNTATASVTVQK
    1        B.anthracis       --------NKAPVAVINGP-YSGNVNEAISFKSDGSKDEDGKIVAYKWEFGDGT-VSNEQNPTHVYTKEGTYTARLTVTDDKGLTNTVTTNVTVQK
    1        B.thuringiensis   --------NKAPVAFINGP-YSGNVNEAISFKSDGSKDEDGKITSYKWEFGDGT-VSNEQNPTHVYTKEGAYTAKLTVTDDKGATNTATATVTVQK
    1        ColH s2a          --------NSLPYGKINGT-YKGTEKEKIKFSSEGSFDPDGKIVSYEWDFGDGN-KSNEENPEHSYDKVGTYTVKLKVTDDKGESSVSTTTAEIKD
    2        C.limosum         --------NKKPIAKINIAKNNVKLNEDIQFNGDESTDTDGNIVSYLWDFGDGK-TSNEVNPNHKYSKEGCYDVKLKVTDDKGAIGEETIKISVVN
    2        C.perfringens     --------NKEPKAVIKSD-SSVIVEEEINFDGTESKDEDGEIKAYEWDFGDGE-KSNEAKATHKYNKTGEYEVKLTVTDNNGGINTESKKIKVVE
    2        C.botulinum       --------NEAPTIKLDFP-KKSNTDEKIKFSSEGSTD-DGKIVSYAWDFGDGE-TSSEKNPTHAYKTPGTYTVKLTVTDDKGLKAEKTSSITINK
    2        B.brevis          --------NKPFVIVMNGP-SKAAVNEEVAFSSNGSSDPDGSIVSYLWEFGDGT-SSSDANPTHVYENQGEYTVKLRVTDDKGAVTAKSMTITIEE
    2        ColG s2           --------NKAPIAKVTGP-STGAVGRNIEFSGKDSKDEDGKIVSYDWDFGDGA-TSRGKNSVHAYKKAGTYNVTLKVTDDKGATATESFTIEIKN
                                *        *         *         *          *          *          *         *          *          *
                               680      690       700        710        720        730        740        750        760        770
                                        →         →         →          →          →                    →          →
                                                                                →          →          →
                                         A         A'        B          B'         C          D         E         F          G
```
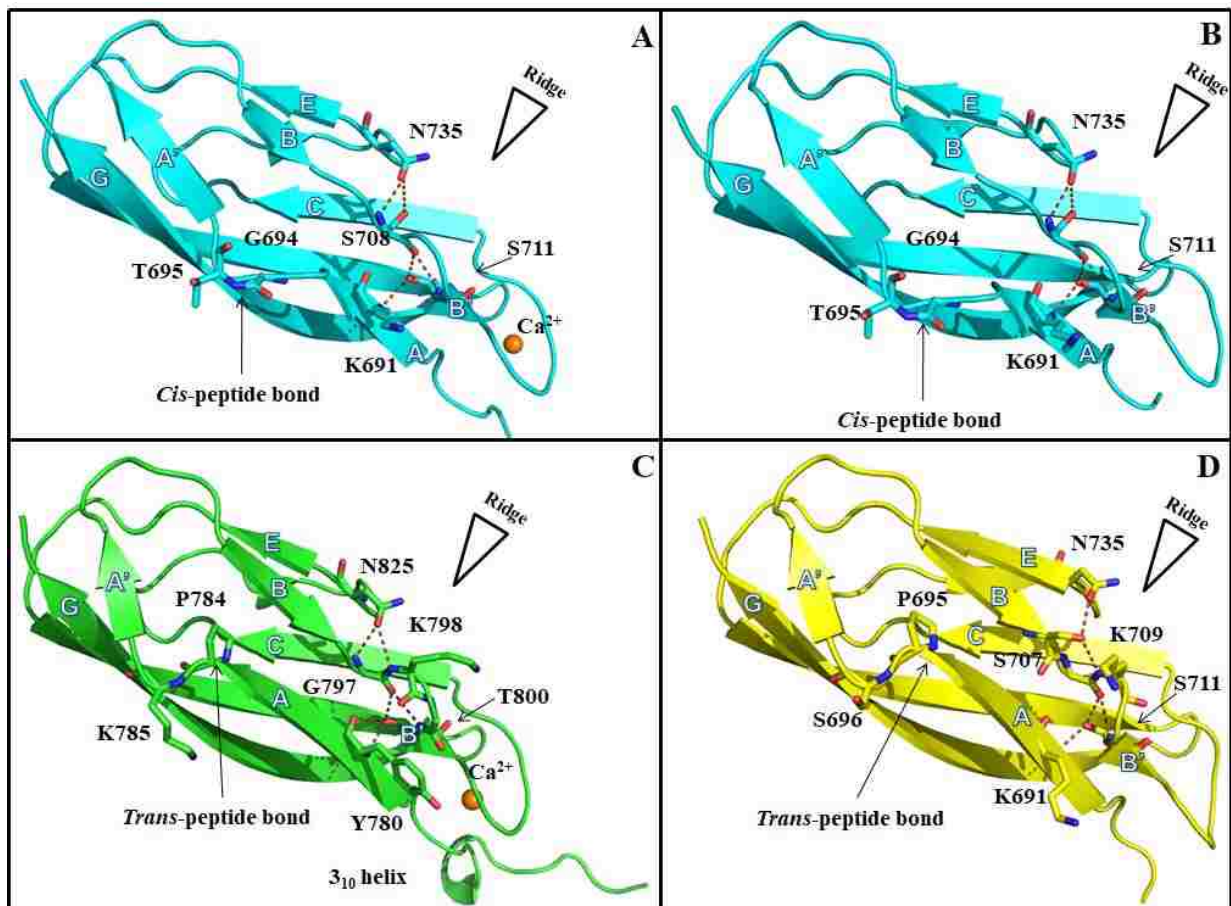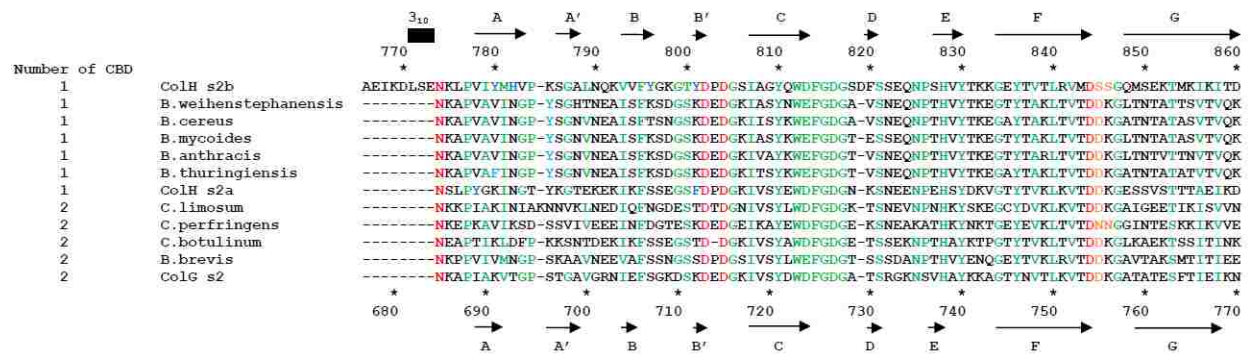
Figure 3: Structure based sequence alignment of PKD-like domains from M9B. Residues responsible for Ca^{2+} binding, for positioning the Ca^{2+} interacting water, architecturally critical residues, and surface aromatic residues are shown in red, orange, green and blue, respectively. Sequence numbering and secondary structure position for s2b is shown at the top of the figure. Secondary structure position for the s2 structure is similar, though the 3_{10} helix is absent. Sequence numbering for s2a and s2, as well as secondary structure position for s2a is shown at the bottom of the figure. Sequence alignment was aided by the use of ClustalW2 (*79*).
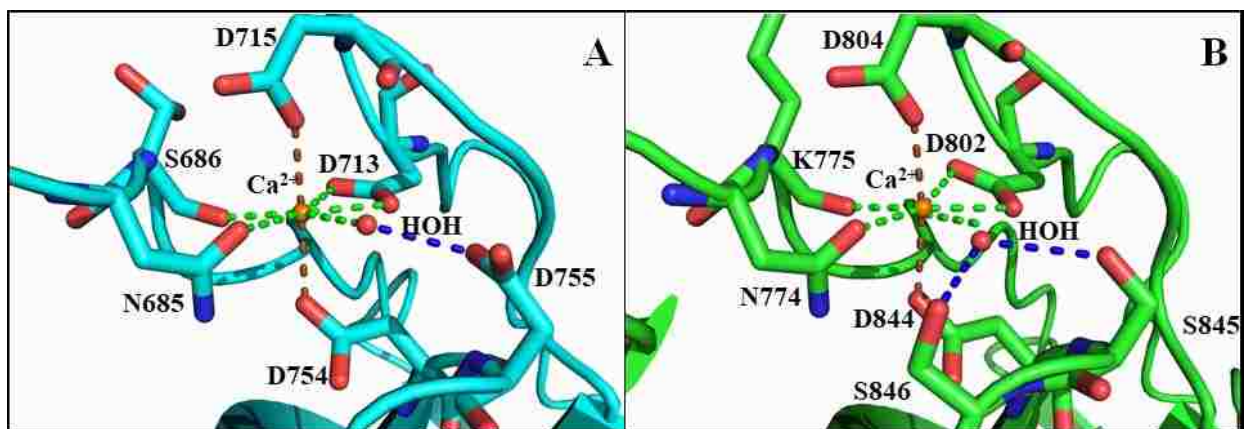
Figure 4: Ca^{2+} coordination in s2a (A) and s2b (B). Seven oxygen atoms from five residues and one water molecule coordinate with Ca^{2+} in a pentagonal bipyramidal geometry. Pentagonal base interactions are indicated using brown dashes, while axial interactions are indicated using yellow dashes. Residue to water interactions are indicated with blue dashes. Either one aspartate (s2a) or adjacent serines (s2b) are responsible for positioning the water molecule along the pentagonal base.
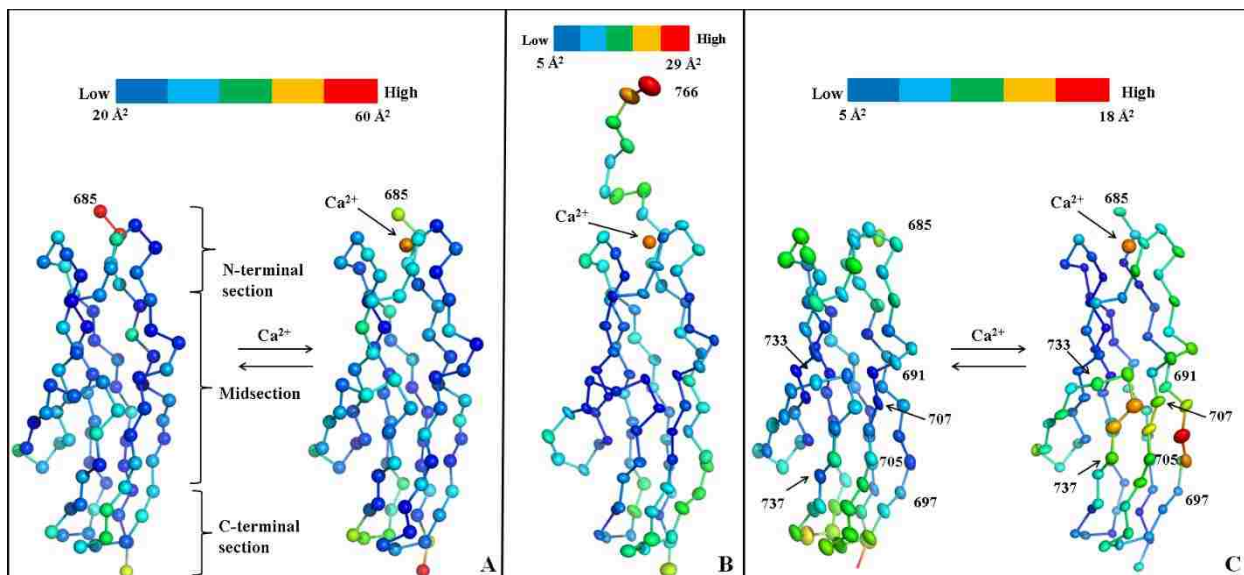
Figure 5: C_α B-factor changes upon Ca^{2+} binding for s2a (A) and s2 (C). C_α B-factor of *holo*-s2b (B) is also shown.
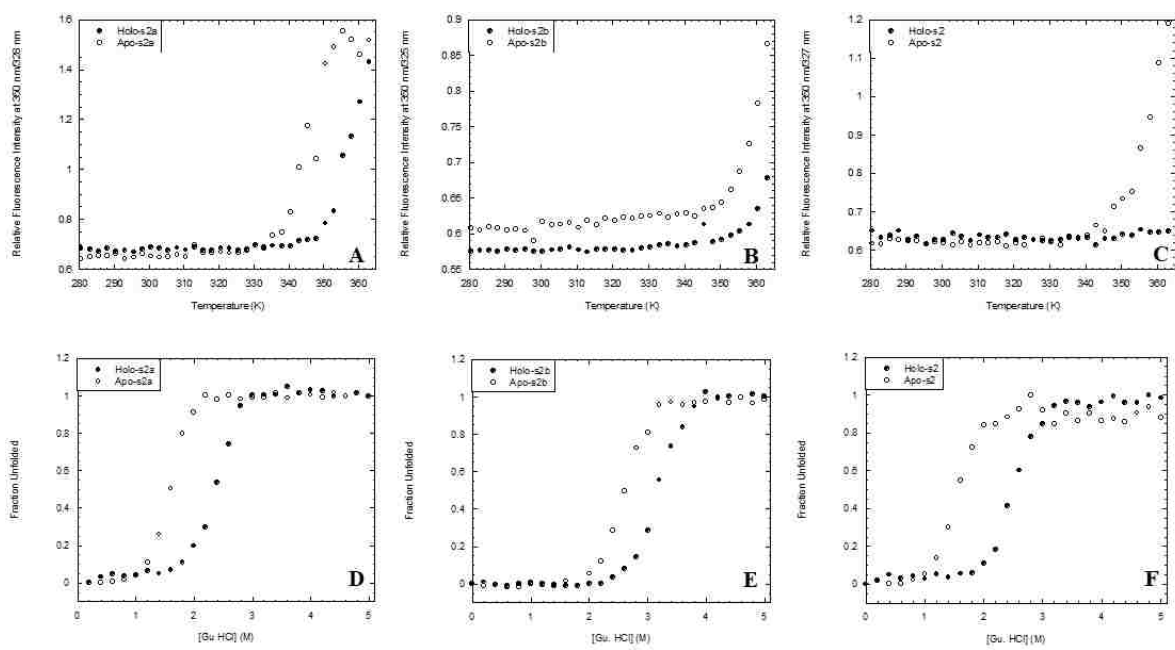
Figure 6: Results of fluorescence-measured equilibrium denaturation of (A, D) s2a, (B, E) s2b, and (C, F) s2 in their *apo*-(open circles) and *holo*-forms (closed circles).
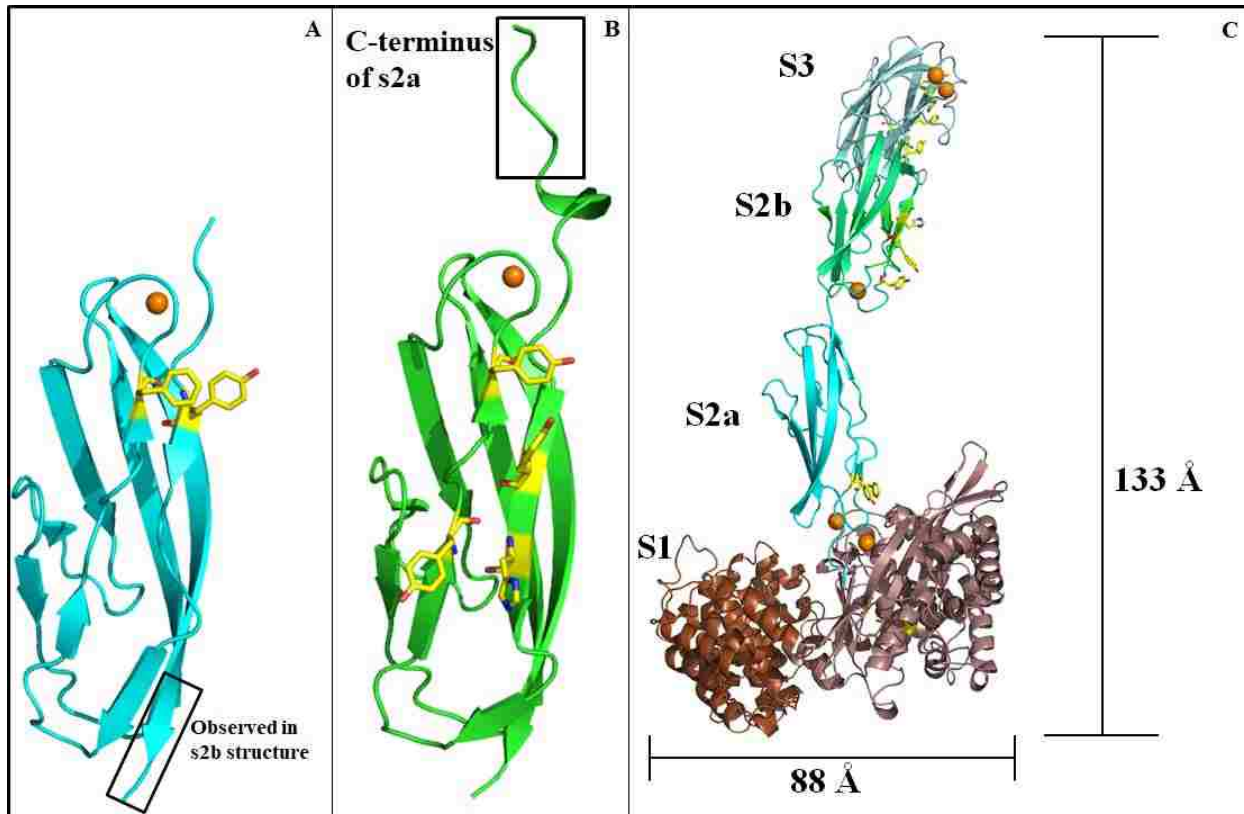
Figure 7: Surface aromatic residues in s2a (A) and s2b (B). The boxed in regions corrspond to residues Ala766-Asp770, which is observed in both s2a and s2b structures and was used to help assemble the full *holo*-ColH structure (C). This structure is assembled from the crystal structures of the peptidase domain of s1, s2a , s2b , and s3, as well as the homology modeled activator domain of s1. Homology modeling was accomplished using SWISS-MODEL (*111*). Surface exposed aromatic residues of the peptidase domain of s1 and s2a-s2b as well as the conserved collagen interacting aromatic residues of s3 are shown in yellow. Ca^{2+} is shown as orange spheres.
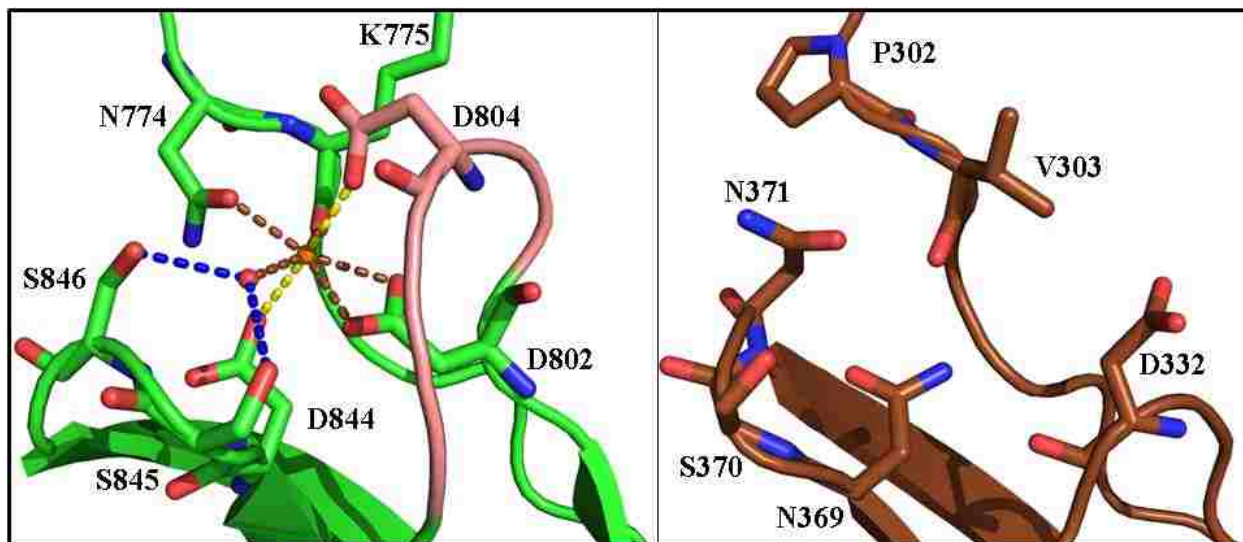
Figure 8: Proposed evolution of the Ca^{2+} binding pocket in bacterial PKD-like domains (*holo*-s2b shown on the left) from the archaea PKD-like domain (1LOQ shown on the right).

Supplemental Material

Tables

Supplemental table 1: Data collection and refinement statistics.

	apo-s2 Form I
Data Collection	
X-ray wavelength (Å)	0.919
Space group	$P2_12_12_1$
a (Å), b (Å), c (Å)	45.0, 49.0, 70.9
β (°)	90.0
γ (°)	90.0
Resolution (Å)	40.3-1.6
Highest resolution bin (Å)	1.64-1.60
Number of reflections	165,013
Redundancies (a)	4.1 (3.6)
Completeness (%) (a)	99.9 (99.6)
$I/\sigma I$ (a)	30.9 (2.1)
R_{meas} (%) (a, b)	7.0 (69.2)

Supplemental table 1 (Cont.): Data collection and refinement statistics.

	apo-s2 Form I
Refinement	
Unique reflections	20,159
R_{cryst} (%) (a, c)	16.2 (26.5)
R_{free} (%) 5% of data (a, d)	19.2 (30.6)
Average B-factor: Main chain A (Å^2)	11.2
Average B-factor: Side chain A (Å^2)	15.7
Average B-factor: Main chain B (Å^2)	11.6
Average B-factor: Side chain B (Å^2)	16.2
Average B-factor: Solvent (Å^2)	28.1
Ramachandran statistics	
Favored (%)	100
Additionally allowed (%)	0
Outliers (%)	0

(a) Data for the highest resolution shell are given in parenthesis

(b) $R_{meas} = \sum_h \sum_i \sqrt{\frac{N_h}{N_h - 1}} |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$

(c) $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 95% of reflection data used for refinement.

(d) $R_{free} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 5% of reflection data excluded from refinement.

Supplemental table 2: Alternate conformations of the PKD-like domains

| Domain	Molecule	Alternate conformations
Apo-s2a	Molecule A	S720, K742, and S762
	Molecule B	D715 and K742
	Molecule C	K697, S720, and N732
	Molecule D	D715, S720, and N732
Holo-s2a	Molecule A	S686, S759, and T763
	Molecule B	S720 and S759
	Molecule C	None
	Molecule D	S762
	Molecule E	S686 and S762
	Molecule F	S759 and S762
	Molecule G	S686 and S759
	Molecule H	S686 and S762
Holo-s2b	Molecule A	K792, V793, S806, S822, S827, and M854
	Molecule B	S786, S806, S822, and S827
Apo-s2 Form I	Molecule A	R702, K717, R732, S736, T761, and S763
	Molecule B	I689, K691, S720, T730, T749, and T761
Apo-s2 Form II	Molecule A	E714, N747, T749, and T761
	Molecule B	K691, E714, V737, and S763

Supplemental table 3: hydrogen bond totals for PKD-like domains in presence and absence of Ca^{2+}.

| Domain	Molecule	NH…O	OH…O	NH…N	CH…O	total
Apo-s2a	Molecule A	45	8	19	95	167
	Molecule B	38	8	19	87	152
	Molecule C	37	7	18	93	155
	Molecule D	39	8	20	92	159
Holo-s2a	Molecule A	45	7	18	86	156
	Molecule B	43	6	18	90	157
	Molecule C	42	7	19	91	159
	Molecule D	46	6	18	88	158
	Molecule E	47	6	18	86	157
	Molecule F	41	7	18	95	161
	Molecule G	40	6	18	89	153
	Molecule H	40	7	18	84	149
Holo-s2b	Molecule A	45	12	20	73	150
	Molecule B	47	6	20	85	158
Apo-s2	Molecule A	46	5	18	77	146
	Molecule B	48	4	18	82	152
Holo-s2	Molecule A	45	6	17	70	138
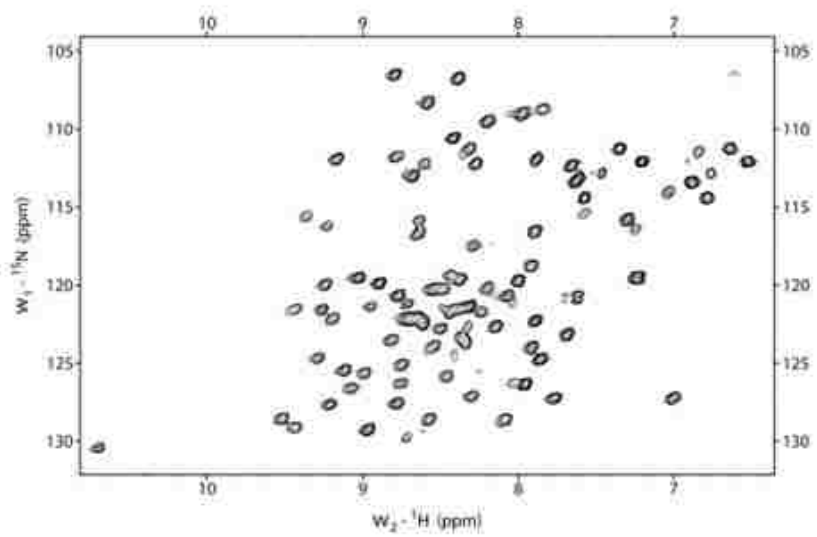
Supplemental figure 1: HSQC spectra for uniformly ^{15}N labeled s2. In the spectra, thirteen residues could not be identified due to extensive band broadening.
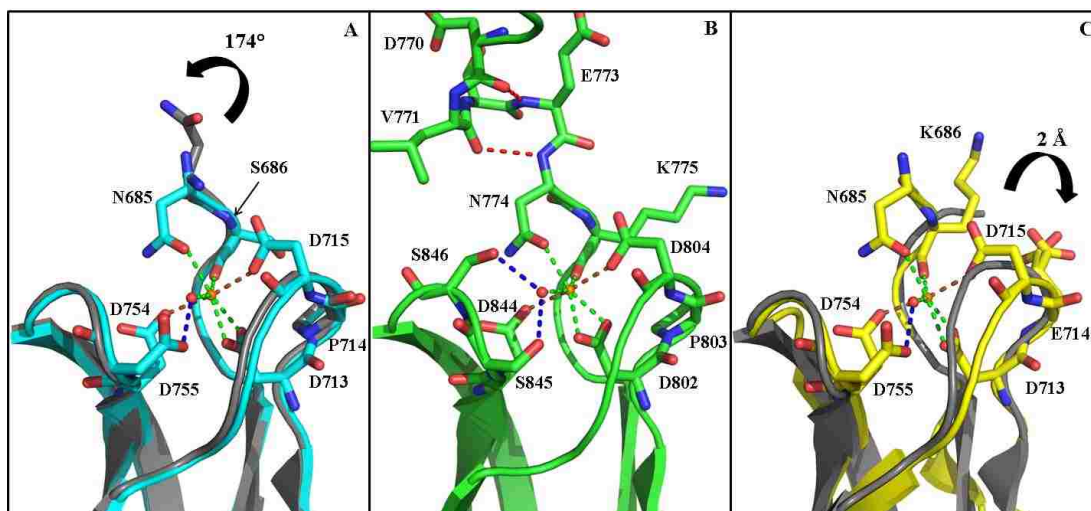
Supplemental figure 2: Ca^{2+}-induced structure rearraingment in the PKD-like domains. The N-terminal loop of s2a (A) is re-oriented as indicated by the rotation along the ψ bond of Asn685. The N-terminal linker of s2b (B) is observed in the crystal structure and indicates that the linker forms a 3_{10} helix (hydrogen bonding indicated by red dashes). Unlike s2a and s2b, the loop (713-717) of s2 (C) moves out to accommodate Ca^{2+}.
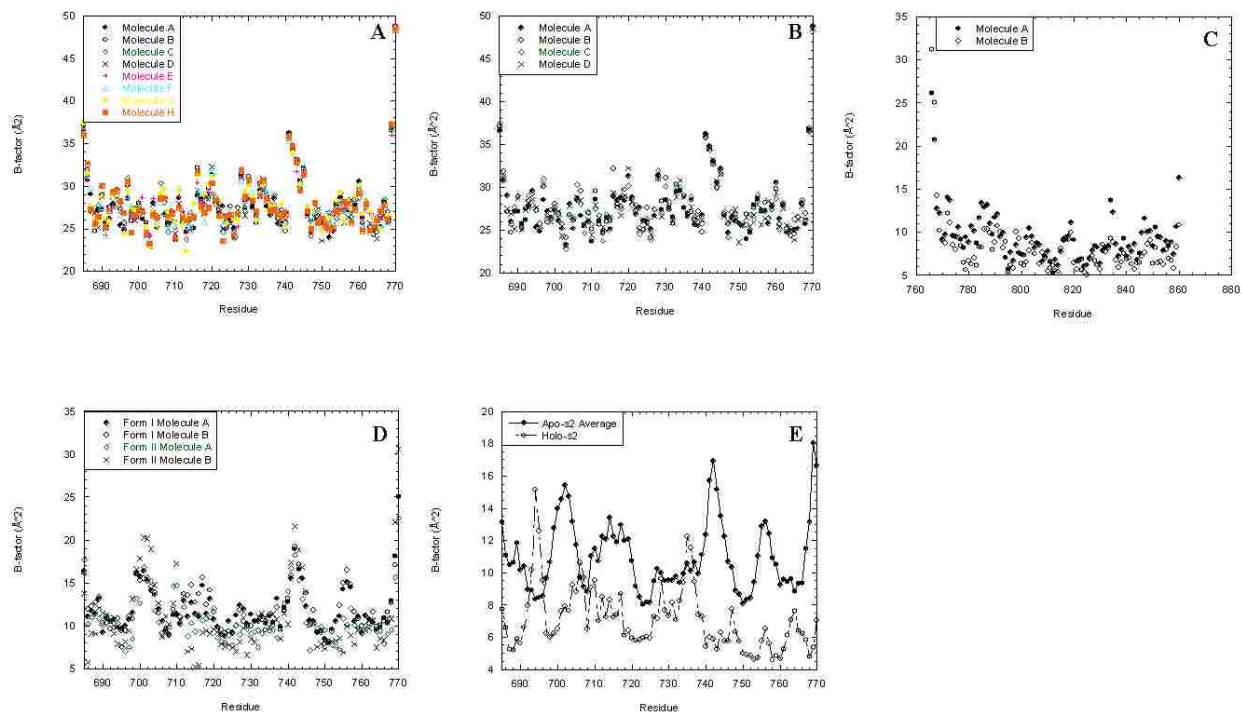
Supplemental figure 3: Per-residue B-factor trend for the PKD-like domains. The trends for *holo*-s2a (A), *apo*-s2a (B), *holo*-s2b (C), and *apo*-s2 (D) are similar. Comparison of the averaged B-factor of *apo*-s2 with *holo*-s2 (E) revealed the distinctly different influence Ca^{2+} has on potential dynamics.

Chapter 4: Activation and binding mechanism of a clostridial tandem collagen-binding domain with pseudo-two-fold symmetry

Abstract

Clostridium histoliticum secrets virulence factors, including highly active collagenases CollG (class I) and ColH (class II), which penetrate animal tissues. After the multi-domain ColG utilizes its tandem collagen-binding domain (CBD) to anchor itself onto insoluble collagen, subsequent degradation of the hierarchical substrate involves processive cleavage and rearrangement of fibrils. In this work, the structure of the calcium bound tandem CBD is presented at 1.9 Å resolution (R_{work} = 15.0%; R_{free} = 19.6%). The pseudo-two-fold arrangement of CBD could allow ColG to wedge between collagen molecules that are 55 Å apart and subsequently aid in fibril rearrangement and processive cleavage. Indeed, between 0.1:1 and 0.5:1 molar ratios of tandem CBD and collagen, it accelerated collagen fiber formation. At 1:1 molar ratios and above, the tandem CBD retarded fibril formation. To toggle between collagen molecules, a tighter binding C-side CBD, may initiate binding. Subsequently, the weaker binding N-side CBD can latch onto a prone collagen molecule to provide the tightest known fibril binding. The conformational change of the tandem CBD is calcium dependent and cooperative as measured by size exclusion chromatography and by SAXS at pCa in the range of 3-6. At pCa > 5, the tandem CBD adopts an extended structure that is easier to be secreted from the bacterium. In the host pCa^{2+} >3, the compact structure seen in the crystal structure is adopted. The binding and activation mode described here will help guide site-directed drug delivery vehicle development.

Introduction

Fibrous collagen is resistant to most proteases due to its tightly packed structure, which shields the peptide bonds. Most clostridial collagenases possess multiple collagen-binding domains (CBD) (Supplementary Fig. 1). The evolutionary advantage for this redundancy is addressed by determining the structure of the tandem CBD segment of collagenase ColG from *Clostridium histolyticum*. The catalytic domain alone is capable of hydrolyzing solubilized collagen while C-terminal CBDs of ColG (s3a, s3b) are necessary for dismantling collagen fibril (*29, 31*).

The C-terminal collagen-binding domains (CBDs) of ColG (s3a, s3b), and ColH (s3), are homologs consisting of approximately 120 amino acids. The domains bind to soluble and insoluble collagen structures. Their role in binding to collagen fibril is essential in dismantling its hierarchical structure (*40, 43*). Truncation of CBD from either full-length ColG or ColH incapacitates their abilities to degrade collagen fibril. Such enzymes can only degrade solubilized collagen or denatured collagen (gelatin). Mutagenesis and collagen-binding studies mapped the binding surface of s3b, while NMR and SAXS studies showed that s3b unidirectionally binds to under-twisted regions of mini-collagen (*45, 66*). High-speed atomic force microscopy has recently revealed ColG's ability to dismantle collagen in real time (*33*). During degradation, ColG moves processively from the fibril's C-terminus to its N-terminus to dismantle the fibril. ColG also initially targeted less ordered regions of the fibril.

Bacterial collagenases require calcium to attain both full catalytic activity and collagen-binding function. The activation of bacterial collagenase involves domain rearrangement triggered by the Ca^{2+} binding (*29, 30, 64, 80, 112*). The full-length ColG is expected to be relatively flexible inside the bacteria where Ca^{2+} concentration is low ($0.2\text{-}0.3 \times 10^{-6}$ M) allowing

the enzyme to be secreted easier (*29*). Upon secretion, the linker chelates to Ca^{2+} (~1.2 mM) in the ECM to adopt a rigid structure. Though it has not been shown for ColG, Ca^{2+} chelation indeed triggers full-length ColH to adopt a compact, less flexible structure (*81*).

The clostridial collagenases have been successfully used for years as a wound debridement. Recently the mixture of ColG and ColH was approved for use in the treatment of excessive connective tissue build up found, for example, in Dupuytren's disease (*113*). In addition to therapeutic use of full-length collagenase for removal of connective tissue, the non-catalytic segments are used for targeted drug-delivery to reduce dosage and to minimize side effects. Initially, Nishi *et al*. developed fusion proteins of s2b-s3 and growth factors. When injected, the fusion proteins remained active at the site of injection for up to 10 days (*46*). While systemic applications utilize fusion proteins consisting of lower affinity collagen-binders (e.g. s3) to treat osteoporosis (*48-50*), and to prevent and to treat alopecia (*47, 96, 114, 115*), tighter collagen-binders (e.g. s2b-s3) are more efficacious in localized wound healing applications when applied at the site of injury with collagen-based bone graft material (*116-118*). Since tandem CBD bind to collagen even tighter than s2b-s3 (*43*), these results suggest a clinical significance for tandem CBD.

Methods

Production, Purification, Crystallization and Structure Determination

Individual s3a, s3b, as well as tandem CBD derived from the *C. histolyticum* ColG were expressed as glutathione S-transferase (GST)-fusion proteins using method as described previously (*40*). Initial conditions suitable to grow crystals of tandem CBD were identified by high-throughput screen (Hampton Research Crystal Screen HT). Subsequent crystallization trials using the initial conditions were carried out using the hanging-drop method. Crystals of tandem

CBD, obtained in the presence of 21-26% PEG 3350, 0.1 M HEPES pH 7.5, 3 mM calcium chloride, were orthorhombic (space group, $P2_12_12_1$), with cell parameters $a = 51.5$ Å, $b = 54.7$ Å, $c = 92.0$ Å. The crystals grew within 24 hours in 37°C but did not grow at lower temperatures. The crystals were temperature sensitive and could not withstand cryogenic temperatures. Therefore, diffraction data were collected by means of in–house X-ray facility at room temperature to 1.9 Å resolution using a Rigaku 007 generator with Cu $K\alpha$ radiation. The data sets were processed with d*TREK (99) (Supplementary Table 1). The structure was solved with the molecular replacement program MolRep from the CCP4 package, by using s3b (PDB code 2O8O) as the search model (52). One tandem CBD was found in an asymmetric unit; and therefore, V_M was 2.5 Å^3/Da and solvent content was 50% (119). Refinement of the tandem CBD was carried out using Refmac_6.1.13 (52). TLS restraints were applied to main chain atoms with each CBD acting as a TLS group. Babinet scaling was used for bulk solvent refinement. Five percent of the data were set aside to monitor R_{free}. The models were manually adjusted between each refinement cycle using MIFit (55). Alternate confirmations were built for Lys818, Glu945, Tyr970, and Arg1005. The Ramachandran plot for the final structure obtained with the RCSB validation server (120) showed 91% of the residues in the core region and 9% in the additionally allowed region, with none in the generously allowed or disallowed regions. The final refinement statistics are shown in Table 1.

Small Angle X-ray Scattering

Suitable buffer conditions for small angle X-ray scattering (SAXS) measurements were identified using Native-PAGE. For the pCa (-log [Ca^{2+}]) analysis the tandem CBD was initially equilibrated into 100 mM NaCl and 2% glycerol. Measured quantity of $CaCl_2$ and 0.2 mM total EGTA achieved pCa values to 3, 4, 5 and 6. The amount of Ca^{2+} needed to reach a given pCa

was determined using MAXCHELATOR (*121*). Measurements were completed for three concentration series per sample. For the pCa series, the concentration of tandem CBD used at pCa 4 and 5 and 6 was 1, 3, and 5 mg/mL. At pCa 3, the concentration series used was 2, 4, and 6 mg/mL. All SAXS data were collected at 10°C at the Advanced Light Source at Berkley National Lab (SIBYLS beamline, 12.3.1) (*122, 123*) using a Pilatus 2M detector. Sample to detector distance was 1.5 m, and X-ray wavelength was 1.127 Å. All data processing was accomplished using primusqt from the ATAS 2.6.1 software package. For CBD at each pCa and the complex with mini-collagen, exposure data from the concentration gradient that were not affected by either aggregation or detector saturation were extrapolated to infinite dilution. Determination of the radius of gyration (R_g), maximum diameter (D_{max}) as well as *ab initio* shape reconstruction of the extrapolated data for each model in the pCa series was carried out using the dammif function in primusqt. The χ values calculated at the end of each run indicated the agreement between the calculated scattering curve and the experimental scattering curve. The R_g, D_{max}, and χ values for each model are summarized in Table 2.

Collagen Fibril Formation

The impact of addition of tandem CBD, S3a and S3b on self-assembly of collagen molecules was monitored by measuring turbidity as an increase in optical density at 450 nm, at 37ºC (Supplementary Fig. 2). On ice, a solution of 2 mg/ml of rat collagen was diluted with 40 mM HEPES buffer pH 7.5, with addition of 300 mM NaCl, 2 mM $CaCl_2$, to final concentration of 0.5 mg/ml (2.4 µM). In the next step collagen binding domain was added in ratios: 0.1 to 1; 0.2 to 1; 0.3:1; 0.5 to 1; 1 to 1; 2 to 1; 3 to 1; 5 to 1 of molar concentration. The turbidity measurement was taken in 96 well plates with 1-minute intervals with spectrophotometer Filter Max F5 (Molecular Devices). From the turbidity curve the following parameters were estimated

t_{lag}-time at the end of lag phase, maximum turbidity, and V the maximum fibril growth rate (Supplementary Table 1). Retardation of collagen fibril formation by s3b (C-side CBD) was as expected (*124*).

Analytical Size Exclusion Chromatography

Size exclusion chromatography was performed at room temperature on a HPLC system equipped with a Superdex 75 column (1 × 30 cm, Pharmacia) at a flow rate of 0.5 ml/min as described (*29*). The following proteins were used as molecular mass standards: bovine serum albumin, 67.0 kDa; chicken ovalbumin, 43.0 kDa; and ribonuclease A, 13.7 kDa (Pharmacia). The measurement was carried out in triplicate. For every apparent mass measured for the tandem CBD at different pCa, standard deviation was less than 0.06 kDa.

Results and Discussion

Structural description of tandem CBD

The crystal structure of tandem CBD consists of s3a, which is described for the first time in this paper, and the previously described s3b (*29*). Both s3a and s3b adopt similar β-sandwich 'jelly-roll' folds composed of ten β-strands. The CBDs are related by a pseudo two-fold rotational symmetry that is stabilized by salt-bridges and hydrogen-bonding interactions. The pseudo symmetry axis, which is perpendicular to the plane of the page in Fig. 1a, positions the collagen-binding pockets in the tandem CBD to be ~55 Å apart. Each domain chelates to two Ca^{2+} as described (*29, 80*). Overlay of 110 equivalent C_α atoms showed that s3a and s3b share a root mean square deviation (r.m.s.d.) of 0.9Å and most significantly deviate at loops with r.m.s.d. of approximately 2.5 Å.

Influence of Ca^{2+} on domain rearrangement

Calcium ions are thought to trigger domain rearrangement that can enhance ColG's efficiency in digesting fibril. The domain rearrangement of tandem CBD triggered by Ca^{2+} binding is examined here. Hydrodynamic radius change monitored by size exclusion chromatography demonstrated that Ca^{2+}-induced domain rearrangement is cooperative. Molecular shape transformation was visualized by small angle X-ray scattering (SAXS, Fig.1b). The SAXS derived envelope for tandem CBD at pCa 6 (1 µM) adopts an elongated, rod-like shape. The protrusion from the envelope resembles the α-helical linker from the molecular envelope of apo-s3b (64). Increasing Ca^{2+} concentration compacts the tandem CBD. At pCa 4 (100 µM) the shape resembles the crystal structure with the exception of a bulge that suggests the N-terminal linker remained dynamic. At pCa 3 (1 mM) the shape agrees well with the crystal structure. The difference in Ca^{2+} concentration inside *Clostridium*, which is likely similar to the concentration inside *Escherichia coli* ($0.2–0.3 \times 10^{-6}$ M; pCa~7)(125), and host ECM (~1.2 mM; pCa~3) (126) could be exploited by the bacteria to facilitate rapid secretion into the host.

Unique influence of tandem CBD on collagen fibril formation

Previously, s3b was shown to bind collagen unidirectionally (45). The pseudo-two-fold arrangement of tandem CBD would allow the domain to bind to two parallelly oriented tropocollagen molecules. Individually, s3a or s3b, retards collagen fibril self-assembly (Supplementary Fig. 2) similar to other collagen-binders (124). However, since tandem CBD could facilitate aligning collagen molecules, we tested its potential to promote fibril self-assembly by monitoring the turbidity of mixtures of tandem CBD and collagen at 450 nm (Fig. 2a). Here, we observed the first instance where a collagen-binder accelerates fibril formation. In the absence of tandem CBD (control), the lag time for fibril formation is 19 min as previously observed (124). The lag time progressively shortens as tandem CBD concentration increases

from 0.1:1 (tandem CBD:collagen) to 0.5:1. At a 0.5:1 ratio of tandem CBD to collagen, the lag time is reduced to 15 min. At low concentrations, tandem CBD indeed aids in collagen fibril self-assembly. As concentration of tandem CBD increases, the lag time also correspondingly increases. At a tandem CBD to collagen ratio of 5:1, the lag time is 30 min. The dual role of the tandem CBD on collagen fibril assembly may provide an important clue to understand how full-length ColG dismantles insoluble fibril. High-speed atomic force microscopy visualized ColG's processive movement and its rearrangement of collagen fibrils (*33*). ColG initiates collagenolysis from disordered regions. Here, the tandem CBD's role is to intercalate into the niche provided in the disordered region and to anchor the collagenase onto the fibril. The subsequent C-terminus to N-terminus processive cleavage could then be facilitated by tandem CBD correctly positioning the catalytic domain. ColG also isolates collagen fibrils and rearranges them to thicken neighboring fibrils. The tandem CBD's ability to facilitate collagen fibril formation may allow ColG to rearrange collagen fibril.

Collagen-binding mode

The opposing positions of the collagen-binding clefts on tandem CBD present unique modes for the domain to latch onto collagen fibril. Whereas collagen fibril is built from a staggered array of triple-helical tropocollagen, and is water insoluble, synthetic mini-collagen, which mimics the tropocollagen structure, is water soluble and also allows solution-based analysis of CBD-collagen interaction. Measured dissociation constants for the interaction between the CBDs and either mini-collagen or fibril tend to agree (*43*). Tandem CBD is the tightest binder to collagen fibril, and is able to bind tighter than the sum of s3a and s3b individually. However tandem CBD binds mini-collagen about as tightly as s3b alone (*43*). Corroborating this, the SAXS-derived shape of the CBD:mini-collagen suggests only one CBD

bound to mini-collagen, even when two-fold excess of mini-collagen was used (Supplementary Table 2). The binding affinities of s3a and s3b to mini-collagen suggest s3b initiates collagen binding. Furthermore, the sequence alignment of multiple CBD binding segments suggests this binding strategy is conserved amongst the collagenases possessing tandem CBDs. The three most critical tyrosine residues for collagen-binding in s3b (970, 994, and 996) are conserved amongst C-side CBDs, while only tyrosine residues 970 and 994 are conserved amongst N-side CBDs (Supplementary Fig. 1b). Apparently, gene duplication of CBD required reduced functionality in the N-side CBD in order to prevent tandem CBD from binding too tightly to collagen fibril. The tighter binding s3b initiates binding and serves the central role, while s3a plays an auxiliary, yet pivotal role in intercalating between collagen molecules in collagen fibril.

The 5.5 nm spacing of binding clefts in tandem CBD excludes the segment from reaching tropocollagen molecules within tightly packed hierarchical arrangements. However, the spacing would allow it to seek either interfibrous spaces or crevices on the surface of damaged collagen fibril (Fig. 2b). The median surface-to-surface distance between fibers in normal skin is only ~3.2 nm. However, the interfibrous space widens to ~6 nm in regenerating skin (*127*). Meanwhile loss of two micro-fibrils from the surface of fibril could also generate an ~6 nm gap (*33*). Bacterial collagenase equipped with tandem CBD could initiate destruction from these most vulnerable regions.

Application of tandem CBD to localized drug delivery

The binding segments can converge fused growth factors to the site of lesion, which minimizes dosage and side effects in rodents (*46, 116-118*). The fusion protein of basic FGF and tandem CBD yielded the highest efficacy in healing bone fracture (*128*). Our results present the structural support for use of tandem CBD in localized therapeutic delivery.

Accession codes

The full atomic coordinates of the tandem CBD and its corresponding structure factor amplitudes have been deposited in the Protein Data Bank (PDB accession code 5IKU). SAXS data and corresponding *ab initio* models have been submitted to the Small Angle Scattering Biological Data Bank (SASBDB accession codes CL2, CM2, CN2, CP2 assigned to envelopes for tandem CBD at pCa 3, 4, 5, and 6 respectively).

Tables

Table 1. Data collection and refinement statistics

Data collection statistic	
Wavelength (Å)	1.5419
Temperature (K)	298
Resolution range (Å)[a]	19.7-1.90 (1.97-1.90)
Space group	$P2_1\,2_1\,2_1$
Unit cell dimension (Å)	
a (Å)	51.5
b (Å)	54.7
c (Å)	92.0
Unit cell angle (°)	$\alpha,\ \beta,\ \gamma = 90$
Total reflections	89,473
Redundancy	4.47 (4.27)
Completeness (%)[a]	94.8 (94.2)
R_{meas} (%)[a, b]	7.1 (49.2)
$I/\sigma I$[b]	11.2 (2.9)

Table 1 (Cont.): Data collection and refinement statistics

Refinement statistic	
Unique reflections	18,987
Solvent molecules	211
R_{work} (%)	15.0 (25.5)
R_{free} (%)[b, c]	19.6 (26.8)
Average B value (Å^2)	37.68
Coordinates ESU based on R_{free} (Å)	0.13
Root mean square deviations	
Bond distance (Å)	0.011
Bond angles (°)	1.87
Chiral centers (Å^3)	0.17
Planar groups (Å)	0.01
B-factor restrains	
Main - chain bond (Å^2)	3.81
Main-chain angle (Å^2)	4.72
Side- chain bond (Å^2)	6.84
Long range B-factor (Å^2)	12.3
Ramachandran statistic	
Most favored region (%)	90.9
Allowed region (%)	9.1

[a] $R_{meas} = \Sigma_{hkl} (N_{(hkl)}/N_{(hkl)}-1)^{1/2} \Sigma_i \left| I_{i(hkl)} - \langle I_{(hkl)} \rangle \right| / \Sigma_{(hkl)} \Sigma_i I_{i(hkl)}$

[b]Data for highest resolution shell are given in parentheses
[c] 5% of data excluded from refinement

Table 2: Small Angle X-ray Scattering Statistics for the tandem CBD at different pCa.

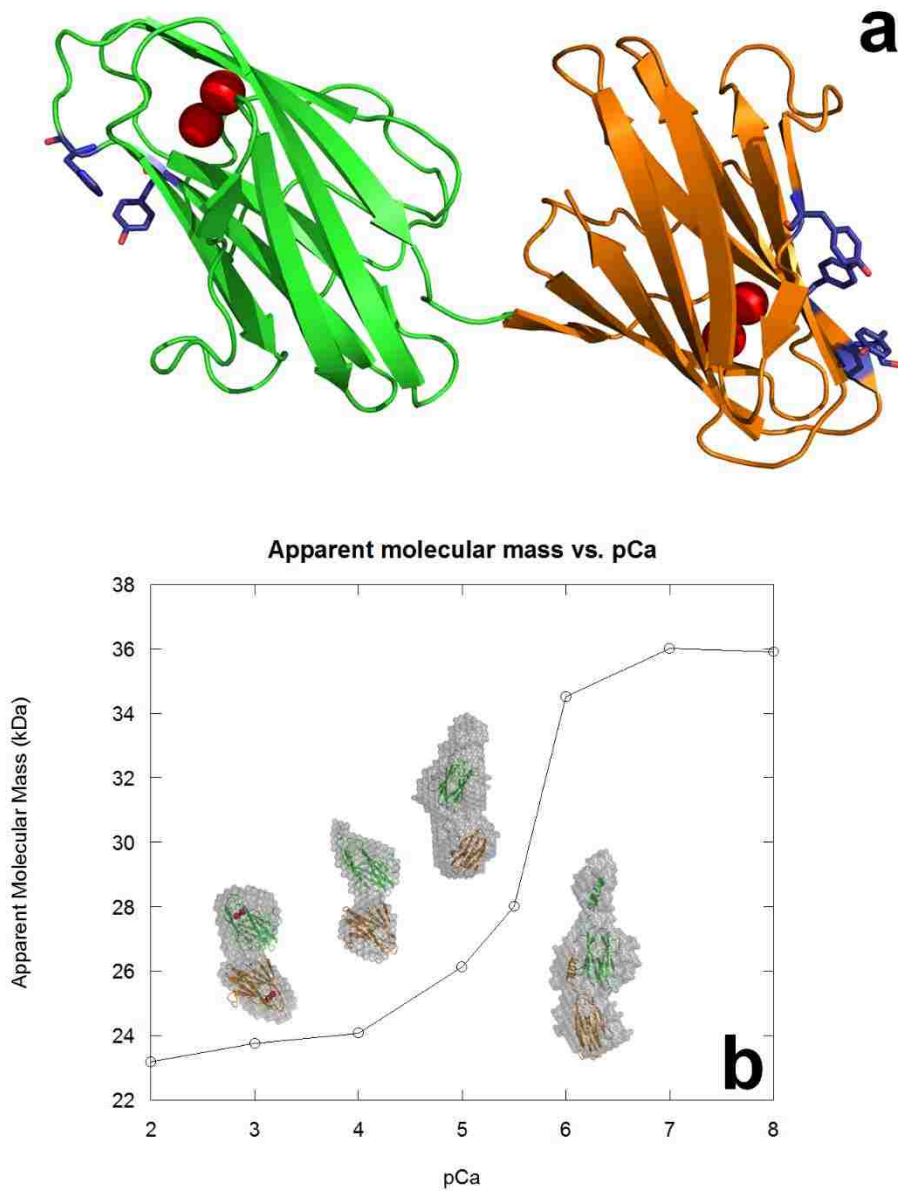| pCa	R_g (Å)	D_{max} (Å)	χ
3.0	24	75	0.8
4.0	25	76	1.0
5.0	25	94	0.8
6.0	27	128	1.2

Fig 1. Crystal structure of tandem CBD and Ca^{2+} induced transformation. (a) The N-side CBD (s3a) and the C-side (s3b) are drawn in green and orange, respectively. Four Ca^{2+} (red) are chelated by the CBDs. Aromatic residues (blue) were previously identified to interact with mini-collagen (*29, 45*). (b) Size exclusion chromatography and SAXS experiments were carried out in HBS-based calcium buffers of varying pCa.
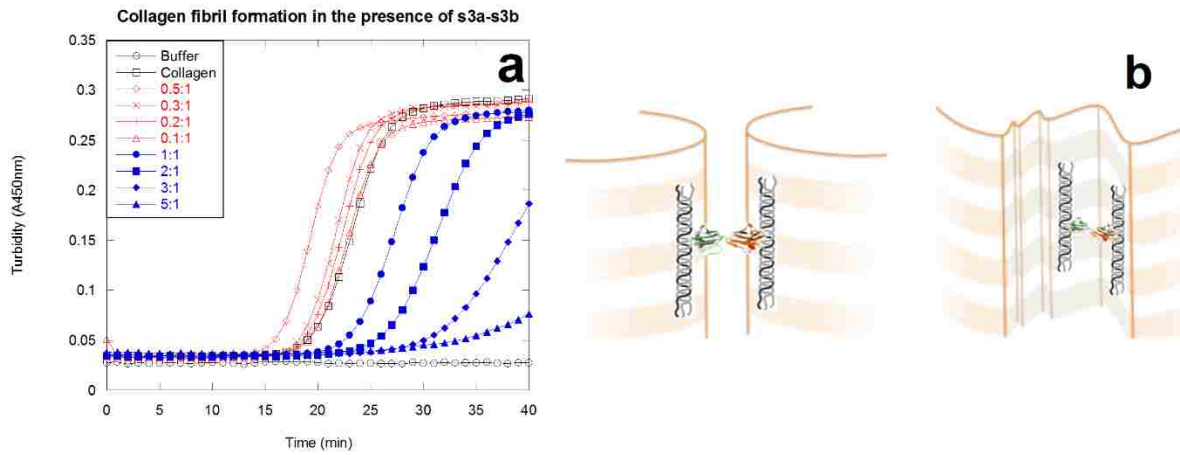
Fig 2. Putative binding mechanism for tandem CBD. (a) Promotion (red) and retardation (blue) of collagen fibril formation by tandem CBD. (b) Schematic drawing of putative collagen niches preferred by tandem CBD.

Table

Supplementary Table 1: Turbidity parameters of self-assembly of collagen in a presence of collagen binding domain

		Tandem CBD to collagen molar ratios							
	collagen	0.1:1	0.2:1	0.3:1	0.5:1	1:1	2:1	3:1	5:1
t_{lag} (min)	19.0	19.0	18.0	17.0	15.0	21.0	24.0	27.0	30.0
V (turbidity/min)	0.019	0.022	0.022	0.021	0.019	0.017	0.014	0.010	0.005
ΔA (turbidity)	0.284	0.266	0.271	0.277	0.268	0.274	0.287	0.291	0.331

		s3a to collagen molar ratios							
	collagen	0.1:1	0.2:1	0.3:1	0.5:1	1:1	2:1	3:1	5:1
t_{lag} (min)	11.6	15.3	13.6	15.9	11.2	17.0	19.0	21.6	15.6
V_{max} (turbidity/min)	0.014	0.027	0.013	0.014	0.016	0.024	0.016	0.022	0.011
ΔA (turbidity)	0.256	0.336	0.282	0.246	0.249	0.270	0.276	0.272	0.293

		s3b to collagen molar ratios							
	collagen	0.1:1	0.2:1	0.3:1	0.5:1	1:1	2:1	3:1	5:1
t_{lag} (min)	19.9	20.3	20.6	20.4	20.5	25.1	32.9	42.6	68.3
V_{max} (turbidity/min)	0.041	0.035	0.034	0.033	0.030	0.027	0.018	0.013	0.008
ΔA (turbidity)	0.289	0.281	0.283	0.278	0.264	0.289	0.314	0.337	0.341

Figures
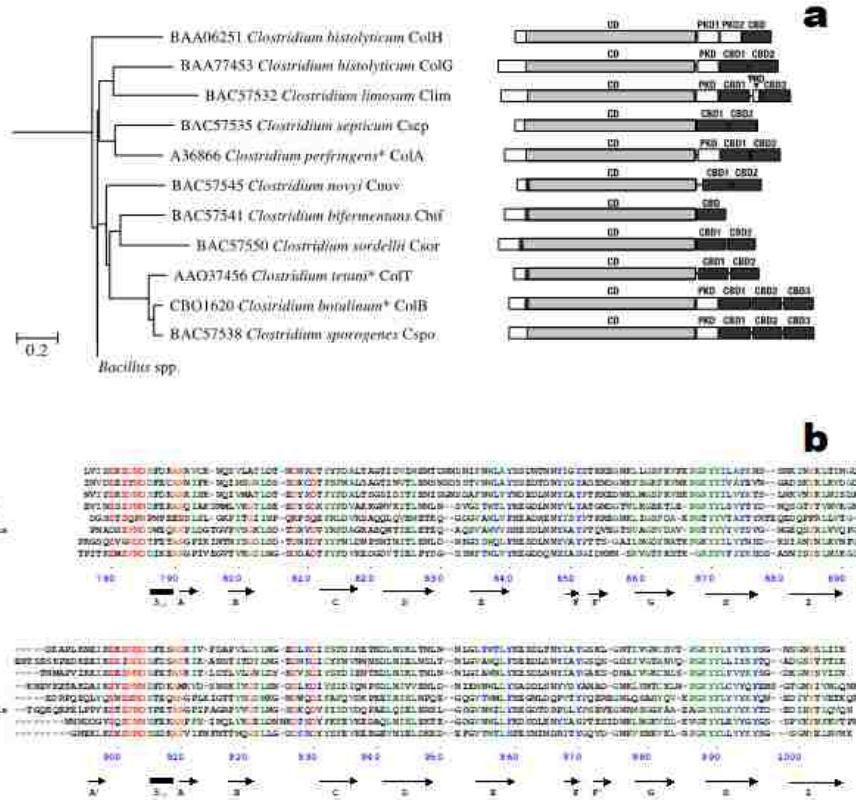
Supplementary Fig. 1 (a) Phylogenic tree of M9B collagenases (InterPro 002169). (b) Sequence alignment of tandem collagen binding domains in M9B collagenases from *Clostridium sporogenes*, *C. botulinum* A3, *C. perfringens*, *Bacillus. brevis*, *Paenibacillus dendritiformis*, *C. sordellii* and *C. histolyticum*. N-side CBD molecules are aligned on the top rows, and C-side CBDs are aligned on the bottom rows. Calcium-binding residues (chelator in red and spotter of chelator in orange), structurally important residues (green) are conserved. Based on biophysical studies completed for s3b (*29* ENREF_29*, 45, 66*), residues that conjectured to interact with mini-collagen are highlighted in blue. One of the most well conserved and critical residues for collagen interaction in s3b is Tyr996. Tyr996 equivalent is conserved in the C-side CBD but not in the N-side CBD.
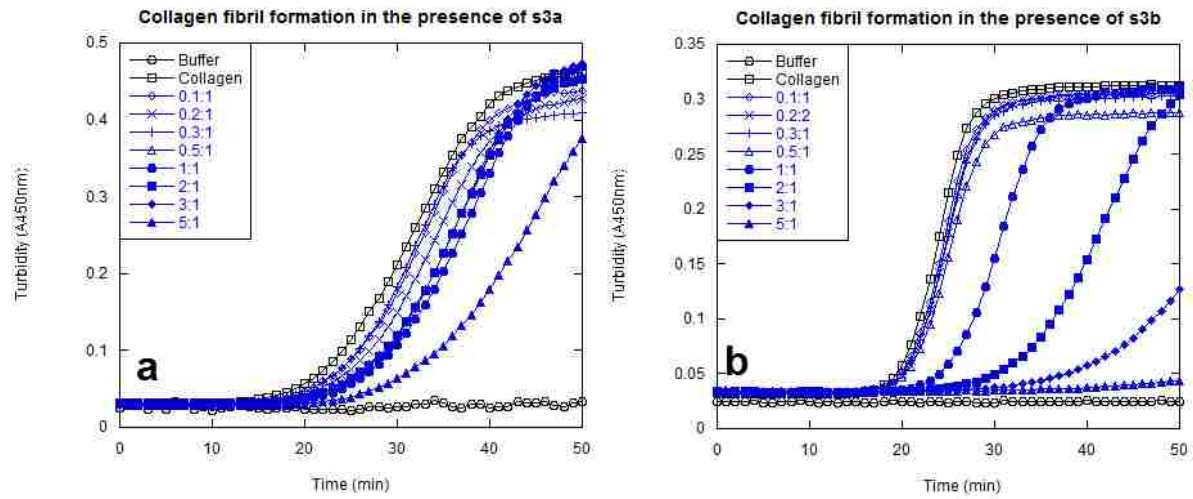
Supplementary Fig. 2 (a) Retardation of collagen fibril formation by s3a (N-side CBD). (b) Retardation of collagen fibril formation by s3b (C-side CBD).

Chapter 5: Conclusion

The crystal structures of Ca^{2+}-bound s2a, s2b, s3, s3a-s3b and Ca^{2+}-absent s2a and s2

facilitate understanding of the division of labor utilized by collagenases ColG and ColH to

degrade collagen. Over time, evolutionary pressure appears to have tuned the role of homologous

domains to maintain Ca^{2+}-triggered activation, yet also initially recognize unique regions on

collagen. Initiating collagen degradation at unique regions would allow the collagenases to

synergistically work towards collagen dismantling. The novel crystal structure of the Ca^{2+}-bound

(*holo*)-s3 closely resembles the higher resolution, re-refined structure of holo-s3b, and shares

both Ca^{2+}-induced stabilization as well as the architecture of the collagen-binding cleft.

However, charge differences on the surface of the collagen-binding cleft may allow the domains

to initiate binding at different regions. The PKD-like domain s2b is not a collagen-binder, but it

has been shown to tighten the interaction between s3 and collagen. Curiously, the role of s2b as a

source of secondary, weak interaction that improves binding to collagen is analogous to the role

of s3a. While s3a is a CBD, its collagen-binding cleft lacks a key tyrosine found in s3b. Both s2b

and s3a allow the main CBD to bind tighter, but not become locked into a single site. The

tandem CBD structure represents the first structure of two domains from collagenase. With this

structure, the Ca^{2+}-induced domain rearrangement could be evaluated using SAXS. The

molecular envelopes obtained using this technique indicate the domain transforms from an

elongated, flexible confirmation to the compact, pseudo-two-fold arrangement seen in the crystal

structure. In correlation with SAXS studies of ColH, the domains of the apo-collagenase inside

the bacterium appear to be connected by flexible linkers that can facilitate secretion from the cell

into the host extracellular matrix. Here, the enzyme must resist proteases secreted by the host

immune system. The domains therefore likely utilize the higher Ca^{2+} concentrations of the

extracellular matrix to trigger domain rearrangement into the compact *holo*-enzyme, which is expected to minimize protease targets.

Development of applications of the targeting segments of ColG and ColH are aided by structural and functional studies that identify the role of targeting segment domains. Systemic delivery of a therapeutic, for example, in the treatment of ostroporosis could benefit from use of collagen-binders that do not bind ultra-tight to collagen. Localized delivery, which would be used in the treatment of ailments such as bone fracture, necessitates minimal collagen-binder distribution. The work described in this dissertation indicates that CBD, PKD-CBD, and tandem CBD are extremely stable proteins that are capable of withstanding *in vivo* conditions. Moreover, these domains are likely to target the damaged regions of collagen that are induced by the ailment. Selection of collagen-binding segment for the therapeutic delivery can be guided by binding affinity of the segment.

References

1.   Nagata, K. (1996) Hsp47: a collagen-specific molecular chaperone, *Trends Biochem Sci 21*, 22-26.

2.   Prockop, D. J., and Kivirikko, K. I. (1995) Collagens: molecular biology, diseases, and potentials for therapy, *Annu Rev Biochem 64*, 403-434.

3.   Ricard-Blum, S., and Ruggiero, F. (2005) The collagen superfamily: from the extracellular matrix to the cell membrane, *Pathol Biol (Paris) 53*, 430-442.

4.   Orgel, J. P., Irving, T. C., Miller, A., and Wess, T. J. (2006) Microfibrillar structure of type I collagen in situ, *Proc Natl Acad Sci U S A 103*, 9001-9005.

5.   Perumal, S., Antipova, O., and Orgel, J. P. (2008) Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis, *Proc Natl Acad Sci U S A 105*, 2824-2829.

6.   Nagase, H., Visse, R., and Murphy, G. (2006) Structure and function of matrix metalloproteinases and TIMPs, *Cardiovasc Res 69*, 562-573.

7.   Van Wart, H. E., and Birkedal-Hansen, H. (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family, *Proc Natl Acad Sci U S A 87*, 5578-5582.

8.   Jung, C. M., Matsushita, O., Katayama, S., Minami, J., Sakurai, J., and Okabe, A. (1999) Identification of metal ligands in the Clostridium histolyticum ColH collagenase, *J Bacteriol 181*, 2816-2822.

9.   Jongeneel, C. V., Bouvier, J., and Bairoch, A. (1989) A unique signature identifies a family of zinc-dependent metallopeptidases, *FEBS Lett 242*, 211-214.

10.   Arnold, L. H., Butt, L. E., Prior, S. H., Read, C. M., Fields, G. B., and Pickford, A. R. (2011) The Interface between Catalytic and Hemopexin Domains in Matrix Metalloproteinase-1 Conceals a Collagen Binding Exosite, *Journal of Biological Chemistry 286*, 45073-45082.

11.   Jozic, D., Bourenkov, G., Lim, N. H., Visse, R., Nagase, H., Bode, W., and Maskos, K. (2005) X-ray structure of human proMMP-1: new insights into procollagenase activation and collagen binding, *J Biol Chem 280*, 9578-9585.

12.   Iyer, S., Visse, R., Nagase, H., and Acharya, K. R. (2006) Crystal structure of an active form of human MMP-1, *J Mol Biol 362*, 78-88.

13.   Overall, C. M. (2002) Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites, *Mol Biotechnol 22*, 51-86.

14.   Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R., and Nagase, H. (2004) Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis, *EMBO J 23*, 3020-3030.

15.   Lauer-Fields, J. L., Chalmers, M. J., Busby, S. A., Minond, D., Griffin, P. R., and Fields, G. B. (2009) Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity, *J Biol Chem 284*, 24017-24024.

16.   Mallya, S. K., Mookhtiar, K. A., and Van Wart, H. E. (1986) Accurate, quantitative assays for the hydrolysis of soluble type I, II, and III 3H-acetylated collagens by bacterial and tissue collagenases, *Anal Biochem 158*, 334-345.

17.   French, M. F., Bhown, A., and Van Wart, H. E. (1992) Identification of Clostridium histolyticum collagenase hyperreactive sites in type I, II, and III collagens: lack of correlation with local triple helical stability, *J Protein Chem 11*, 83-97.

18.   Duarte, A. S., Correia, A., and Esteves, A. C. (2014) Bacterial collagenases - A review, *Crit Rev Microbiol*.

19.   Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) MEROPS: the peptidase database, *Nucleic Acids Res 38*, D227-233.

20.   Lee, J. H., Kim, G. T., Lee, J. Y., Jun, H. K., Yu, J. H., and Kong, I. S. (1998) Isolation and sequence analysis of metalloprotease gene from Vibrio mimicus, *Biochim Biophys Acta 1384*, 1-6.

21.   Lee, C. Y., Su, S. C., and Liaw, R. B. (1995) Molecular analysis of an extracellular protease gene from Vibrio parahaemolyticus, *Microbiology 141 (Pt 10)*, 2569-2576.

22.   Takeuchi, H., Shibano, Y., Morihara, K., Fukushima, J., Inami, S., Keil, B., Gilles, A. M., Kawamoto, S., and Okuda, K. (1992) Structural gene and complete amino acid sequence of Vibrio alginolyticus collagenase, *Biochem J 281 (Pt 3)*, 703-708.

23.   Kim, S. K., Yang, J. Y., and Cha, J. (2002) Cloning and sequence analysis of a novel metalloprotease gene from Vibrio parahaemolyticus 04, *Gene 283*, 277-286.

24.   Luan, X., Chen, J., Zhang, X. H., Li, Y., and Hu, G. (2007) Expression and characterization of a metalloprotease from a Vibrio parahaemolyticus isolate, *Can J Microbiol 53*, 1168-1173.

25.   Hickman, F. W., Farmer, J. J., 3rd, Hollis, D. G., Fanning, G. R., Steigerwalt, A. G., Weaver, R. E., and Brenner, D. J. (1982) Identification of Vibrio hollisae sp. nov. from patients with diarrhea, *J Clin Microbiol 15*, 395-401.

26.   Thompson, F. L., Hoste, B., Vandemeulebroecke, K., and Swings, J. (2003) Reclassification of Vibrio hollisae as Grimontia hollisae gen. nov., comb. nov., *Int J Syst Evol Micr 53*, 1615-1617.

27. Teramura, N., Tanaka, K., Iijima, K., Hayashida, O., Suzuki, K., Hattori, S., and Irie, S. (2011) Cloning of a novel collagenase gene from the gram-negative bacterium Grimontia (Vibrio) hollisae 1706B and its efficient expression in Brevibacillus choshinensis, *J Bacteriol 193*, 3049-3056.

28. Lecroisey, A., and Keil, B. (1979) Differences in the degradation of native collagen by two microbial collagenases, *Biochem J 179*, 53-58.

29. Wilson, J. J., Matsushita, O., Okabe, A., and Sakon, J. (2003) A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation, *EMBO J 22*, 1743-1752.

30. Philominathan, S. T., Matsushita, O., Gensure, R., and Sakon, J. (2009) Ca2+-induced linker transformation leads to a compact and rigid collagen-binding domain of Clostridium histolyticum collagenase, *FEBS J 276*, 3589-3601.

31. Eckhard, U., Schonauer, E., Nuss, D., and Brandstetter, H. (2011) Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis, *Nat Struct Mol Biol 18*, 1109-1114.

32. Breite, A. G., McCarthy, R. C., and Dwulet, F. E. (2011) Characterization and functional assessment of Clostridium histolyticum class I (C1) collagenases and the synergistic degradation of native collagen in enzyme mixtures containing class II (C2) collagenase, *Transplant Proc 43*, 3171-3175.

33. Watanabe-Nakayama, T., Itami, M., Kodera, N., Ando, T., and Konno, H. (2016) High-speed atomic force microscopy reveals strongly polarized movement of clostridial collagenase along collagen fibrils, *Sci Rep 6*, 28975.

34. Craveur, P., Joseph, A. P., Poulain, P., de Brevern, A. G., and Rebehmed, J. (2013) Cis-trans isomerization of omega dihedrals in proteins, *Amino Acids 45*, 279-289.

35. Jabs, A., Weiss, M. S., and Hilgenfeld, R. (1999) Non-proline cis peptide bonds in proteins, *J Mol Biol 286*, 291-304.

36. Croll, T. I. (2015) The rate of cis-trans conformation errors is increasing in low-resolution crystal structures, *Acta Crystallogr D Biol Crystallogr 71*, 706-709.

37. Spiriti, J., and van der Vaart, A. (2010) Mechanism of the calcium-induced trans-cis isomerization of a non-prolyl peptide bond in Clostridium histolyticum collagenase, *Biochemistry 49*, 5314-5320.

38. Hurst, L. C., Badalamente, M. A., Hentz, V. R., Hotchkiss, R. N., Kaplan, F. T., Meals, R. A., Smith, T. M., and Rodzvilla, J. (2009) Injectable collagenase clostridium histolyticum for Dupuytren's contracture, *N Engl J Med 361*, 968-979.

39. Wang, Y. K., Zhao, G. Y., Li, Y., Chen, X. L., Xie, B. B., Su, H. N., Lv, Y. H., He, H. L., Liu, H., Hu, J., Zhou, B. C., and Zhang, Y. Z. (2010) Mechanistic insight into the

function of the C-terminal PKD domain of the collagenolytic serine protease deseasin MCP-01 from deep sea Pseudoalteromonas sp. SM9913: binding of the PKD domain to collagen results in collagen swelling but does not unwind the collagen triple helix, *J Biol Chem 285*, 14285-14291.

40.   Matsushita, O., Jung, C. M., Minami, J., Katayama, S., Nishi, N., and Okabe, A. (1998) A study of the collagen-binding domain of a 116-kDa Clostridium histolyticum collagenase, *J Biol Chem 273*, 3643-3648.

41.   Yeats, C., Bentley, S., and Bateman, A. (2003) New knowledge from old: in silico discovery of novel protein domains in Streptomyces coelicolor, *BMC Microbiol 3*, 3.

42.   Yan, B. Q., Chen, X. L., Hou, X. Y., He, H., Zhou, B. C., and Zhang, Y. Z. (2009) Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium Pseudoalteromonas sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains, *Extremophiles 13*, 725-733.

43.   Matsushita, O., Koide, T., Kobayashi, R., Nagata, K., and Okabe, A. (2001) Substrate recognition by the collagen-binding domain of Clostridium histolyticum class I collagenase, *J Biol Chem 276*, 8761-8770.

44.   Toyoshima, T., Matsushita, O., Minami, J., Nishi, N., Okabe, A., and Itano, T. (2001) Collagen-binding domain of a Clostridium histolyticum collagenase exhibits a broad substrate spectrum both in vitro and in vivo, *Connect Tissue Res 42*, 281-290.

45.   Philominathan, S. T., Koide, T., Hamada, K., Yasui, H., Seifert, S., Matsushita, O., and Sakon, J. (2009) Unidirectional binding of clostridial collagenase to triple helical substrates, *J Biol Chem 284*, 10868-10876.

46.   Nishi, N., Matsushita, O., Yuube, K., Miyanaka, H., Okabe, A., and Wada, F. (1998) Collagen-binding growth factors: production and characterization of functional fusion proteins having a collagen-binding domain, *Proc Natl Acad Sci U S A 95*, 7018-7023.

47.   Katikaneni, R., Ponnapakkam, T., Suda, H., Miyata, S., Sakon, J., Matsushita, O., and Gensure, R. C. (2012) Treatment for chemotherapy-induced alopecia in mice using parathyroid hormone agonists and antagonists linked to a collagen binding domain, *Int J Cancer 131*, E813-821.

48.   Ponnapakkam, T., Katikaneni, R., Miller, E., Ponnapakkam, A., Hirofumi, S., Miyata, S., Suva, L. J., Sakon, J., Matsushita, O., and Gensure, R. C. (2011) Monthly administration of a novel PTH-collagen binding domain fusion protein is anabolic in mice, *Calcif Tissue Int 88*, 511-520.

49.   Ponnapakkam, T., Katikaneni, R., Nichols, T., Tobin, G., Sakon, J., Matsushita, O., and Gensure, R. C. (2011) Prevention of chemotherapy-induced osteoporosis by cyclophosphamide with a long-acting form of parathyroid hormone, *J Endocrinol Invest 34*, e392-397.

50. Ponnapakkam, T., Katikaneni, R., Suda, H., Miyata, S., Matsushita, O., Sakon, J., and Gensure, R. C. (2012) A single injection of the anabolic bone agent, parathyroid hormone-collagen binding domain (PTH-CBD), results in sustained increases in bone mineral density for up to 12 months in normal female mice, *Calcif Tissue Int 91*, 196-203.

51. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

52. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240-255.

53. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes, *Acta Crystallographica Section D-Biological Crystallography 62*, 859-866.

54. Sheldrick, G. M. (2008) A short history of SHELX, *Acta Crystallogr A 64*, 112-122.

55. McRee, D. E. (1999) *Practical protein crystallography*, 2nd ed., Academic, San Diego, Calif. ; London.

56. Weigand, S., B. Stillwell, W. Guise, J. Quintana, and D. Keanne. (2009) Flexibility and high throughput: supporting SAXS users at a joint industrial academic beamline, *Adv X-Ray Anal 52*, 58-68.

57. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Cryst. 25*, 495-503.

58. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering, *Biophys J 80*, 2946-2953.

59. Volkov, V., and D. Svergun (2003) Uniqueness of ab initio shape determination in small-angle scattering, *J. Appl. Cryst 36*, 860-864.

60. Kozin, M., and D. Svergun. (2000) A software system for rigid-body modelling of solution scattering data, *J. Appl. Cryst 33*, 775-777.

61. Svergun, D., C. Barberato, and M. Koch. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *J. Appl. Cryst 28*, 768-773.

62. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophys J 89*, 1237-1250.

63. Palma, P. N., Krippahl, L., Wampler, J. E., and Moura, J. J. (2000) BiGGER: a new (soft) docking algorithm for predicting protein interactions, *Proteins 39*, 372-384.

64. Sides, C. R., Liyanage, R., Lay, J. O., Jr., Philominathan, S. T., Matsushita, O., and Sakon, J. (2012) Probing the 3-D structure, dynamics, and stability of bacterial collagenase collagen binding domain (apo- versus holo-) by limited proteolysis MALDI-TOF MS, *J Am Soc Mass Spectrom 23*, 505-519.

65. Kirberger, M., Wang, X., Deng, H., Yang, W., Chen, G., and Yang, J. J. (2008) Statistical analysis of structural characteristics of protein Ca2+-binding sites, *J Biol Inorg Chem 13*, 1169-1181.

66. Philominathan, S. T., Koide, T., Matsushita, O., and Sakon, J. (2012) Bacterial collagen-binding domain targets undertwisted regions of collagen, *Protein Sci 21*, 1554-1565.

67. Berman, H. M., Westbrook, J. D., Gabanyi, M. J., Tao, W., Shah, R., Kouranov, A., Schwede, T., Arnold, K., Kiefer, F., Bordoli, L., Kopp, J., Podvinec, M., Adams, P. D., Carter, L. G., Minor, W., Nair, R., and La Baer, J. (2009) The protein structure initiative structural genomics knowledgebase, *Nucleic Acids Res 37*, D365-368.

68. Ding, J., Choe, H. W., Granzin, J., and Saenger, W. (1992) Structure of ribonuclease T1 complexed with zinc(II) at 1.8 A resolution: a Zn2+.6H2O.carboxylate clathrate, *Acta crystallographica. Section B, Structural science 48 (Pt 2)*, 185-191.

69. Zhou, Y., Morais-Cabral, J. H., Kaufman, A., and MacKinnon, R. (2001) Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution, *Nature 414*, 43-48.

70. Parkinson, G. N., Lee, M. P., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature 417*, 876-880.

71. Gros, P., Kalk, K. H., and Hol, W. G. (1991) Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium, *The Journal of biological chemistry 266*, 2953-2961.

72. Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L., and Woldike, H. F. (1990) Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1-A resolution of two enzymes from Aspergillus, *Biochemistry 29*, 6244-6249.

73. Betzel, C., Pal, G. P., and Saenger, W. (1988) Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 A resolution, *Acta Crystallogr B 44 (Pt 2)*, 163-172.

74. Hubbard, S. J., and J.M. Thronton (1993) NACCESS, *Department of Biochemistry and Molecular Biology, University of College, London*.

75. Bella, J. (2010) A new method for describing the helical conformation of collagen: dependence of the triple helical twist on amino acid sequence, *J Struct Biol 170*, 377-391.

76. Miyoshi, S. (2006) Vibrio vulnificus infection and metalloprotease, *J Dermatol 33*, 589-595.

77. Nonaka, T., Fujihashi, M., Kita, A., Saeki, K., Ito, S., Horikoshi, K., and Miki, K. (2004) The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain, *J Biol Chem 279*, 47344-47351.

78. Yu, M. S., and Lee, C. Y. (1999) Expression and characterization of the prtV gene encoding a collagenase from Vibrio parahaemolyticus in Escherichia coli, *Microbiology 145 (Pt 1)*, 143-150.

79. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res 22*, 4673-4680.

80. Bauer, R., Wilson, J. J., Philominathan, S. T., Davis, D., Matsushita, O., and Sakon, J. (2013) Structural comparison of ColH and ColG collagen-binding domains from Clostridium histolyticum, *J Bacteriol 195*, 318-327.

81. Ohbayashi, N., Matsumoto, T., Shima, H., Goto, M., Watanabe, K., Yamano, A., Katoh, Y., Igarashi, K., Yamagata, Y., and Murayama, K. (2013) Solution structure of clostridial collagenase h and its calcium-dependent global conformation change, *Biophys J 104*, 1538-1545.

82. Eckhard, U., Schonauer, E., and Brandstetter, H. (2013) Structural basis for activity regulation and substrate preference of clostridial collagenases g, h, and T, *J Biol Chem 288*, 20184-20194.

83. Adhikari, A. S., Glassey, E., and Dunn, A. R. (2012) Conformational dynamics accompanying the proteolytic degradation of trimeric collagen I by collagenases, *J Am Chem Soc 134*, 13259-13265.

84. Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Toccafondi, M., Lauer, J. L., and Fields, G. B. (2012) Structural basis for matrix metalloproteinase 1-catalyzed collagenolysis, *J Am Chem Soc 134*, 2100-2110.

85. Nagase, H., and Fushimi, K. (2008) Elucidating the function of non catalytic domains of collagenases and aggrecanases, *Connect Tissue Res 49*, 169-174.

86. Welgus, H. G., Jeffrey, J. J., and Eisen, A. Z. (1981) Human skin fibroblast collagenase. Assessment of activation energy and deuterium isotope effect with collagenous substrates, *J Biol Chem 256*, 9516-9521.

87. Welgus, H. G., Jeffrey, J. J., and Eisen, A. Z. (1981) The collagen substrate specificity of human skin fibroblast collagenase, *J Biol Chem 256*, 9511-9515.

88. Fields, G. B. (2013) Interstitial collagen catabolism, *J Biol Chem 288*, 8785-8793.

117

89. Eckhard, U., and Brandstetter, H. (2011) Polycystic kidney disease-like domains of clostridial collagenases and their role in collagen recruitment, *Biol Chem 392*, 1039-1045.

90. McCarthy, R. C., Breite, A. G., Green, M. L., and Dwulet, F. E. (2011) Tissue dissociation enzymes for isolating human islets for transplantation: factors to consider in setting enzyme acceptance criteria, *Transplantation 91*, 137-145.

91. Fujio, A., Murayama, K., Yamagata, Y., Watanabe, K., Imura, T., Inagaki, A., Ohbayashi, N., Shima, H., Sekiguchi, S., Fujimori, K., Igarashi, K., Ohuchi, N., Satomi, S., and Goto, M. (2013) Collagenase H is Crucial for Isolation of Rat Pancreatic Islets, *Transplantation 96*, S141-S142.

92. Akimoto, M., Takeda, A., Matsushita, O., Inoue, J., Sakamoto, K., Hattori, M., Kounoike, N., and Uchinuma, E. (2013) Effects of CB-VEGF-A injection in rat flap models for improved survival, *Plast Reconstr Surg 131*, 717-725.

93. Uchida, K., Matsushita, O., Naruse, K., Mima, T., Nishi, N., Hattori, S., Ogura, T., Inoue, G., Tanaka, K., and Takaso, M. (2013) Acceleration of periosteal bone formation by human basic fibroblast growth factor containing a collagen-binding domain from Clostridium histolyticum collagenase, *J Biomed Mater Res A*.

94. Saito, W., Uchida, K., Ueno, M., Matsushita, O., Inoue, G., Nishi, N., Ogura, T., Hattori, S., Fujimaki, H., Tanaka, K., and Takaso, M. (2013) Acceleration of bone formation during fracture healing by injectable collagen powder and human basic fibroblast growth factor containing a collagen-binding domain from Clostridium histolyticum collagenase, *J Biomed Mater Res A*.

95. Ponnapakkam, T., Katikaneni, R., Sakon, J., Stratford, R., and Gensure, R. C. (2013) Treating osteoporosis by targeting parathyroid hormone to bone, *Drug Discov Today*.

96. Katikaneni, R., Ponnapakkam, T., Matsushita, O., Sakon, J., and Gensure, R. (2014) Treatment and prevention of chemotherapy-induced alopecia with PTH-CBD, a collagen-targeted parathyroid hormone analog, in a non-depilated mouse model, *Anticancer Drugs 25*, 30-38.

97. Philominathan, S. T., Matsushita, O., Jordan, J. B., and Sakon, J. (2008) 1H, 13C and 15N resonance assignments of Ca2+ bound collagen-binding domain derived from a clostridial collagenase, *Biomol NMR Assign 2*, 127-129.

98. Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N., and Fontes, C. M. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains, *J Biol Chem 281*, 8815-8828.

99. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection, *Acta Crystallogr D Biol Crystallogr 55*, 1718-1725.

100. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat Struct Biol 6*, 458-463.

101. Merritt, E. A. (2012) To B or not to B: a question of resolution?, *Acta Crystallogr D Biol Crystallogr 68*, 468-477.

102. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21.

103. Hemmingsen, J. M., Gernert, K. M., Richardson, J. S., and Richardson, D. C. (1994) The tyrosine corner: a feature of most Greek key beta-barrel proteins, *Protein Sci 3*, 1927-1937.

104. Schormann, N., Murrell, J. R., Liepnieks, J. J., and Benson, M. D. (1995) Tertiary structure of an amyloid immunoglobulin light chain protein: a proposed model for amyloid fibril formation, *Proc Natl Acad Sci U S A 92*, 9490-9494.

105. Eyal, E., Yang, L. W., and Bahar, I. (2006) Anisotropic network model: systematic evaluation and a new web interface, *Bioinformatics 22*, 2619-2627.

106. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions, *Protein Sci 3*, 1984-1991.

107. Kellis, J. T., Jr., Todd, R. J., and Arnold, F. H. (1991) Protein stabilization by engineered metal chelation, *Biotechnology (N Y) 9*, 994-995.

108. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic Acids Res 34*, W116-118.

109. Bycroft, M., Bateman, A., Clarke, J., Hamill, S. J., Sandford, R., Thomas, R. L., and Chothia, C. (1999) The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease, *EMBO J 18*, 297-305.

110. Jing, H., Takagi, J., Liu, J. H., Lindgren, S., Zhang, R. G., Joachimiak, A., Wang, J. H., and Springer, T. A. (2002) Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins, *Structure 10*, 1453-1464.

111. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Res 42*, W252-258.

112. Bauer, R., Janowska, K., Taylor, K., Jordan, B., Gann, S., Janowski, T., Latimer, E. C., Matsushita, O., and Sakon, J. (2015) Structures of three polycystic kidney disease-like domains from Clostridium histolyticum collagenases ColG and ColH, *Acta Crystallogr D Biol Crystallogr 71*, 565-577.

113. Gaston, R. G., Larsen, S. E., Pess, G. M., Coleman, S., Dean, B., Cohen, B. M., Kaufman, G. J., Tursi, J. P., and Hurst, L. C. (2015) The Efficacy and Safety of Concurrent Collagenase Clostridium Histolyticum Injections for 2 Dupuytren Contractures in the Same Hand: A Prospective, Multicenter Study, *J Hand Surg-Am 40*, 1963-1971.

114. Katikaneni, R., Ponnapakkam, T., Seymour, A., Sakon, J., and Gensure, R. (2014) Parathyroid hormone linked to a collagen binding domain promotes hair growth in a mouse model of chemotherapy-induced alopecia in a dose-dependent manner, *Anticancer Drugs 25*, 819-825.

115. Katikaneni, R., Seymour, A. W., Gulati, R., Ponnapakkam, T., and Gensure, R. C. (2015) Therapy for Alopecia Areata in Mice by Stimulating the Hair Cycle with Parathyroid Hormone Agonists Linked to a Collagen-Binding Domain, *J Investig Dermatol Symp Proc 17*, 13-15.

116. Saito, W., Uchida, K., Matsushita, O., Inoue, G., Sekiguchi, H., Aikawa, J., Fujimaki, H., and Takaso, M. (2015) Acceleration of callus formation during fracture healing using basic fibroblast growth factor-kidney disease domain-collagen-binding domain fusion protein combined with allogenic demineralized bone powder, *J Orthop Surg Res 10*, 59.

117. Uchida, K., Matsushita, O., Nishi, N., Inoue, G., Horikawa, K., and Takaso, M. (2017) Enhancement of periosteal bone formation by basic fibroblast-derived growth factor containing polycystic kidney disease and collagen-binding domains from Clostridium histolyticum collagenase, *J Tissue Eng Regen Med 11*, 1165-1172.

118. An, B., Lin, Y. S., and Brodsky, B. (2016) Collagen interactions: Drug design and delivery, *Adv Drug Deliv Rev 97*, 69-84.

119. Matthews, B. W. (1968) Solvent content of protein crystals, *J Mol Biol 33*, 491-497.

120. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures, *Journal of Applied Crystallography 26*, 283-291.

121. Bers, D. M., Patton, C. W., and Nuccitelli, R. (2010) A Practical Guide to the Preparation of Ca2+ a Buffers, *Calcium in Living Cells 99*, 1-26.

122. Dyer, K. N., Hammel, M., Rambo, R. P., Tsutakawa, S. E., Rodic, I., Classen, S., Tainer, J. A., and Hura, G. L. (2014) High-throughput SAXS for the characterization of biomolecules in solution: a practical approach, *Methods in molecular biology 1091*, 245-258.

123. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A., and Tainer, J. A. (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source, *Journal of Applied Crystallography 46*, 1-13.

124. Okano-Kosugi, H., Matsushita, O., Asada, S., Herr, A. B., Kitagawa, K., and Koide, T. (2009) Development of a high-throughput screening system for the compounds that inhibit collagen-protein interactions, *Anal Biochem 394*, 125-131.

125. Holland, I. B., Jones, H. E., Campbell, A. K., and Jacq, A. (1999) An assessment of the role of intracellular free Ca2+ in E. coli, *Biochimie 81*, 901-907.

126. Maurer, P., and Hohenester, E. (1997) Structural and functional aspects of calcium binding in extracellular matrix proteins, *Matrix Biol 15*, 569-580; discussion 581.

127. Kuwaba, K., Kobayashi, M., Nomura, Y., Irie, S., and Koyama, Y. (2001) Elongated dermatan sulphate in post-inflammatory healing skin distributes among collagen fibrils separated by enlarged interfibrillar gaps, *Biochem J 358*, 157-163.

128. Bauer, R., Sakon, J., Tanaka, K., Janowska, K., Roeser, J. R., Matsushita, O., and Uchida, K. (2017) Collagen-binding agent compositions and methods of using the same., The Board of Trustees of the University of Arkansas, USA Provisional Patent 62/457,410.