Collagen and Collagen-Associated Proteins:
Studies on the Type XV Trimerization Domain, Prolyl 3-hydroxylase and the Prolyl 3-hydroxylase:Cartilage Associated Protein:Cyclophilin B Complex

By

Jacqueline Alani Wirz

A Dissertation

Presented to the Department of Biochemistry & Molecular Biology
and the Oregon Health & Science University School of Medicine
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

July 20th, 2010

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Jacqueline Alani Wirz

has been approved

_____
Hans Peter Bächinger
Mentor

_____
Lynn Sakai
Chair

_____
Michael Chapman

_____
David Farrens

_____
Peter Rotwein

TABLE OF CONTENTS

Acknowledgements

The sun has left and forgotten me
It's dark, I cannot see
Why does this rain pour down
I'm gonna drown
In a sea
Of deep confusion

Somebody told me, I don't know who
Whenever you are sad and blue
And you're feelin' all alone and left behind
Just take a look inside and you will find

You gotta hold on, hold on through the night
Hang on, things will be all right
Even when it's dark
And not a bit of spark
Sing-song sunshine from above
Spreading rays of sunny love

Just hang on, hang on to the vine
Stay on, soon you'll be divine
If you start to cry, look up to the sky
Something's coming up ahead
To turn your tears to dew instead

And so I hold on to his advice
When change is hard and not so nice
You listen to your heart the whole night through
Your sunny someday will come one day soon to you

"Hang On Little Tomato"
Pink Martini

Abstract

Collagen is one of the most abundant proteins in the human body and plays an important role in the structural stability of many tissues. Collagen-associated proteins are essential for the correct biosynthesis of collagen. Defects in either collagen or collagen-associated proteins result in a myriad of diseases, underscoring the importance of both classes of molecules. Although much work has been performed on the triple helical domains found in all collagens, less is known about their non-collagenous domains. Here we report the crystal structure and biochemical characterization of the non-collagenous domain responsible for the trimerization of human type XV collagen. The structure is stabilized by the presence of hydrophobic cores in the individual monomeric chain as well as in the trimer assembly. The trimer is also exceptionally stable and forms at picomolar concentrations. Many non-collagenous proteins are associated with collagens, including a variety of enzymes and chaperones. We present data demonstrating that the resident endoplasmic reticulum protein complex consisting of prolyl 3-hydroxylase 1 (P3H1), cartilage associated protein (CRTAP) and cyclophilin B (CypB) performs three functions: the complex is a prolyl 3-hydroxylase, a molecular chaperone and a protein disulfide isomerase. To perform these functions, the complex is capable of binding both denatured and triple helical collagen. Attempts to overexpress P3H1 or CRTAP for further studies were unproductive as the proteins were either insoluble or not in biologically relevant conformations. The P3H1●CRTAP●CypB complex is purified out of chick embryos, and experiments were performed to enhance protein yield. Although some progress was made, the overall stability of the complex remains low, confounding attempts to perform more detailed structural analysis.

List of Figures and Tables

Chapter 5: P3H1 and P3H1●CRTAP●CypB Protein Production:
Recombinant Protein Expression and Protein Preparation from Whole Chick Extracts

Chapter 6: Summary, Future Directions and Conclusions

Abbreviations

BM, basement membrane

CCD, charge coupled device

CLE-1, *Caenorhabditis elegans* collagen XVIII homologue

COL, collagenous

CRTAP, cartilage associated protein (Casp)

CypB, cyclophilin B

Dmp, *Drosophila melanogaster* multiplexin

FACIT, fibril associated collagens with interrupted triple helices

FKBP, FK506 binding protein

GAG, glycosaminoglycan

GRCB, gene rich cluster, B

Gros1, growth suppressor on chromosome 1

GSH/GSSH, reduced/oxidized glutathione

GuHCl, guanidine hydrochloride

HSP47, heat shock protein 47

Leprecan, leucine proline-enriched proteoglycan

LH, lysyl hydroxylase

MAD, multiwavelength anomalous dispersion

MCA, mehtyl coumaryl amide

MLAT4, myxoid liposarcoma associated protein 4

MMP, matrix metalloproteinase

MR, molecular replacement

Multiplexin, collagens with multiple triple helical domains with interruptions

NC, non-collagenous

OI, osteogenesis imperfecta

PPIase, peptidyl-prolyl *cis-trans* isomerase

PBS, phosphate-buffered saline

pN type III collagen, type III collagen containing the amino terminal propeptide

P3H, prolyl 3-hydroxylase

P3H1, prolyl 3-hydroxylase 1

P3H2, prolyl 3-hydroxylase 2

P3H3, prolyl 3-hydroxylase 3

P4H, prolyl 4-hydroxylase

PDI, protein disulfide isomerase

rER, rough endoplasmic reticulum

SC65/No55, synaptonemal complex protein, nucleolar protein No55

Sf9, *Spodoptera frugiperda*

TCEP, tris(2-carboxyethyl) phosphine HCl

TD, trimerization domain

TPR, tetratricopeptide

TSPN-1, N-terminal module of thrombospondin-1

3Hyp, 3-hydroxyproline

CHAPTER 1

Introduction

Introduction to Scientific Questions about Collagen and Collagen-Associated Proteins

Collagen is the most abundant protein in the human body, playing an important role in the structural stability of connective tissues. Collagen has been implicated in a variety of other cellular functions including cell adhesion, chemotaxis, tissue remodeling, morphogenesis and wound healing. The importance of these molecules is clearly demonstrated by a wide spectrum of diseases caused by defects in collagen. Proper biosynthesis and processing of collagen requires many additional proteins such as hydroxylases and chaperones. Mutations in these proteins also result in a variety of diseases, further underscoring the significance of collagen and collagen-associated proteins.

Much work has been performed to better understand the structure of collagen molecules; however, several questions still remain. Trimerization of individual collagen chains is an important initial step in collagen biosynthesis. Although studies have identified many regions of collagen molecules that are important for proper self-assembly, only a few detailed structural analyses of these domains have been reported. In this thesis, studies on the trimerization domain of type XV collagen were performed to advance understanding of the structure and biochemical properties of these pivotal domains.

Collagen biosynthesis is a complex process that involves many different rough endoplasmic reticulum resident proteins. Numerous post-translational modifications, such as 3-hydroxylation of a proline residue, are performed as collagen chains are translocated into the rough endoplasmic reticulum. C*is-trans* isomerization of prolyl

peptide bonds is also necessary before folding and assembly can begin. Collagen chains are only marginally stable and have a tendency to aggregate, a propensity that is not surprising given that collagen eventually forms insoluble complexes outside of the cell. Collagen chaperones are necessary to prevent premature aggregation. In this thesis, the prolyl 3-hydroxylase 1 (P3H1) cartilage associated protein (CRTAP) and cyclophilin B (CypB) complex was found to incorporate peptidyl-prolyl isomerase and chaperone activity in addition to its previously characterized enzymatic activity. Additionally, studies on the P3H1 protein and the P3H1●CRTAP●CypB complex were performed in an attempt to further understand the structure of these proteins.

A review of the relevant literature follows to provide a better understanding of collagen and collagen-associated proteins covered in this work.

Collagen

The collagen family of proteins exhibits a broad range of structural and biological functions. All collagens are composed of three separate polypeptide chains called α-chains which can form homo- or hetero-trimers. α-chains are identified by the presence of one or more triple helical collagenous (Col) domains that are created by a repeated Gly-X-Y sequence in addition to globular, non-collagenous (NC) domains. To date, 29 different types of collagen with 43 different α-chains have been identified (1-3)'[*]. The number, size and distribution of the collagenous and NC domains vary between collagen types. Collagen is classified into six general categories: fibrillar collagens, network

---

[*] Collagen nomenclature identifies different types of collagen with roman numerals according to the chronology of their discovery. Individual α chains are labeled as αX[Y], where X identifies the chain isoform and Y indicates the type of collagen.

collagens, membrane associated collagens with interruptions, fibril associated collagens with interruptied triple helices (FACIT), anchoring fibrils and collagens with multiple triple helical domains with interruptions (multiplexin).

Many notable scientists have modeled the structure of the collagen triple helix, including Francis Crick, Linus Pauling and G.N. Ramachandran (4-6). Early fiber diffraction studies yielded pertinent information about the quaternary structure of the triple helix (7). However, atomic resolution structures of full length collagen are difficult to achieve, as the large size and insolubility of the molecule make traditional methods of structure determination difficult. Researchers have inferred more detailed information on the atomic structure of collagen using synthetic peptides and short collagen sequences expressed recombinantly.

The collagen triple helix is made of three left handed poly-proline type II helices (in which all peptide bonds are in the *trans* conformation), which are wrapped together into a right handed super helix (8). Due to tight stereochemical constraints, a repeating pattern of Gly-X-Y must be maintained. The X and Y positions are frequently occupied by proline and hydroxyproline, respectively. These residues, which are conformationally restrictive, stabilize the triple-helix entropically. Additional stabilization is provided by recurring $N-H_{Gly}$ to $O=C_X$ interstrand hydrogen bonds. 4-hydroxylation of prolines in the Y-position dramatically enhances the thermal stability of triple helices (8). Although stabilization was initially attributed to the formation of water bridges, it has been subsequently found that favorable ring pucker produces a stereoelectric effect that stabilizes triple helical collagen (9).

The collagen triple helix differs from other macromolecular helices such as the DNA double helix in that collagen assembles α-chains in a staggered arrangement. Three different chains can arrange into six staggered alignments. The exact assignment of stagger in natural collagens is unknown; however, stagger may play an important role in protein:protein interactions between collagens and other proteins in the extracellular matrix.

Collagen biosynthesis and processing

The biosynthesis and processing of collagen involves many steps. As the collagen chain grows, it undergoes multiple post-translational modifications important for the stability and function of the mature collagen molecule. Hydroxylases, glycosyltransferases, proteinases and an oxidase are required for proper processing of fibrillar collagen (10). Although not covered in this body of work, detailed literature on the glycosyltransferases, proteinases and oxidase can be found elsewhere (11; 12).

The three hydroxylases are prolyl 3-hydroxylase (P3H), prolyl-4 hydroxylase (P4H), and lysyl hydroxylase (LH). P3H and P4H modify proline residues to 3-hydroxyproline (3Hyp) and 4-hydroxyproline (4Hyp), respectively. 3Hyp is found in almost all collagens in a -Gly-3Hyp-4Hyp-Gly- sequence (13-16). 3Hyp residues occur at a lower frequency than 4Hyp: only 1-2 residues occur per chain in type I and II collagens and 3-6 residues per chain in type V and XI collagens (17-19; 13). Type IV collagen contains the highest level of hydroxylation, in which up to 10% of all hydroxyprolines are modified to 3Hyp (13; 20-25). Structurally, a polypeptide containing 3Hyp in the X position of a Gly-X-4Hyp sequence adopts a typical 7/2

superhelical symmetry found in other collagen structures (26). It was recently shown that conversion of a Pro to a 3Hyp residue in the X position of a Gly-X-Y sequence slightly increases the stability of the triple helix (27).

A few roles for 3Hyp modifications have been proposed. 3Hyp has been hypothesized to participate in supramolecular assembly (13). In this model, 3Hyp forms hydrogen bonds with adjacent chains and may thus play a role in aligning collagen molecules for intermolecular crosslinks. However, experimentation is needed to verify this hypothesis. Additionally, endogenous 3Hyp found in urine has been identified as a potential marker for cancer screening: urine samples from cancer patients have been shown to have lower levels of reactivity to an antibody against a type IV peptide containing a 3Hyp residue (28). Further studies are necessary to develop these results.

4-hydroxylation of proline residues is a key element in the folding of the triple helix, as the modification greatly enhances collagen stability (8). Triple helical collagen is not a substrate for P4H: an unfolded conformation is required for catalysis (29). P4H is a 2-oxoglutarate- and Fe(II)-dependant dioxygenase. The biologically active form is made from two α and two β subunits. In humans, several different isoenzymes are found in the rough endoplasmic reticulum (rER) with distinct α subunits. All isoenzymes have the same β subunit, which is identical to protein disulfide isomerase (PDI) (30). A novel family of three cytoplasmic P4Hs are involved with the regulation of the hypoxia-inducible transcription factor HIF; however, these enzymes have no PDI subunit and have different enzymatic requirements (31). LH modifies lysine residues to hydroxylysine; following lysyl hydroxylation, some lysine residues are further modified by the addition of galactose and glucose by glycosyltransferases.

After modification, chains are selected and aligned by association of trimerization domains, which are generally attributed to the NC domains located towards the C-terminus of the α-chain for most collagens. Trimerization domains will be discussed in more detail below. At this point, correct intramolecular disulfide bonds form to stabilize the trimeric intermediate state, although disulfide bonds are not necessary for triple helix formation (32). The process of chain selection and trimerization may occur at the luminal surface of the rER membrane (33).

Triple helix formation initiates from the C-terminus and propagates towards the N-terminus in a zipper-like fashion (34). The formation of the triple helix itself is a slow process when compared to the folding of globular proteins, since many of the prolyl and hydroxyprolyl peptide bonds undergo *cis-trans* isomerization (35; 36). Prolines, unlike other peptide bonds in the unfolded state, can exist in both *cis* and *trans* isomers. However, only *trans* peptide bonds are incorporated into the triple helix. This process is catalyzed by *cis-trans* isomerases such as the cyclophilins and the FK506 binding proteins (FKBPs) (37; 38). Isomerization has been found to be the rate limiting step in triple-helix formation both *in vitro* and *in vivo* (35; 36; 39)

The folding of collagen requires special chaperones to prevent aggregation. Heat shock protein 47 (HSP47) and FKBP65 have been shown to be molecular chaperones. In this thesis, the protein complex comprised of P3H1●CRTAP●CypB has also been shown to exhibit chaperone activity (40-42). HSP47 and FKBP65 have been shown to stabilize the collagen triple helix. However, the P3H1●CRTAP●CypB complex does not appear to have any significant effect on collagen stability (42).

As processing continues, collagens exit the rER and continue down the secretory pathway once the triple helix has formed. Procollagen molecules that are successfully modified and folded are then transported to the Golgi apparatus. Procollagen moves across the Golgi stacks without leaving the lumen of the Golgi cisternea (43). The N- and C-propeptides are then enzymatically cleaved, generating mature collagen molecules that subsequently form aggregates (44). Mature collagen is then packaged into vesicles and secreted into extracellular space (45; 46).

P3H1, P3H2, P3H3 and the P3H1●CRTAP●CypB Complex

The protein responsible for 3-hydroxylation of proline residues was not conclusively identified until 2004 (47). Prior to its discovery, prolyl 3-hydroxylase activity was partially purified from rat kidney cortex (48). The extract was capable of forming 3-hydroxyproline (3Hyp) residues in a reaction distinct from P4H activity. 3-hydroxylation was shown to require Fe^{2+}, O_2, 2-oxoglutarate and ascorbate. Longer polypeptide chains were shown to be better substrates for enzymatic activity, while triple-helical collagen was not found to be an acceptable substrate. Further analysis of enzyme activity using chick procollagen confirmed separation between P3H and P4H activity and a clear preference for 4-hydroxylation prior to 3-hydroxylase activity was also identified (49). In 1979, the protein responsible for 3-hydroxylation was partially purified from a chick embryo extract using a denatured collagen column (50). The molecular mass was estimated to be roughly 160 kDa by gel filtration. Additionally, the partially purified protein appeared to be a glycoprotein, since enzyme bound to concanavalin A coupled agarose was eluted using methyl α-D-mannoside in a manner consistent with other

glycoproteins. Although the enzyme was purified 5000-fold over the ammonium sulfate

fractionated enzyme derived from rat kidney cortex, it was neither sufficiently pure nor

concentrated for absolute identification.

The protein responsible for P3H activity was finally identified in 2004 (47).

Chick embryo rER extracts were partially purified by affinity chromatography on a

gelatin Sepharose column, which has been used to identify proteins that associate with

unfolded collagen (40; 51). A previously unidentified 90 kDa protein was found to

possess prolyl 3-hydroxylase activity. The protein was cloned, sequenced and named

P3H1. Sequence analysis identified P3H1 as the chicken homologue of leprecan (leucine

proline-enriched proteoglycan). Leprecan was originally isolated in 1999 and was shown

to be a chondroitin sulfate proteoglycan associated with the basement membrane,

although no enzymatic activity was associated with the protein at that time (52). The

human homolog to leprecan was independently identified as growth suppressor on

chromosome 1 (Gros1) (53). It is interesting to note that leprecan was identified as a

putative 2-oxoglutarate- and Fe(II)-dependent dioxygenase in 2001, three years before its

P3H activity was directly shown (54).   Two other genes were identified as P3H

homologs; myxoid liposarcoma associated protein 4 (*MLAT4*) and gene rich cluster B

(*GRCB*) correspond to P3H2 and P3H3, respectively. Overall, in *gallus gallus,* P3H2 is

67% homologous to P3H1, while P3H3 is 65% homologous (Figure 1).

P3H isoforms share a similar primary structure (Figure 2) (47). All isoforms have

an rER retention signal (KDEL) at the C-terminal end.   The remainder of the molecule

can be roughly divided into two portions: a unique N-terminal domain and a C-terminal

dioxygenase domain. The dioxygenase domain shares the highest sequence homology

among the P3H isoforms (Figure 1).   Two histidines and an aspartic acid residue that are critical for catalysis are conserved not only among the P3H homologues, but also among P4H and LH enzymes. A conserved positively charged residue involved in binding 2-oxoglutarate is found at a position +10 with respect to the second iron binding histidine: in P3H1 and lysyl hydroxylase, this is an arginine residue, whereas in P4H it is a lysine (55).   The N-terminal domains share less homology; however, all family members contain four CXXXC repeats and multiple tetratricopeptide (TPR) domains which have been shown to be important for protein-protein interactions (56). The N-terminal portion of P3H isoforms are unique and share homology with only two other proteins, the cartilage-associated protein (CRTAP) and the synaptonemal complex protein (SC65, also known as No55). P3H1 is co-purified with CRTAP and CypB (47). Together these proteins form a 1:1:1 complex (57). Mutations in P3H1, CRTAP or CypB lead to severe or lethal forms of recessive osteogenesis imperfecta (OI), a connective-tissue disorder (57-62), demonstrating the importance of the complex in human disease.

It is of interest to note that earlier work on partially purified 3-hydroxylase activity identified a proteoglycan, and initial characterization of rat leprecan identified it as a chondroitin sulfate proteoglycan (48; 50; 52). Rat leprecan localized to rER and Golgi organelle as well as to the extracellular basement membrane. The rER KDEL retention signal was retained in both the intracellular protein as well as the secreted proteoglycan. Chick P3H1, however, was not shown to be a proteoglycan. Additionally, an antibody specific for chick P3H1 did not stain tissues where basement membrane predominates, unlike previous reports on rat leprecan. Differences in species,

developmental stages, preparation protocols and antibodies most likely contribute to these discrepancies.

Northern blot analysis and real-time quantitative PCR on adult mouse tissues displayed both overlapping and unique expression of P3H family members (63; 64). All P3H family members were expressed in skeletal muscle, kidney, heart and liver. P3H1 expression was highest in the lung, liver and cartilage. P3H2 was strongly expressed in the kidney and was also detected in the spleen, eye and lung. P3H3 expression was found in the brain, an organ where P3H1 and P3H2 transcripts were apparently absent.

Embryonic expression of the P3H1, P3H2 and P3H3 genes shared some overlap, but also displayed distinct expression differences (47; 64; 65). All three were moderately expressed at E17. P3H1 displayed highest expression at E7, whereas P3H2 and P3H3 were expressed at highest levels at E11. *In situ* hybridization detected P3H1 expression in prechondrogenic skeletal elements, whereas P3H2 expression was observed in cells forming intervertebral discs and in the lens of the eye. P3H3 displayed more generalized expression that overlaps with P3H1 and P3H2. Additionally, all P3H family members were co-expressed in a variety of embryonic mouse tissues, including the notochord, limb buds and developing kidneys.

Immunofluorescence using an antibody specific for P3H1 has localized the enzyme to tissues that express fibrillar collagens. These include tendon, cartilage, skin and large blood vessels. However, P3H1 antibodies did not show staining in skeletal and cardiac muscle, kidney cortex or liver parenchyma (47). P3H2 was immunolocalized to tissues rich in basement membranes, with highest expression in the kidney.

Ultrastructural analysis by immunoelectron microscopy further localized P3H2 to kidney tubular cells (65). Immunolocalization of P3H3 is unknown.

Mice null for P3H1 (*Lepre1-/-*) displayed abnormalities in bones, tendon and skin (66). Null mice were smaller, grew slower and had decreased body fat. They developed progressive kyphoscoliosis. Bone defects included rhizomelia and decreased bone density, which was a result of delayed ossification. P3H1 null tendons displayed ultrastructural defects in overall shape and diameter of collagen fibers, with an increase in small diameter fibrils. The fibers displayed an axial twist and increased branching as well as an increase in overall heterogeneity. Although similar amounts of collagen were present in the dermis, *Lepre1-/-* skin was appreciably thinner that that of wild type mice. Collagen produced by null mice was overmodified and had increased levels of both 4Hyp and glucosyl galactosyl hydroxylysine resides. Overmodification is likely a result of decreased secretion rates in null fibroblasts which expose collagen chains to post-translational enzymes for a longer period of time. Increased 4Hyp and glycosylation levels led to a slight increase in the melting temperature of type I collagen. Hydroxylation at the P986 site in type I collagen was eliminated. However, 3Hyp residues were not eliminated at other known sites of 3-hydroxylation, which indicates a role for the P3H2 and P3H3 isoenzymes (unpublished results).

Mutations in P3H1 lead to severe OI, a connective-tissue disorder (57-62). OI is characterized by bone fragility, low bone mass, short stature, bowing of the long bones and kyphoscoliosis (67; 68). Classic OI is caused by dominant negative structural mutations in type I collagen (68-70). However, a subset of patients with lethal and severe OI do not have genetic defects in type I collagen genes (61; 62; 71-74). Despite the lack

of genetic abnormalities in type I collagen, these patients exhibit overmodified collagen. Proteins associated with collagen were therefore hypothesized to be involved with non-classic OI phenotypes. Indeed, 17 mutant *LEPRE1* alleles were identified as the cause of lethal to severe recessive OI, now classified as OI VIII. These mutations led to an absence or loss of function of P3H1. Type VIII OI probands had rhizomelia, *in utero* fractures and poorly mineralized bones. Patients also exhibited severely decreased levels of prolyl 3-hydroxylation and slightly elevated levels of lysyl hydroxylation and glycosylation (61). Secretion of type I collagen was also increased relative to wild type probands.

Less is known about the biological role of the P3H2 and P3H3 isoforms. P3H2 has been shown to preferentially hydroxylate type IV collagen, in keeping with its strong localization to kidney which is rich in type IV collagen (65). P3H3 localization and hydroxylation activity have yet to be determined. Mouse models are being developed in our laboratory which will provide new systems for further analysis. In humans, P3H2 and P3H3 have been shown to be down-regulated in certain cancers (75). Ectopic expression of P3H2 and P3H3 in cell lines with methylation-dependent transcriptional silencing of the endogenous genes were shown to suppress cell proliferation, indicating a potential tumor suppressor function. Mutations in P3H2 or P3H3 have not been identified ir nor associated with OI.

Mutations in CRTAP and CypB also led to defects in mouse and human models similar to those found in mutant P3H1 models. CRTAP null mice had recessive osteochondrodysplasia characterized by osteopenia, rhizomelia, progressive kyphosis and an overall growth deficiency. *Crtap-/-* mice also had decreased levels of 3Hyp and

overmodification of collagen (57; 76). Additionally, null mice displayed abnormalities in connective tissues consistent with dysregulation of collagen: the skin of null mice was thin, had disorganized layers and collagen fibrils that were increased in diameter with respect to wild type mice. As a result, the skin was noticeably lax and weaker when measured with a load-to-failure assay. Loss of CRTAP correlated to an almost complete loss of 3Hyp in type I, II and V collagens. However, the level of 3-hydroxylation in type IV collagen was similar between wild type and *Crtap-/-* mice, indicating that P3H activity does not depend on CRTAP in some types of collagen. CypB null mice (*Ppib-/-*) developed kyphosis and severe osteoporosis (77). CypB null mice also had reduced body size, thin skin with lower mechanical strength, overmodified collagen and abnormal collagen fibrils. Unlike the *Lepre1-/-* model, fibrils from *Ppib-/-* mice were on average wider than similar samples from wild type control mice.

CRTAP mutations have also been shown to lead to severe or lethal OI, designated as OI type VII (57; 58; 72). Patients displayed severe osteoporosis with rhizomelia, abnormal bone modeling and neonatal fractures. Those who survive into childhood developed extreme symptoms of chondrodystrophy and extreme growth deficiency. In most cases, levels of 3Hyp were reduced in patients with type VII OI . However, in one well-studied *CRTAP* missense mutation, levels of 3Hyp were near normal despite a severe phenotype (58; 60).

Recently it was reported that *PPIB* mutations also cause severe OI (62). 3Hyp levels were reduced, although the levels were higher than those seen in probands carrying *CRTAP* and *LEPRE1* mutations. However, a form of OI resulting from a mutation in *PPIB* has been identified wherein probands have moderate OI without rhizomelia or

extreme growth delay. Moreover, these probands displayed normally folded collagen and have normal 3Hyp levels (78).

Thus, mutations in any member of the P3H1•CRTAP•CypB complex result in OI. It was originally hypothesized that the lack of 3Hyp at specific proline residue in type I collagen led to the OI phenotype. However, defects in CRTAP and CypB, neither of which can perform 3-hydroxylation, generally lead to a decrease in 3Hyp levels (57; 62; 73). This suggests that recessive OI is caused by a dysfunctional P3H1•CRTAP•CypB complex rather than a lack of 3Hyp residues. This hypothesis is strengthened by cases of *CRTAP* and *PPIB* mutants that develop a strong OI phenotype despite having moderately reduced or even normal 3Hyp levels. Additionally, work in our lab indicates that a mutation in the CypB of the American Quarter Horse leads to a skin disease that exhibits changes in collagen fibril structure similar to those seen in P3H1 null mice. However, the collagen of affected horses did not have reduced levels of 3-hydroxylation. Absolute levels of 3Hyp, therefore, do not necessarily correlate to disease states.

Analysis of the protein levels of individual members of the complex using P3H1, CRTAP and CypB null or knock-down cell lines have shown that P3H1 and CRTAP are mutually stabilizing and that CRTAP and CypB are required to maintain intracellular levels of P3H1 (77; 79). The overall stability of the complex may therefore play a more important role than its specific 3-hydroxylase activity. In fact, work performed in this thesis demonstrates that the P3H1•CRTAP•CypB complex functions as a molecular chaperone and a *cis-trans* isomerase. These additional functions may play a more important role than 3-hydrxylase activity in OI disease pathology. Mutations in P3H1, CRTAP or CypB have been shown to lead to the overmodification of collagen. A

defective complex may also contribute to the OI phenotype by downstream impairment of secreted proteins in addition to changes of P3H1•CRTAP•CypB function in the rER. Further experimentation will be necessary to clarify the role of the P3H1•CRTAP•CypB complex in disease states.

The growing body of literature on P3H proteins and the P3H1●CRTAP●CypB complex indicate that these proteins play an important role in collagen biosynthesis and connective-tissue diseases. However, the underlying structural components that facilitate these activities are unknown.

Trimerization Domains

Before triple helical collagen can form, a self-assembly step must first select and bind α-chains. Unlike base-pair formation in the DNA helix, Gly-X-Y sequences do not form specific residue-to-residue interactions that guide helical formation. Thus, the triple helical regions of collagen molecules do not intrinsically posses the ability to specifically organize α-chains. Selection and association of α-chains must be a specific and controlled process, since hybrid α-chains between different collagen types are generally not found in nature, despite the fact that most collagen-producing cells simultaneously synthesize different types of procollagen molecules. The mechanism of chain association must therefore be attributed to NC domains (80). NC domains are numbered from the C- to the N-terminus with the NC domain closest to the C-terminus labeled NC1. All collagen molecules contain two or more NC domains. Since triple helix formation usually proceeds from the C-terminus (36), the NC1 domain is most likely responsible for trimerization for most collagens.

Studies on the classic fibril-forming collagens I and III have indeed found that the NC1 domain, also known as the C-propeptide, is essential for trimerization (81; 82). The C-propeptide is roughly 250 amino acids in length and is highly conserved throughout the fibrillar collagens (83). The C-propeptide has eight conserved cysteine residues: four cysteines in the N-terminus of the C-propeptide form intermolecular crosslinks, while the remaining four residues form intramolecular disulfide bridges (84). Mutations or deletions in the C-propeptide of type I collagen result in impairment of proper α-chain assembly, demonstrating the importance of C-propeptides in chain association (85; 86). Although no atomic-level structures of type I or III collagen C-propeptides exist, a low resolution model of the human type III C-propeptide was generated by a variety of biophysical techniques (87; 88). The C-propeptide trimer had a cruciform structure with three major lobes and one minor lobe, and it was hypothesized that the major lobes represent a globular fold that contains the intramolecular crosslinked portion of the α-chain while the minor lobe was made of intra-molecular disulfide bridges that join the triple helical region to the C-propeptide. An energy-minimized model of the type I C-propeptide yields a similar lobed structure (80). The type I C-propeptide model has distinct subdomains, and the domain responsible for trimerization has been localized to the very C-terminal portion of the C-propeptide and is referred to as G2. Recombinantly expressed $\alpha1(I)G2$ and $\alpha2(I)G2$ domains successfully formed trimers as shown by laser light scattering and gel filtration chromatography (89). This result directly contradicts previous predictions that a putative coiled-coil motif found in the N-terminal portion of the C-propeptide was responsible for oligomerization (90).

Network-forming collagens (types IV, VIII and X) and are characterized by their ability to form macromolecular assemblies (10). Oligomerization of network collagens is organized by C-terminal NC domains. The crystal structures of type IV, VIII and X collagens have been solved and provide insight into the mechanisms of network assembly (91-94). Type IV collagen is the major component of basement membranes and is composed of six different α-chains that assemble into three heterotrimeric protomers. The assemblies further assemble into three distinct supramolecular networks (95). Two crystal structures of the type IV [$(\alpha 1)_2(\alpha 2)_1$] NC1 domain have been solved (91; 92). Both structures displayed a hexameric NC2 (the second NC domain from the C-terminal end) complex that was composed of two trimers. A large planar interface joined the two trimers. In the structure derived from human placenta, additional stabilization of the dimer of trimers was provided by a novel crosslink between a Met and a Lys residue of α1 and α2 chains of opposite trimers. This explains previous reports of non-reducible cross-links in type IV assemblies (96). Both trimers were composed primarily of β-sheets. Significant stabilization was conferred by a finger-like hairpin loop that inserts into a six stranded β-sheet of the immediately adjacent chain.

Collagens VIII and X are short chain collagens which have highly similar NC1 domains (97). These NC1 domains are roughly 130 residues in length, are extremely resistant to thermal and SDS-induced denaturation, form stable trimers and are necessary for the assembly of collagen triple helices as well as supramolecular assemblies (98; 99). Both NC1 domains have been crystallized, and their overall structures were shown to be quite similar (93; 94). Each subunit consisted of a ten-stranded β-sandwich with jellyroll topology. The trimer core was almost entirely hydrophobic near the N-terminus of the

trimer, and became progressively more hydrophilic towards the top of the trimer, forming

a solvent-filled central channel. The type X NC1 trimer was stabilized by a buried cluster

of four calcium ions. Three strips of exposed aromatic residues on the surface of the

NC1 trimer are most likely involved in supramolecular assembly into polygonal lattices.

Similar strips were present in the NC1 domain of collagen VIII, although the buried

calcium cluster was not found.

FACIT collagens IX and XIX, which are characterized by short collagenous

domains that are interrupted by noncollagenous domains, have recently been shown to

trimerize via their NC2 domain (100; 101). The C-propeptide of fibril-forming collagens

is around 250 residues in length, whereas the NC1 domain of FACIT collagens is much

shorter, between 19 and 37 residues in length. FACITs have high sequence homology at

their collagenous 1 (COL1)/NC1 junctions; several studies suggested that the COL1 and

NC1 domains were involved in chain selection and trimerization (102). A study in 2008

showed that the absence of the COL1 and NC1 domains in type IX collagen did not

inhibit triple helix formation, although the COL2-NC2 region alone was not sufficient for

trimerization (103). McAlinden found that the NC2 domain of FACIT collagens contain

a repeating heptad sequence typical of coiled coils (90). When fused to a reporter

molecule containing a collagen sequence that does not spontaneously trimerize, coiled

coil heptad repeats were able to induce trimerization. Boudko and coworkers later found

that the NC2 domain of the homotrimeric collagen XIX and the heterotrimeric collagen

IX was responsible for chain selection and trimerization. In keeping with the hypothesis

set forth by McAlinden, the CD spectra of the NC2 domains of type IX and XIX were

primarily α-helical. The type IX trimerization domain is of particular interest; as a

heterotrimer, it may provide insight into the mechanism of chain selection that ultimately

determines the stagger of a collagen triple helix. However, atomic resolution structures

of any FACIT NC2 domain have yet to be determined.

For all remaining collagens, the NC1 domain is believed to be responsible for

chain selection and trimerization (80). Transmembrane collagens are most likely the only

exception: N-terminal NC domains were shown to be necessary for triple helix formation,

consistent with studies that have shown that propagation of transmembrane collagen

proceeds from the N- to the C-terminus (104). Further experimentation is necessary to

firmly identify and characterize trimerization domains of numerous other collagens.

Multiplexins

Type XV and XVIII collagens contain multiple triple-helix domains with

interruptions and are the only two known members of the multiplexin family (105). They

have long N- and C-terminal NC domains with a central, highly interrupted domain that

contains collagenous domains alternating with NC domains of varying length (Figure 3a).

Comparison of the cDNA-deduced primary structures of human type XV and XVIII

collagen reveal homologies in both their non-collagenous and collagenous sequences , a

similarity which extends to their genomic organization. Furthermore, many of the

collagenous domains in these two collagens are homologous in size and sequence (106-

114). The N-terminal NC domain of type XV and XVIII collagen contains a ~200

amino acid module which is homologous to the N-terminal module of thrombospondin-1

(TSPN-1), a domain that is also found in several other collagens (115; 116). The amino

acid residues involved in heparin binding of the TSPN-1 domain are not conserved in any collagen type, the significance of which has yet to be determined (107).

Multiplexins share highest homology in their NC1 domains, which are organized into three different subdomains: a trimerization domain, a protease-sensitive hinge and an endostatin or endostatin-like domain (Figure 3b) (117).   The trimerization domain is roughly 60 residues long in both type XV and XVIII collagens. The trimerization domain displays relatively low sequence identity (~32%); however, as demonstrated in this body of work, their three dimensional structures are quite similar. The hinge region is most variable between the multiplexins (118). The type XVIII hinge is roughly 70 residues long and contains multiple protease-sensitive sites that are involved in the generation of endostatin-containing fragments. The type XV hinge region is significantly shorter (~20 residues) and contains only two known protease sites. The endostatin and endostatin-like domains of type XVIII and XV collagens share the highest level of identity at 61%. Endostatin and endostatin-like domains are globular proteins that have been implicated in angiogenesis inhibition. Much work has been performed on these molecules that exceeds the scope of this chapter – a more detailed analysis of endostatin and endostatin-like protein function can be found elsewhere (119-123). Structurally, mouse and human endostatin and the mouse endostatin-like protein have been shown to share highly similar overall folds, consistent with the high level of sequence similarity (124; 125; 119).

Both type XV and XVIII collagens have been located in basement membrane zones, although there were differences in their expression patterns and tissue distribution. Homologues of multiplexin collagen have been identified in *C. elegans* and *D.*

melanogaster, indicating a long evolutionary lineage. A more detailed discussion of type XV and XVIII collagens is covered below.

Type XV Collagen

Type XV collagen was initially identified from human placenta cDNA by Myers et al. in 1992 (126) and the full human $\alpha1(XV)$ gene was reported by Kivirikko et al and Muragaki et al. in 1994 (106; 107). The human type XV $\alpha1$ chain was shown to be 1388 residues in length, consisting of a 25-residue signal peptide, a 530 residue N-terminal NC domain, a 577 residue interrupted region and a 256-residue C-terminal NC domain. The interrupted region contains nine collagenous domains interrupted by eight NC domains. Additionally, five short interruptions of 2 or 3 residues were found in four of the collagenous domains (107). The molecule has eight conserved cysteines: α-chains are covalently linked by interchain disulfide bonds between two cysteines in the collagenous region (127). Additionally, type XV collagen has several putative glycosylation sites. Indeed, type XV collagen was shown to be a chondroitin sulfate proteoglycan in several tissues (127), and was subsequently shown to carry chondroitin sulfate or chondroitin sulfate and heparan sulfate GAG chains (128). The chondroitin to heparan sulfate ratio varied depending on tissue type, with the highest amount of heparan sulfate found in kidney. The GAG chains were attached to the N-terminal NC domain (127).

Type XV collagen mRNAs were found in every human tissue tested, but had strongest expression in heart, skeletal muscle, smooth muscle and placenta. Moderate expression was found in the kidney, pancreas and placenta. Type XV transcripts were absent in the lung, brain and liver. Fetal expression patterns were slightly different, with

stronger activity shown in the developing kidney than that of adult kidney. Additionally, weak expression was detected in the brain and lung which was not seen in adult samples (129; 130).

Type XV protein immunolocalized to capillaries and the BM zones of skeletal and cardiac muscles, the dermal-epidermal junction, the small intestine and placenta (131; 132). Staining outside of BM zones was observed in the fibrillar collagen matrix of papillary dermis and placental villi (131). Type XV protein levels changed during kidney development (131). Primitive glomeruli exhibited no immunostaining, but an increase in signal was detected with glomerular development in fetal kidney. The immunostaining levels in mature kidney were weaker than during fetal development. Type XV collagen was virtually absent from the BM zones of malignant glandular elements in moderately differentiated human colonic adenocarcinomas, and accumulated abnormally in interstitial space (133). Type XV collagen also accumulated in the interstitium during invasive fibrotic processes (131; 134). Immunogold ultrastructural analysis by electron microscopy found that type XV collagen was almost exclusively associated with the fibrillar collagen network in close proximity to but not directly associated with basement membrane. Type XV collagen was found on the surface of thick-banded collagen fibrils and also appeared to act as a bridge between fibrils (128; 135).

Consistent with localization by electron microscopy, mice null for *Col15a1* did not display significant disruption of their basement membrane (BM) (136). Null mice did not have any major developmental or reproductive problems and were indistinguishable from their wild type littermates. After a few months, however, progressive histological changes developed. Markers of muscular disease such as the presence of degenerative

muscle fibers, variability in fiber size and centrally located nuclei were found in adult tissue samples. Null mice were also more vulnerable to exercise-induced muscle injury. After imposed mechanical stress, type XV null mice developed cardiovascular defects and skeletal myopathy. Histological analysis showed collapsed capillaries and endothelial cells. Additionally, although vascular development was normal, ultrastructural analysis of null mice revealed abnormal microvessels in the heart and skeletal muscle (136). Since type XV collagen has been shown to link BM to the surrounding fibrillar matrix, these studies suggest that type XV collagen plays a structural role in skeletal muscle cells and microvessels by stabilizing these cells with surrounding connective tissues.

Collagen XV has been purified from human umbilical cords and visualized by electron microscopy (135). Most molecules were found in complex configurations not previously described for a collagen molecule. Knots, figure-eight and pretzel-like shapes were most common, while a small minority of the molecules were present in a more elongated conformation. Type XV monomers were shown to self-assemble into higher-order structures, most likely through intermolecular interactions found in the interrupted domain.

Mutations in type XV collagen have not been associated with any human disease thus far, although changes in type XV expression were associated with kidney fibrosis and human colonic adenocarcinomas as outlined above.

Unlike endostatin, the biological role of type XV endostatin-like protein has not been extensively studied. Endostatin-like protein exhibits some antiangiogenic and antimigratory effect on endothelial cells, but has no effect on non-endothelial cells (137;

138). Interestingly, unlike the type XVIII NC1 domain, the type XV NC1 domain does

not appear to be substantially converted into smaller fragments (118).

Type XVIII Collagen

The α1 chain of human type XVIII collagen was identified in 1994 (139). Human

type XVIII collagen, in contrast to type XV collagen, exists as two variants expressed

from two different promoters (109; 140) (Figure 3). The two chains have different signal

peptides of 23 and 33 residues and 493 and 303 residue N-terminal NC domains,

respectively (109). The collagenous region is 688 residues in length which contains ten

collagenous domains with nine interruptions, while the C-terminal NC domain is 312

residues long. The collagenous domain also contains five smaller interruptions of 2-5

amino acids (108; 139) Both the interrupted collagenous domain and the C-terminal NC

domain are identical between the two variants. Additionally, the transcript can undergo

alternative splicing affecting a 43-residue stretch at the junction of the NC11[*] domain and

the beginning of the collagenous sequence. α1(XVIII) chains have eight conserved

cysteines residues, although two cysteines residues are located in the NC10 domain

which is the least conserved region and is lost during alternative splicing.

Chick type XVIII collagen was shown to constitutively carry three heparan sulfate

chains out of 8 potential GAG sites (141). Human type XVIII collagen has four potential

N-linked glycosylation sites; however, the exact extent of glycosylation is unknown

(142).

---

[*] The paper in which this work was published uses a different nomenclature to number NC domains. Their nomenclature has been adjusted to the modern system which labels from the C- to the N-terminus.

Type XVIII mRNA variants were differentially expressed in several tissues. The long variant was highly expressed in both fetal and adult liver by hepatocytes and occurs only in minor amounts elsewhere. The short variant, although not significantly expressed in the liver, was found in most basement membranes including blood vessels, various epithelial structures and muscular structures. The short variant was highly expressed in fetal kidney and brain (109). Expression patterns of XVIII isoforms changed during early lung and kidney development, implicating type XVIII collagen involvement in organ morphogenesis and patterning (143).

Type XVIII protein localized to the BM zones of several embryonic and adult tissues, particularly those associated with blood vessels and the eye (144). Both variants localized to BM zones of vascular and epithelial tissues, while the longer isoforms is almost exclusively found in liver (109; 145).   However, type XVIII collagen was not found to be a universal BM collagen; most tubular basement membranes in the kidney, for example, were negative. Specialized capillaries stained strongly for type XVIII collagen, specifically glomerular capillaries in kidney, sinusoid capillaries in liver, alveolar wall capillaries in lung and the sinus capillaries in spleen (132). Immunogold labeling with a human XVIII antibody clearly localized collagen XVIII to the BM of the eyes (146). Additionally, clusters of collagen XVIII co-localize with fibrillar structures in the adjacent matrix of most BMs.

Type XVIII collagen was strongly expressed in liver, and has been implicated in BM changes associated with liver cirrhosis (147). Tumor hepatocytes were shown to be a major source of the both the long and short variants of type XVIII collagen in human hepatocellular carcinomas (147). Increased expression levels of the long variant were

shown to correlate with smaller tumor size; likewise, decreased levels of expression were associated with larger tumor size. The short variant was also identified in tumor tissues of metastatic liver cancers.

Col18a1 null mice were viable and reproduce normally (146). As with *Col15a1* mice, there were no significant gross disruptions in the BM. However, abnormalities in the retinal vasculature were apparent: mice displayed delayed regression of blood vessels along the surface of the retina and abnormal outgrowth of retinal vessels. Additionally, the vitreal matrix separated from the inner limiting membrane resulting in defective irises (148; 149). Immunogold labeling previously localized type XVIII collagen to collagen fibrils connected with the inner limiting membrane: collagen XVIII probably plays an important role in anchoring vitreal collagen fibrils, which may explain why null mice experience vitreoretinal degeneration and separation (146).

Human type XVIII collagen defects result in Knobloch syndrome, an autosomal recessive disorder characterized by various eye defects such as high myopia, macular abnormalities, vitreoretinal degeneration and retinal detachment (143; 150). Patients with Knobloch syndrome experience increasingly severe ocular abnormalities that progress to bilateral blindness. Premature stop codons, insertions and deletions were identified in various exons and introns of the *COL18A1* gene. A loss of immunodetectable collagen XVIII was detected in skin biopsies from Knobloch syndrome patients (151).

Endostatin, a fragment derived from the NC1 domain of type XVIII collagen, has been the target of much research since it was originally identified as a specific inhibitor of endothelial cell proliferation *in vitro* and a potent inhibitor of angiogenesis *in vivo* (152). The protein was shown to be present in a variety of tissues and also as a

circulating fragment in human plasma hemofiltrate (117; 137; 153). In a variety of experiments, endostatin exhibited a broad spectrum of angiogenesis inhibition, had low toxicity and did not induce acquired drug resistance (120). On a cellular level, endostatin has been shown to specifically induce endothelial apoptosis as well as to block growth factor-induced proliferation and migration of endothelial cells (154; 155). Numerous studies have shown that a variety of animal and human tumors were inhibited by administration of recombinant endostatin protein; however, there are conflicting reports about the efficacy of endostatin anti-tumor activity (156; 157). Investigations into the biological effect of endostatin are complicated by a variety of issues including differences in protein expression, solubility and mechanism of delivery – for a comprehensive review, see Folkman (2006) (121). Endostatin has also been implicated in other cellular processes such as wound healing (158) and axon guidance (159).

Endostatin was shown to be produced by proteolytic activity in the hinge region of the NC1 domain of type XVIII by cathepsin-L, elastase and matrilysin (160-162). Pre-processing within the NC1 domain by matrix metalloproteinases (MMPs) may be necessary *in vivo* before endostatin can be processed (122). Trimeric type XVIII NC1 domain and monomeric endostatin were shown to act antagonistically in certain cellular processes: the trimeric NC1 domain of type XVIII collagen induced motility of both endothelial and non-endothelial cells, while the monomeric endostatin inhibited motility. Artificial dimerization of endostatin changed its activity to that of the full length NC1 domain (163). Thus, proteolytic processing of type XVIII collagen to NC1 and endostatin fragments may play a regulatory role in endostatin function.

Type XV/XVIII Collagen in *C. elegans and D. melanogaster*

Although there are over twenty types of collagens in mammals, only type IV and type XV/XVIII collagens are conserved within the *C. elegans* and *D. melanogaster* genomes (159; 164). Like mammalian XVIII, the nematode and fruit fly multiplexin homologues (CLE-1 and Dmp, respectively) exist in multiple forms that have distinct tissue specific expression patterns. Additionally, CLE-1 was able to bind heparin in a manner consistent with mammalian endostatin. CLE-1 is broadly distributed in basement membranes with highest expression in the nervous system. Deletion of the CLE-1 NC1 domain resulted in cell migration and axon guidance defects, which were partially rescued by ectopic expression of the trimeric NC1 domain. Addition of the monomeric endostatin did not rescue the defects. Dmp is involved with motor axon pathfinding. Transgenic over expression of monomeric Dmp endostatin as well as full length Dmp were able to rescue axon defects, although trimeric endostatin did not. These data indicate that processing of multiplexin collagens play an important role in a variety of cellular functions in a manner that is well conserved amongst several organisms.

Thesis Overview

In this thesis, I describe the results of my investigation into trimerization domains and collagen-associated proteins. Chapter 2 covers the principles of macromolecular X-ray crystallography techniques used to solve the structure of the type XV trimerization domain. Chapter 3 presents the structure in addition to biochemical characterization which shows the type XV trimerization domain to be an extremely stable domain capable of forming at very low concentrations. In Chapter 4, the biochemical characterization of

the P3H1●CRTAP●CypB complex is presented: my specific contribution focused on analysis of complex binding to folded type I collagen using fibril formation and surface plasmon resonance assays. Chapter 5 presents a detailed analysis of P3H1 protein over expression systems and purification from chick embryos. These studies provide insight into the structural basis of collagen trimerization, advance the biochemical characterization of the P3H1●CRTAP●CypB complex and provide a foundation of work to improve expression and purification of P3H1 and the P3H1●CRTAP●CypB complexfor future studies.

```
                  1
cP3H1     (1)   -MALLLPLLPLLVWAAGPPGPLGPPEHSEPPLPAEPPDALFAAGAEAYARGDWPAVVLQMERALRARAAI
cP3H2     (1)   MAPGSRSWGAVLLLAAMLPAACG--SCGADGGPLEPFDALYASGVEAYYGGDFAGAARCLEQALRSRREL
cP3H3     (1)   MASLVCLLLLLLLPAATSSVAG---------RLVPYDLLYADGVRAYFARDWGRAAELLQRALYSYAGL

                  71        #       #                                    #       #
cP3H1     (70)  RARSVFCRLRCANATAVVPTDGLEPTLHDLLFFRGLLRRAACLRGCGPTQP----SRYRLGEELEREFRK
cP3H2     (69)  RAERLFCRRRCRGQVRLAALG--AGPAGELPFFGALLRRAGCLRSCEEPRLGAA-SRHRAAEEVRSDFQR
cP3H3     (61)  RAARLACRDACLREAAFLGEPP-AAGPWEAALFDRVLQRADCLQHCLGSRLGAAPSAHRASLVIQRDFER

                  141
cP3H1     (136) RSPYNYLQVAYFKINKVAKAVAAAHTFFVANPEHVEMKQNLEYYQMMAGVRESDFADLEARPHMTEFRLG
cP3H2     (136) RVPYSYLQRAYIQLNKLEEAANAAHTFFMANPEHMEIQQDIENYKTTAGKVS--LIDLEAKPHMEDYSAG
cP3H3     (130) REPYNYLQVAFFQLKKLDQAVSAAHTFFVANPQHLQMREDIEKYRRMSGVKSDNFRDLEAAPHWEAYEAG

                  211                               #       #                         #       #
cP3H1     (206) VRFYSEEQPAAAVLHLEKALQEYFVADTECRALCEGPYDYEGYNYLEYNADIFQAMTDHYMQVLSCKQGC
cP3H2     (204) VRHYDKEEYGLAITFLERALEGYYAEDEDCQIMCEGPQRFEEHEYLEYKAGLYEAIADHYMQVLACKHDC
cP3H3     (200) MQHYNADEYVQAVDRLEESLREALSALEECRALCEGP--LEDEEEEDLQPGIYEAIAAHYIQVLKCRQQC

                  281
cP3H1     (276) VTEIASQPGREKPLEDFLPSHFNYLQFAYYNNGNYEKAIECAKTYLLFFPNDEVMNQNLAYYTAVLGEDL
cP3H2     (274) IREIATRSGRISPIENFLPLHYDYLQFAYYRVGDYVKALECARSYLLFHPDDEDVLENAAYYEGLLEGTV
cP3H3     (268) VLEIATKPGWISATEDFIPSHLDILQFAYDQVGNPALAAECAASYLLFYPTDEPVLEKMKQYRAELGQDT

                  351
cP3H1     (346) ARPIE-PRKEIQAYRQRSLMEKELLFFSYDVFGIPFVDPDTWTPEEVIPKRLREKQKVERETAARISEEI
cP3H2     (344) DPATIKPRKEAKALLRRHKLESHLLRVAAVGLGFTYTEPNYWKRYGARQDEHSVPSSISSE----P----
cP3H3     (338) AVTAR---ENIRHYVQRSLMEKKLIYYATEHLGGTFNDPQPWSVFFPLLPREDQK----K----------

                  421
cP3H1     (415) GNLMKEIETLVEEKAKESAEMSKFIREGGPIVYEGASVTMNSKSLNGSQRVVVDGVLSAEECRELQRLTN
cP3H2     (406) -----EDGPRLSLTKKPTPKPDRELKEGGPILYSDVKFVYNSQQLNGTQRVLLDNVISEEQCRELHRVAS
cP3H3     (391) ------------QKQETLDMEERGKRGS-LPFEGITVTMDSQQMNGTQRVVFDRVLTESECKDLLRLTE

                  491
cP3H1     (485) AAASAGDGYRGKTSPHTPSELFYCGVTVLKALKLGQEGKVPLQSAHLYYNVTEKVRHMMESYFRLEVPLHF
cP3H2     (471) GIMLAGDGYRGKTSPHTPNERFESCATVLKALKYGYEGRVPLKSARLFYDISEKARRIVESYFMLNSTLYF
cP3H3     (447) AARGTGDGYRARRSPHTPHERFEGLTVLKAMQLAQNGDVDWRDAKILLQASEKSRKIIESYFTPGKKLHF

                  561                          *  *
cP3H1     (555) SYSHLVCRTAIDEKQEGRSDNSHEVHVDNCILNAEEALVCVKEPPAYTRDYSAILYLNGDFEGGAFYFTE
cP3H2     (541) SYTHLVCRTALSGQQERRNDLSHPIHADNCILDPEANECWKEPPAYTRDYSAILYMNADFEGGEFIFTE
cP3H3     (517) SFTHLVCRTAVEEEQEGRMDLSHPVHADNCILDPEGKECWKEPPAYVYRDYSGILYLNDDFRGGSLFFTD

                  631                          *                    •
cP3H1     (625) LDAKTQTAEVQPQCGRAVGFSSGSENPHGVKAVTKGQRCALALWFTLDPRHSERERVQADDLVKMLFRTE
cP3H2     (611) MDAKTVTASIKPKCGRMVSFSSGGENPHGVKAVTKGQRCAVALWFTLDPLYREIERIQADEVIAMLD---
cP3H3     (587) IDTVTVTANVHPKCGRLVAFSSGKENPHGVRAVSQGRRCAIALWYTHSVEHAEQERVKAEELMEQRAVEE

                  701                                                               755
cP3H1     (695) EVDLLQETSAEQEPTAAT----------------------STAGLHAAGKDEL-
cP3H2     (678) -----QEHVGRSE----------------------------MNINPKDEL-
cP3H3     (657) NQPDGEEYHGADRSSRSSPEPLVSGGGTRPTSQKHKSVSERTQHPKKLRARDEF-
```

Figure 1. Alignment P3H1, P3H2 and P3H3. Identical residues are shown in white with black shading, conservative residues in black with gray shading, blocks of similar residues are shown in red, and weakly similar residues in blue. The repeating CXXXC motif is marked with a "#". Residues conserved in the active site domains of P3H, P4H and LS proteins are marked with a "*", while similar residues are marked with a "•".
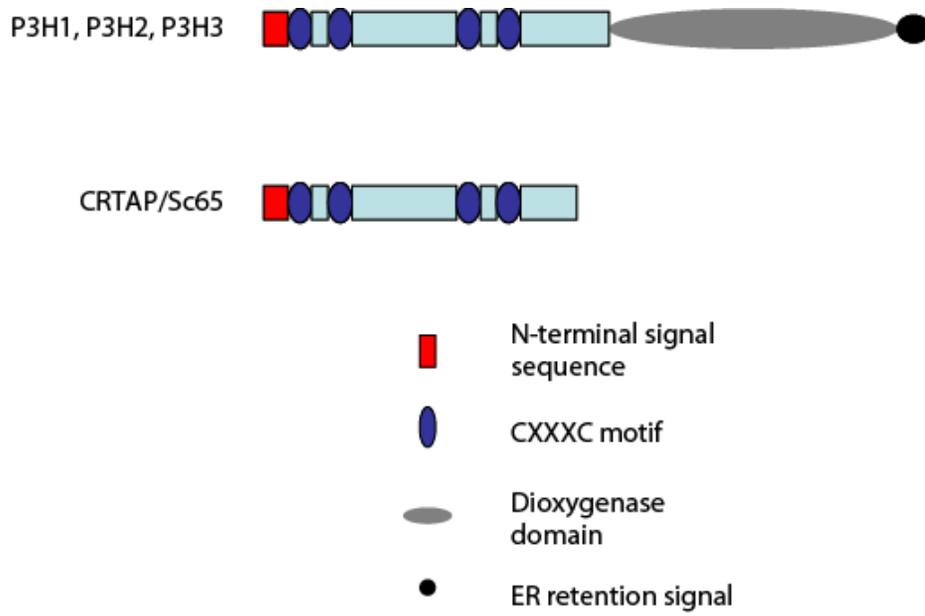
Figure 2. Schematic of the P3H and CRTAP/Sc65 proteins.
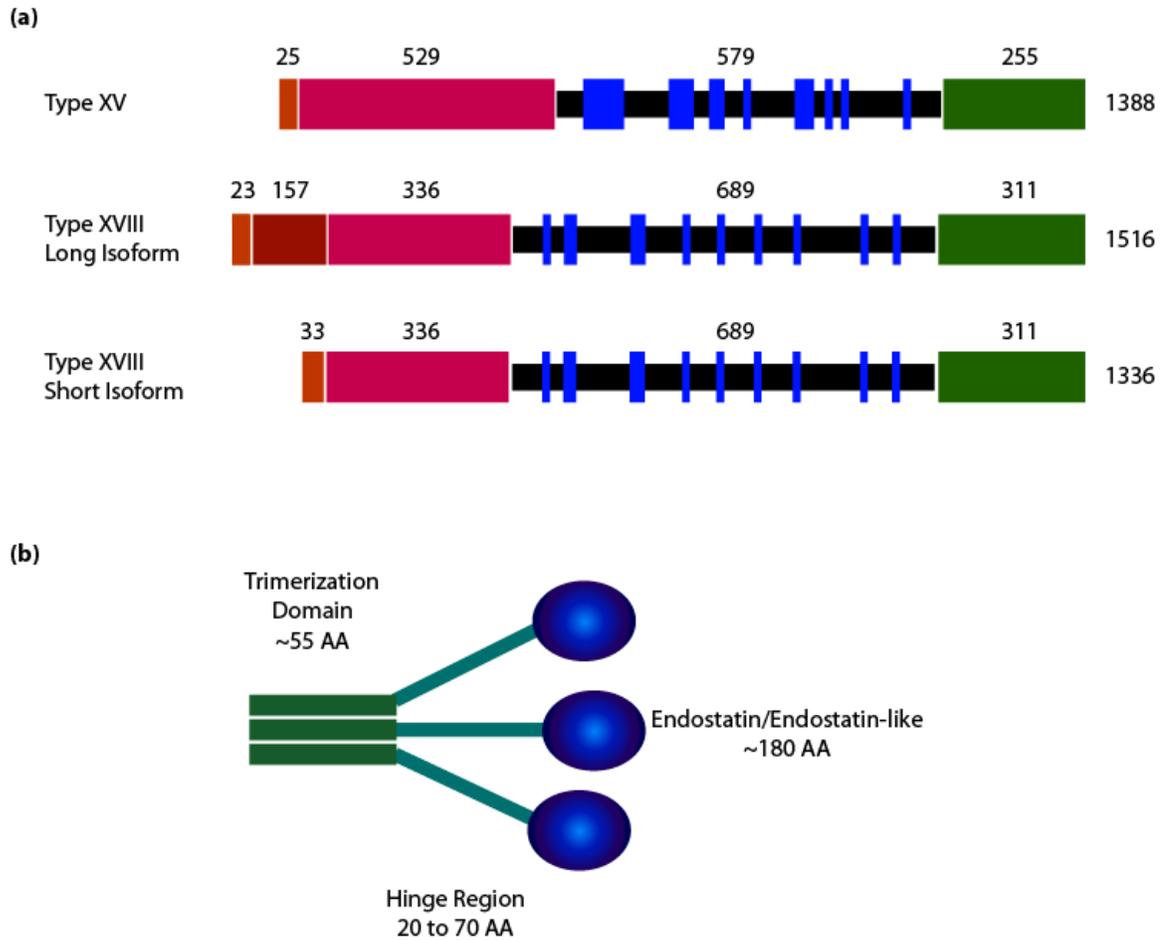
Figure 3. Multiplexin Schematic. (a) Orange represents the signal peptide, red/pink the N-NC domain, black the interrupted collagen domain, blue the interrupting NC regions and green the NC1 domain. Numbers represent the length of the segments in amino acids. (b) The NC-1 domain consists of a trimerization domain (green), a hinge region (teal) and a globular endostatin or endostatin like domain (blue).

CHAPTER 2

Structure Determination by X-Ray Crystallography

Structure Determination by X-ray Crystallography

Macromolecular X-ray crystallography is a method by which atomic resolution structures are determined. X-rays are applied to an ordered macromolecular crystal and the angles and intensities of diffracted rays are used to produce a three-dimensional picture of electron density within the crystal. A molecular model is subsequently developed to best fit the electron density. X-ray crystallography is one of a few techniques that structural biologists can use to probe the structure of proteins at high resolution. This chapter covers the basic concepts of macromolecular crystallography and outlines the key steps in determining a protein crystal structure. Several excellent references are available for a more detailed description of the theory behind and practice of crystallographic techniques (165-168).

When the human eye visualizes an object, light rays are scattered by the object and enter the eye where the lens reconstructs and focuses the image on the retina. In order to resolve an object, the wavelength must not be much larger than the object. Visible light is electromagnetic radiation with wavelengths between 400 and 700 nm; it cannot produce an image of individual atoms in macromolecules where bonded atoms are only 0.15 nm (1.5Å) apart. X-rays, which have wavelengths between 0.1 and 100 Å, are ideally sized to visualize protein structures. X-rays are diffracted primarily by electrons; therefore the diffraction patterns obtained by exposure to X-rays are best described as a representation of electron density.

There are, however, two problems with X-ray determination of macromolecular structure. First, a single molecule is a very weak diffractor of X-rays. Crystallographers enhance the overall diffraction signal by using macromolecular crystals. A crystal

contains numerous ordered molecules in identical orientations, such that each molecule scatters identically. Scattered rays from all molecules in the crystal can interfere constructively, producing detectable X-ray diffraction. The second major problem with the use of X-rays is that they cannot be focused by lenses. Crystallographers collect the angles and intensities of diffracted rays and use this information to mathematically reconstruct an image of the molecule. In effect, the crystallographer's computer acts as lens to compute and display the macromolecular image.

Geometric Principles of Diffraction

Crystals are ordered, three-dimensional arrays of macromolecules that repeat periodically in all directions. Crystals can be divided into identical unit cells which are characterized by three axial lengths ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) and three interaxial angles (α, β, γ) (Figure 1a). These six cell parameters are used to define the seven basic crystal systems: triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal and cubic. The lowest level of symmetry, where all six parameters are different, is classified as triclinic. As symmetry increases, the number of independent parameters decreases. Cubic crystals exhibit the highest level of symmetry, where $\mathbf{a} = \mathbf{b} = \mathbf{c}$ and $\alpha = \beta = \gamma$.

Each unit cell can also be described by its lattice type. Every lattice contains a lattice point at each corner of the unit cell and can be classified as primitive lattice, designated by the letter P. There are several other non-primitive crystal systems that contain two or more lattice points per unit cell. For example, a body-centered internal lattice has an additional lattice point in the center of the cell. Face-centered lattices have

additional lattice points centered on some or all of the unit cell faces. The combination of crystal and lattice systems creates 14 lattices known as the Bravias lattices.

There are a variety of symmetry operators that can be used to describe the relationships between equivalent positions in a crystal. Combinations of these operations result in 32 point groups. The 14 Bravias lattices can be combined with the 32 possible point groups to create 230 unique arrangements of points called space groups. Space groups are described by a symbol system where the capital letter indicates the lattice type and the remaining symbols represent symmetry operations. Due to the chiral nature of amino acids, mirror symmetry is not allowed since biologically irrelevant molecules would be created. Therefore, only translations, rotations and screw axes (translations with rotation) are allowed which reduce the number of biologically relevant space groups to 65. Unit cell symmetry can be described by sets of equivalent positions, which are locations within the unit cell that can be superimposed on each other by symmetry operations. Symmetry can be exploited as a means to reduce the amount of data collected if short collection times are necessary or, alternatively, can provide higher redundancy datasets.

There are usually several ways to choose the unit cell in a lattice. It is always possible to define a triclinic, primitive cell regardless of symmetry. However, unit cells are chosen to take advantage of symmetry elements which can greatly simplify subsequent data collection and analysis. In most cases, the unit cell will contain several identical molecules or complexes in an arrangement that is produced and can be described by symmetry elements. The smallest unique region that can be related to equal assemblies in the unit cell is called the asymmetric unit. Each asymmetric unit must have

an identical environment. For example, in the case of the hexagonal crystal type XV

trimer, a single subunit was found in the asymmetric unit and the remaining two subunits

were generated by symmetry operations. Crystallographic generation of the trimer was

possible because each subunit was in an identical conformation. The trimerization

domain of type XV collagen was also crystallized as a cubic crystal. In this case, six

individual chains were found in the unit cell. Since each chain was in a slightly different

environment, a larger asymmetric unit was formed and limited the ability to use

symmetry elements to describe the relationships between subunits within the crystal.

The unit cell can be divided into a series of regularly repeating sets of planes, the

most obvious of which are the planes that define the unit cell faces (Figure 1b). Miller

indices hkl are related to the **a**, **b** and **c** axes of the unit cell and define all equivalent,

parallel planes. The h index gives the number of parallel planes that intersect the unit cell

along the **a** axis, k for the **b** axis and l for the **c** axis. Indices are written in parentheses

when referring to a set of planes; therefore, planes with indices hkl are the (hkl) planes.

The plane along the **bc** face of a unit cell (normal to the x-axis) would have a (100) index

because there is one intersection along the **a** axis in the x direction. Likewise, planes for

the **ab** and **ac** faces would be indexed (001) and (010), respectively. The (100) set of

parallel planes would therefore describe the set of planes along the **bc** face of each unit

cell, and is not specific to a single unit cell within the crystal. This distinction is

important since each set of parallel planes can be treated as a single diffractor that

produces a unique reflection rather than viewing each plane as an individual diffraction

element. As the Miller indices increase in number, there are a growing number of

intersections along each axis resulting in a decrease in interplanar spacing, d. That is to

say, Miller indices of high numeric value describe closely spaced planes within the unit cell.

The father-son team of W.H. and W.L. Bragg were the first to describe X-ray diffraction for crystals as reflections from lattice planes (169). They were able to create a simple equation that detailed the connection between X-ray wavelength, interplanar spacing and the angle at which the X-rays are constructively reflected. The Braggs were awarded the Nobel Prize in Physics in 1915 for their tremendous contribution to the new field of X-ray crystallography. In mathematical terms, Bragg's law simply states that a set of parallel planes with index *hkl* and interplanar spacing d_{hkl} produces a constructively interfering diffraction beam when X-rays of wavelength λ hit the planes at an angle θ only if θ meets the condition

$$2d_{hkl} \sin \theta = n\lambda \,,$$

where *n* is an integer (Figure 2). $2d_{hkl} \sin \theta$ is the extra distance a second incident ray would travel between two parallel planes. In order for constructive interference of the diffracted X-rays to occur, $2d_{hkl} \sin\theta$ would have to equal a whole number of wavelengths to keep all reflected waves in phase with each other. If Bragg's law is satisfied, rays emerging from successive planes are in phase and interfere constructively to produce a strongly diffracted beam. X-rays that are incident with the same planes at an angle other than θ would not be in phase and would therefore produce no reflection. The intensity of the resulting reflection is proportional to the number of electrons and their distribution along the set of parallel, equivalent planes.

The angle of diffraction θ is inversely related to the interplanar spacing d_{hkl}. Since sinθ is measure of the deviation of diffracted beams from the direct beam, crystals with

large interplanar spacing will have smaller angles of diffraction. Smaller unit cells, in contrast, give larger angles of diffraction which produce fewer measurable reflections, for a given detector area. The spacing of reflections is inversely related to the dimensions of the unit cell.

A theoretical construct called the reciprocal lattice can be used to better understand the relationship between the real space lattice of the crystal and the observed diffraction pattern. The reciprocal lattice is constructed by taking normals to all direct lattice planes from an arbitrary origin and terminating them at a distance equal to $1/\,d_{hkl}$ from the origin. Each plane (*hkl*) produces a point in reciprocal space designated *h, k, l*. The resulting reciprocal lattice has spacing that is inversely proportional to the interplanar spacing within the unit cell, with units inverse to the real cell dimensions.

P.P. Ewald linked Bragg's law to reciprocal space and diffraction (170). If Bragg's reflection angle is equal to θ, then the total angle of deflection of a constructive reflection from the incident X-ray beam is equal to 2θ. Whenever a crystal is rotated about an origin such that a reciprocal lattice point comes into contact with a theoretical circle of radius $1/\lambda$, Bragg's law is satisfied and a reflection occurs (Figure 3). The Ewald sphere can be rotated around an origin such that an area around the origin with potential to diffract can be strictly defined as a limiting sphere of $2/\lambda$ (Figure 4). The total number of reflections that can theoretically be produced is equal to the number of reciprocal lattice points within the limiting sphere. Ewald's model of diffraction therefore implies that the directions of each reflection as well as the number of potential reflections rely only upon the unit cell dimensions.

Unlike the number and position of reflections which are solely dictated by the dimensions of the unit cell, the intensity of a reflection is related to the *content* of the unit cell. Because X-rays are scattered by electrons, the intensity of diffraction data is related to the electron density within the unit cell. Therefore, a relationship between the reflection intensity and electron density must be produced to successfully interpret diffraction data. Each diffracted X-ray is a wave and thus periodic in nature. Simple wave equations can be written in the form

$$f(x) = F \cos 2\pi(hx + \alpha)$$

or

$$f(x) = F \sin 2\pi(hx + \alpha),$$

where $f(x)$ specifices the vertical height of the wave at the horizontal position x as measured in wavelengths, F represents the amplitude of the wave, h the frequency and α the phase. The phase defines the position of the wave with respect to the origin of the coordinate system. A complex periodic function can be described by

$$f(x) = F_h[\cos 2\pi(hx) + i \sin 2\pi(hx)]$$

in which the phase α is implicit in the complex coefficient, F_h. Since complex number theory defines $\cos\theta + i\sin\theta = e^{i\theta}$ and in crystallography θ is substituted with $2\pi(hx)$, a wave can be redefined as

$$f(x) = F_h e^{2\pi i(hx)}$$

The French mathematician Fourier showed that complex periodic functions can be described as the sum of multiple wave equations. The Fourier series that describes a diffracted X-ray is called a structure factor equation, F_{hkl}, in which each term gives the contribution of an individual atom to the final reflection. Thus each reflection is the

result of diffractive contributions of each atom in the cell. The structure factor is represented by

$$F_{hkl} = \Sigma_j f_j e^{2\pi i (hx_j + ky_j + lz_j)}$$

where f_j is the scattering factor of the jth atom, (x_j, y_j and z_j) are the coordinates of the jth atom in the unit cell and *hkl* denotes the indices of each reflection in reciprocal space.

Alternatively, F_{hkl} can be expressed as the sum of contributions from each volume element of electron density in the unit cell, such that electron density at (x, y, z) can be represented by the average electron distributions ρ(x, y, z) of the area. Thus,

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i (hx + ky + lz)} dV$$

where the V represents the volume of the unit cell.

Fourier demonstrated that for any function $f(x)$ there exists another function $F(h)$ such that

$$F(h) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i (hx)} dx,$$

where $F(h)$ is the Fourier transform (FT) of $f(x)$. Therefore, in order to compute $F(h)$, one simply multiples $f(x)$ by the function $e^{2\pi i (hx)}$ and integrates the function with respect to x. The units of the variable h are reciprocal to the units of x: in this way, crystallographers use the Fourier transform to mathematically relate real and reciprocal space. Examination of the structure factor equation reveals that F_{hkl} is the FT of ρ(x, y, z). The Fourier transform is also reversible, such that

$$f(x) = \int_{-\infty}^{+\infty} F(h) e^{-2\pi i (hx)} dh.$$

Back transformation of the structure factor equation yields an equation describing the electron density:

$$\rho(x, y, z) = \left(\frac{1}{V}\right)\Sigma_h \Sigma_k \Sigma_l F_{hkl}\, e^{-2\pi i(hx+ky+lz)}$$

This transform is a triple sum rather than a triple integral because the F_{hkl} terms represent

discrete reflections of the diffraction pattern. Since $F_{hkl} = |F_{hkl}|e^{i\alpha_{hkl}}$, the electron density

can be represented as

$$\rho(x, y, z) = \left(\frac{1}{V}\right)\Sigma_h \Sigma_k \Sigma_l |F_{hkl}|\, e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}}$$

The structure factor amplitude $|F_{hkl}|$ can be obtained from the measurements of reflection

intensities, since

$$I_{hkl} \propto |F_{hkl}|^2$$

However, the phase α_{hkl} can not usually be obtained from diffraction reflections.

Therefore, in order to compute the electron density $\rho(x,y,z)$, the phase of each reflection

must be determined. This is known as the phase problem. There are several methods for

determining phases in crystallography experiments. Molecular Replacement (MR) was

performed in this thesis. In MR, initial phases are calculated from a known structure with

some structural similarity to the target protein, which is often predicted by sequence

homology. MR will be described in more detail in subsequent sections.

Determining a Structure by Macromolecular X-Ray Crystallography

The process of solving a protein structure using macromolecular X-ray crystallography

requires several steps:

  1) Growth of high quality crystals,

2) X-ray data collection

3) Determination of phases, and

4) Refinement and validation.

1. Growth of High Quality Crystals

Ironically, the first step in structure determination is the least well understood portion of the entire process. Crystals are essentially formed by controlled precipitation of protein out of solution. Although progress has been made towards developing a detailed understanding of this procedure, successful crystallization is still largely a trial-and-error procedure. High quality protein is extremely important, as impurities may prevent growth of quality crystals. Crystals of proteins are grown by slow, controlled precipitation from an aqueous solution. Precipitants must be used under conditions that do not denature the protein; polyethylene glycol polymers are often used as they are powerful precipitants and weak denaturants.

Vapor diffusion is a widely used technique in which the protein is mixed with a precipitant solution (mother liquor) and allowed to equilibrate in a sealed container with a large aqueous reservoir of precipitant solution. Vapor diffusion within the closed system results in net transfer of water from the protein/mother liquor drop to the aqueous reservoir. Since the reservoir is much larger than the protein drop, the final concentration of precipitant in the drop will be very near to that of the reservoir. As water leaves the protein drop, the volume of the drop decreases which results in an increase in protein and precipitant concentration. Once protein and precipitant reach high enough concentrations, molecules can form clusters from which crystals grow, a process called

nucleation. After nucleation has begun, subsequent precipitation should be controlled to maximize crystal growth. A balance between nucleation and growth must be found: too much nucleation results in small crystals, while uncontrolled growth results in crystals with high levels of internal disorder. As crystals grow, the local protein concentration decreases, ideally moving drop conditions away from nucleation into growth. Once the system reaches equilibrium, vapor diffusion ceases.

Optimal crystallization conditions are dependant on a variety of parameters including protein concentration, precipitant, pH, temperature and the use of additives. Several commercial screens are available to rapidly search for initial conditions. After conditions are found that yield crystals, conditions can be optimized until well ordered crystals of good size have been obtained. Despite all of the advances in screening technology, macromolecular crystallization remains a mostly empirical procedure.

2. X-ray Data Collection

To generate X-rays in a laboratory suitable for protein crystallography, electrons are produced by a heated filament and accelerated by an electric field towards a metal target. Collision of a high-energy electron with an electron from a low-lying orbital displaces the lower energy electron. An electron from the higher energy orbital subsequently drops into the resulting vacancy and emits excess energy as an X-ray. Electron bombardment of target metals yield characteristic emission spectra related to the energy-level spacing of that element. X-rays can then be further filtered to obtain a monochromatic wavelength X-ray. Copper anodes are a common target metal used as a source of experimental X-rays. Electrons dropping from the L shell of copper to replace

displaced K electrons (an L $\rightarrow$ K or K_α transition) emit X-rays of $\lambda = 1.54$ Å. Wavelenths of 1.39 Å are produced when electrons dropping from the M shell (M $\rightarrow$ K or K_β transition); however, X-rays at these wavelengths are eliminated by the use of a Ni filter. The experiments featured in this thesis use a RIGAKU RU200-H rotating copper anode generator.

Although not used in these studies, synchotron radiation has increasingly been used as a powerful and tunable X-ray source. Particle storage rings circulate electrons or positrons at velocities near the speed of light. Powerful magnets are used to keep the particles moving in a circular fashion. A charged body like an electron emits energy when forced into a curved motion, and the energy is emitted as X-rays. Undulators, wigglers and monochromators at ports tangental to the particle storage ring allow crystallographers access to monochromatic X-rays that can be tuned to selected wavelengths. The tunability of X-rays is a particularily valuable feature when used in multiwavelength anomalous diffraction (MAD) phasing. MAD techniques were not used in these studies, although crystals suitable for this form of crystallography were grown, should further experimentation be desired.

Once suitable crystals have been grown, the next step is to obtain initial images. Protein crystals are sensitive to X-rays, and can suffer from radiation damage during data collection. Experimentation at cryogenic temperatures greatly reduces damage due to X-ray exposure. Simply freezing a crystal will result in damage due to ice formation; crystallographers use a variety of cryoprotectants to prevent ice formation and keep a crystal intact. Crystals can be either grown directly into cryoprotectant or dipped in a cryoprotectant prior to being frozen and exposed to X-rays. The choice of

cryoprotectant, much like the formulation of crystallization conditions, is a matter of trial and error. Once suitable cryo conditions have been found, the crystal is then mounted onto a goniometer which allows any reciprocal lattice point to rotate through the sphere of reflections.   The crystal is cooled in a stream of 100ºK gas derived from a reservoir of liquid nitrogen. X-rays are then directed towards the cryo-cooled crystal, and the resulting diffraction image is captured with a charged coupled device (CCD). CCDs are essentially photon counters that accumulate electrons in direct proportion to the amount of light exposure they receive. Since CCDs cannot directly detect X-rays, phosphors that produce visible light in response to X-ray absorption are used. With an X-ray source, cryo-cooled goniometer crystal mount and a CCD area detector, initial images can be collected.

The first images collected are used to determine crystal orientation, cell parameters and space group. Spatial organization of the diffracted rays are related to the physical properties of the unit cell via the reciprocal lattice, which means that the dimensions of the unit cell can be calculated if the crystal to detector distance and X-ray wavelength is known. Historically, photographs along zero-planes were used to calculate unit-cell dimensions, but these calculations are now swiftly performed with modern data collection software. Additionally, some preliminary measures of diffraction quality can be observed. The mosaicity of a crystal is a measure of its internal disorder, and can be estimated along with the overall resolution of the dataset using initial images. Because crystals are not perfectly packed, the resulting reflections are not points but are spherical in shape. To capture the entire reflection, the reflection must be measured over a small angular range by rotating the crystal to cover the entire peak width.

Once the unit cell dimensions and an estimate of mosaicity have been determined, the symmetry of the cell is examined. Recall from 'Geometric Principles of Diffraction' that symmetry within the unit cell will create a number of equivalent reflections. Since these points are redundant, symmetry thus reduces the total number of number of reflections that need to be recorded. Symmetry elements are identified by systematic absences in the diffraction pattern, locations where symmetry-related pairs effectively cancel each other out. All possible symmetry elements and combinations of symmetry elements have their representative systematic absences tabulated for each space group. The process of identifying and assigning parameters and symmetry elements is called indexing, which ultimately assigns each reflection a Miller index based on the crystal's space group.

A final parameter that can be estimated from the unit cell volume is the number of molecules in the unit cell (Z) (171). B.W. Matthews found that for most protein crystals, the ratio of the unit cell volume to the protein molecular weight is between 1.7 and 3.5 $\text{Å}^3/\text{Da}$. This number is defined as V_m, and helps determine how many molecules are contained within the asymmetric unit:

$$V_m = \frac{volume}{MW * Z * X}$$

such that the *volume* represents the volume of the unit cell, MW is the molecular weight of the macromolecule, Z is the number of asymmetric units in the unit cell (which is related to the number of symmetry operations in the space group) and the unknown variable X which represents the number of molecules in the asymmetric unit (171). The most probable values of X are those which give V_m coefficients within the empirically observed range.

After initial parameters have been defined, a data collection strategy is designed. The goal of a good strategy is to generate an efficient and complete data set. A balance between data quality, completeness of the dataset, data redundancy and crystal stability must be found. After the data has been collected, images must be processed to improve consistency and have corrections applied by procedures called integration and scaling. The first step is to merge and scale all of the obtained images, where redundant reflections are identified and all reflections are scaled to a consistent intensity range. Post refinement then optimizes unit cell dimensions, crystal orientation, experimental parameters and mosaicity by maximizing the agreement of equivalent partial reflections. Following post refinement, usable data from partially measured reflections can be recovered. The partiality of a reflection can be used as a correction factor to covert measured intensity to an estimated intensity for a fully recorded reflection. R_{merge}, a measure of overall data quality, can be calculated by comparing symmetry related reflections, which would be identical under ideal conditions. For n independent reflections and i observations of a given reflection,

$$R_{merge} = \frac{\Sigma_{hkl}\Sigma_i |I_i - \langle I \rangle|}{\Sigma_{hkl}\Sigma_i I_i}$$

Intensity data are then converted into structure factor amplitudes,

$$|F_{hkl}| = \sqrt{I_{hkl} / Lp)}$$

where L is the Lorentz factor that account for the rate at which a reflection passes through Ewald's sphere and p is the polarization factor that accounts for variations in reflection efficiency that vary with reflection angle.

3. Phase Determination

Recall from 'Geometric Principles of Diffraction' that the contour map of electron density $\rho(x,y,z)$ can be represented as a Fourier sum using the structure factors F_{hkl}. Each structure factor is a wave equation that has a frequency, amplitude and phase. The frequency is known and the amplitude can be calculated from the intensity of each reflection; the phase alone remains to be determined. Correct phases are critical to structure determination. It has been shown that incorrect structure factor amplitudes combined with correct phases may lead to an interpretable map, whereas correct structure factor amplitudes combined with incorrect phases lead to meaningless maps (165). A physical understanding of phase will better demonstrate its importance. The Fourier transform of individual reflections represents the average electron density that falls between sets of Miller planes. The electron density itself is a periodic rise and fall of electron density that may not pass precisely through the origin of the unit cell. That is to say, the peak of electron density does not necessarily fall along the faces of the Miller planes. Only if the phase is zero does the electron density peak lie along the planar face. If the phase is 180º, then the peak electron density falls directly in between Miller planes. Phase is essential to precisely orient the electron density within the planes. Since phase is specific for each set of Miller planes, each reflection has its own phase; therefore, the phase must be solved for every single reflection.

If a similar structure is known, the phases from the known protein can be used as initial estimates for phases of the target protein. This method is known as molecular replacement (MR) (172). A brief description of MR follows. A more in-depth discussion

on MR techniques including a more detailed description of the relevant mathematical functions employed can be found elsewhere (167; 173; 174).

MR has the advantage of only requiring a native data set, but care must be taken to avoid phase bias. A preliminary model of the unknown crystal structure is obtained by first rotating and then translating the model molecule in the crystal lattice. Once the correct orientation and position are found, the phases from the model can be combined with the observed structure factor amplitudes. The resulting structure factors can then be used for further refinement.

Molecular replacement begins with a Patterson function, which is defined as

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l \left| F_{hkl} \right|^2 e^{-2\pi i (hu + kv + lw)} ,$$

where the coordinates (u, v, w) locate a point on the Patterson map. The amplitude of each term is a complex conjugate of the structure factor which is proportional to the reflection intensity in a manner that is independent of phase. The Patterson map displays peaks corresponding to vectors between atoms, which means that a molecule with N atoms will have N(N-1) peaks. For a small number of molecules, a Patterson map can be used to solve for the location of objects within a unit cell; in fact, Patterson maps are used to solve for the locations of heavy atoms in alternative phasing techniques. For most macromolecules, however, Patterson maps are far too noisy to generate high resolution structural information.

In MR, the phasing model needs to be correctly oriented and positioned. This is performed by performing two three-dimensional searches, in which the first search finds the rotational angles κ, φ, ψ and the second to determine the three translational components x, y and z. In the rotation function, the Patterson map of a rotated search

model is compared to a Patterson map calculated from the experimental data of the unknown structure. The rotation function will exhibit maxima where the two Patterson maps have many coincident peaks. In selecting mostly interamolecular vectors, the Patterson is limited to peaks within a radius equal to the dimensions of the probe model. The rotation function is successful because intramolecular vectors are dependent only on molecular rotation, not molecular translation.

The original rotation function as defined by Rossmann and Blow required a significant number of calculations. In 1971 Crowther outlined a new approach which describes the Patterson using spherical harmonics and uses a fast Fourier transformation for quick calculation. This calculation, also called the Crowther fast rotation function, has been refined by Jorge Navaza in the molecular replacement suite AMoRE which was used in this thesis (173).

If there are molecules related by symmetry within the unit cell, a translation search will also need to be performed. A translation search is also required if several molecules are present in the asymmetric unit. Location of the first copy fixes the origin and subsequent searches are performed across the entire unit cell to locate the remaining molecules in the asymmetric unit. For the translation search, the now correctly oriented model can be positioned by translating the model within the unit cell. Translation functions use sets of intermolecular vectors, which depend on the position of the molecules in the unit cell. Since self vectors remain the same, they can be subtracted to simplify calculations. There are several translation functions to choose from, but the overall goal is to maximize overlap between the observed and calculated datasets. Successful translation searches can be monitored using an *R*-factor,

$$R = \frac{\Sigma \left\| \mathbf{F}_{obs} \right| - \left| \mathbf{F}_{calc} \right\|}{\Sigma \left\| \mathbf{F}_{obs} \right\|}$$

where $|\mathbf{F}_{obs}|$ are native dataset amplitudes and $|\mathbf{F}_{calc}|$ are calculated amplitudes based on the model in its trial location. If the observed and calculated intensities are similar, their *R* value will be small.

4. Refinement and Validation

        Once solutions for rotation and translation functions have been obtained, they are applied to the probe model to create a starting model, which is oriented and positioned in the unit cell of the target protein and used for refinement. The phases from the model are then used to calculate electron density maps using the experimentally derived intensities. Refinement is an iterative process where the atomic model is adjusted to maximize its agreement with the electron density maps. The final atomic model should be consistent with the electron density and have regular molecular structure and stereochemistry. When a structure is determined by MR, the relative atomic positions are good but their absolute positions will need adjustment. Refinement generally begins with rigid body refinement in which the molecule is rotated and translated as a single unit. This may be performed with subdomains as the orientation of the molecule improves. Rigid body refinement is usually carried out with low resolution data, since the gross secondary and tertiary structural features are of primary interest. Once rigid body refinement has reached convergence, positional and temperature factor optimization can begin. The resolution at this point will be extended to the resolution limit.

        The R-factor is a measure of convergence between the model. A completely random solution would have a value of 0.59, while a perfect model would have an R-

factor of zero. Another more unbiased way of judging the convergence of the model and the observed data is with R_{free}, where a random portion of the model has been withheld from the calculation of improved phases and is used to judge the unbiased quality of the model. At the end of the refinement process, when the R-factor has converged to an acceptable point, ordered water molecules may be added to the model. The resulting model can be validated against established parameters.

a)

b)
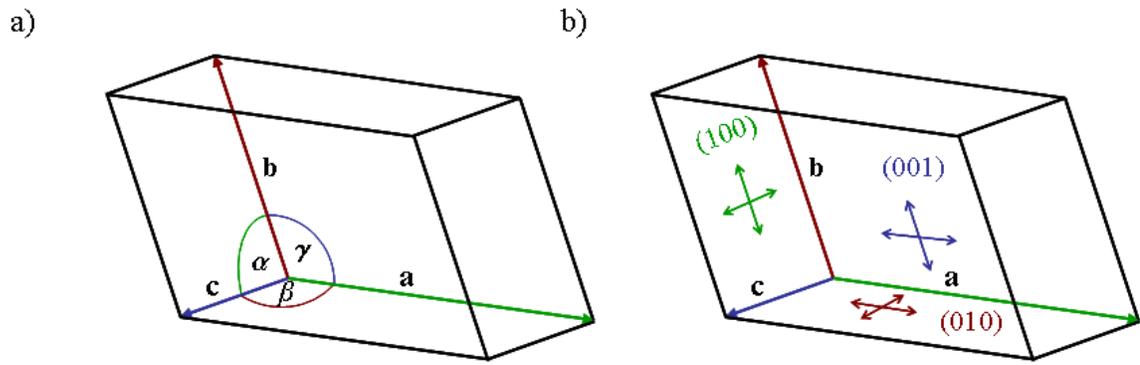
Figure 1. Unit cell parameters (a) General unit cell with edges **a**, **b**, **c** and angles α, β, γ.
(b) Indicies of faces. The (100) plane is parallel to the **bc** face and normal to the *x*-axis,
the (010) plane is parallel to the **ac** face and the (001) plane is parallel to the **ab** face.
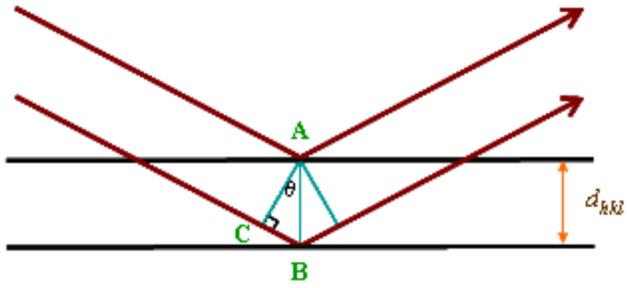
$$\sin\theta = BC/AB$$

$$BC = AB\sin\theta = d_{hkl}\sin\theta$$

$$\text{Difference in path length} = 2BC$$

For constructive interference:
$$2BC = n\lambda$$

$$2BC = 2d_{hkl}\sin\theta$$

Therefore, $\mathbf{2d_{hkl}\sin\theta = n\lambda}$

Figure 2. Derivation of Bragg's Law.
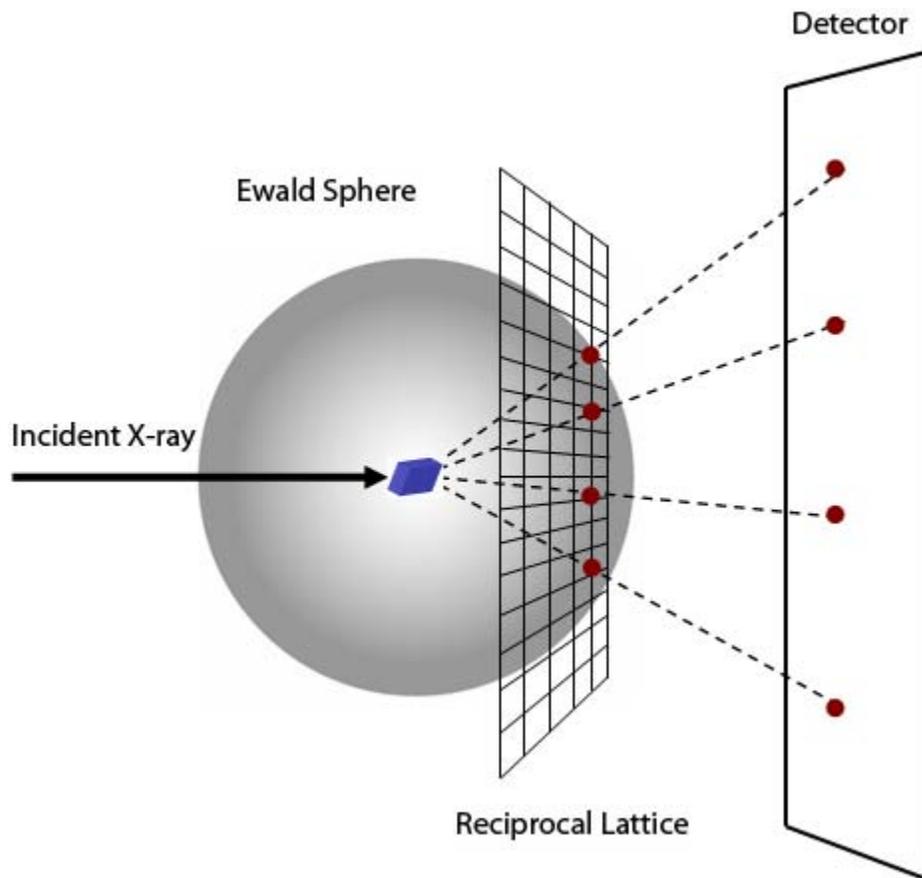
Figure 3. Depiction of the Ewald Sphere. As incident X-rays strike a crystal, the diffracted rays intersect with a hypothetical reciprocal lattice construct. When the diffracted rays come in contact with a reciprocal lattice point, constructive interference occurs and a diffraction spot can be detected.
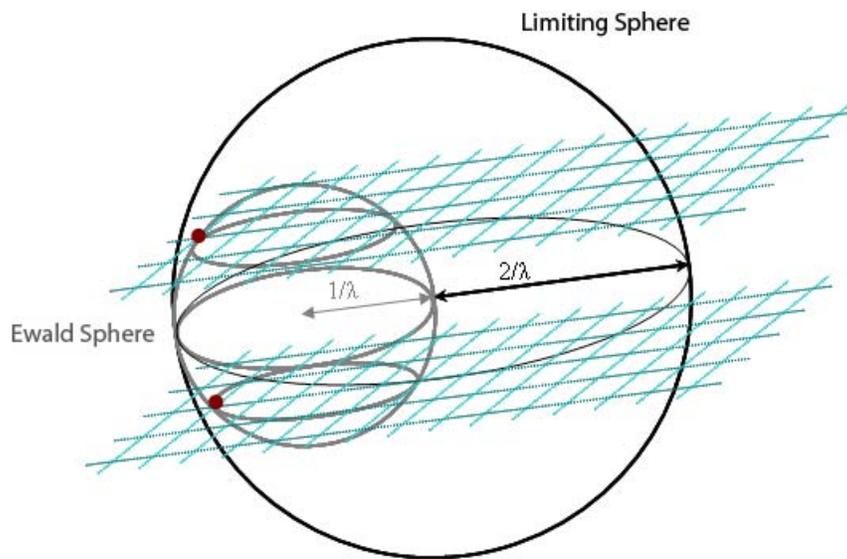
Figure 4. Depiction of the Ewald sphere and the Limiting Sphere. The reciprocal lattice is shown in teal, the Ewald sphere is shown in grey. Any reciprocal lattice point within the limiting sphere can form a reflection.

CHAPTER 3

Crystal Structure of the Human Collagen XV Trimerization Domain: A Potent Trimerizing Unit Common to Multiplexin Collagens

Jacqueline A. Wirz[1,2], Sergei P. Boudko[1,2], Thomas F. Lerch[2], Michael S. Chapman[2] and Hans Peter Bächinger[1,2*]

[1] Research Department of Shriners Hospital for Children, 3101 SW Sam Jackson Pk. Rd., Portland, OR 97239, USA

[2] Department of Biochemistry and Molecular Biology, Oregon Health and Science University, 3191 SW Sam Jackson Pk. Rd., Portland, OR 97239, USA

Explanation of Contribution

       I performed all experiments included in this experiment. Drs. Boudko and Lerch provided technical assistance.

Abstract

Correct folding of the collagen triple helix requires a self-association step which selects and binds α-chains into trimers. Here we report the crystal structure of the trimerization domain of human type XV collagen. The trimerization domain of type XV collagen contains three monomers each composed of four β-sheets and an α-helix. The hydrophobic core of the trimer is devoid of solvent molecules and is shaped by β-sheet planes from each monomer. The trimerization domain is extremely stable and forms at picomolar concentrations. It is found that the trimerization domain of type XV collagen is structurally similar to that of type XVIII, despite only 33% sequence identity. High structural conservation indicates that the multiplexin trimerization domain represents a three dimensional fold that allows for sequence variability while retaining structural integrity necessary for tight and efficient trimerization.

Introduction

Collagens are a diverse family of proteins that constitute the major structural component of the extracellular matrix. Although collagen proteins are identified by the presence of one or more triple-helical domains, each α-chain also contains two or more non-collagenous domains. To date, 28 different types of collagen have been identified (175). Further classification according to supramolecular structure assigns collagens to fibril, fibril-associated containing interrupted triple helicies (FACIT), beaded filament, anchoring fibril, network-forming, transmembrane or multiple triple helicies with interruptions (multiplexin) families (8). Each type of collagen is composed of one, two or three different α chains that can assemble into hetero- or homo-trimers. Assembly of

all collagens begins with a self-association step that selects, aligns and trimerizes α-chains.

The mechanism of trimerization is attributed to non-collagenous (NC) domains (80). Studies on classic fibril-forming collagens I and III found that the carboxy-terminal NC (NC1) domains were essential to trimerization (81; 82). C-terminal NC domains govern chain selection for all collagen families except the transmembrane collagens, which are believed to trimerize via amino-terminal NC domains (104). Fibril associated-collagens IX and XIX have recently been shown to trimerize via their NC2 domains (101; 176). The NC2 domains of the other FACIT family members are also most likely responsible for self association. For all remaining collagens, the NC1 domain is believed to be responsible for chain selection and trimerization.

Crystal structures for the NC1 domains of network-forming collagens IV, VIII and X have been solved (91; 94; 92; 93). Along with the NC1 domain of the multiplexin type XVIII collagen (177), these represent the only known atomic structures for collagen trimerization domains. Although both the multiplexin and network-forming collagen trimerization domains are composed primarily of β-sheets, they share no overall structural homology.

Collagens XV and XVIII are the only known members of the multiplexin family (111). Both collagens are homo-trimers composed of a single α-chain that contains a central highly interrupted collagenous domain flanked by N- and C-terminal non collagenous domains (126; 139; 112). The collagenous domain of human type XV is divided into nine collagenous domains with eight non-collagenous repeats. Human type XVIII exists as two variants with differing N-terminal NC domains, but have identical

interrupted collagenous and NC1 domains. There are ten collagenous sequences alternating with nine non-collagenous repeats in the type XVIII interrupted collagenous domain (109). Type XVIII is a heparan sulfate proteoglycan (178) and type XV can carry chondroitin/dermatan sulfate alone or chondroitin/dermatan sulfate and heparan sulfate chains (128). Type XVIII collagen localizes in epidermal and vascular basement membrane, similar to other heparan sulfate proteoglycans found in basement membranes (149). Type XV is not an integral basement membrane component; rather, it localizes to areas peripheral to basement membranes and associates directly with collagen fibers/fibrils in a manner consistent with other chondroitin sulfate proteoglycans (128; 179).

Collagen XV null mice develop collapsed capillaries and degenerating endothelial cells in the heart and skeletal muscles and increased sensitivity to exercise-induced muscle damage (136). However, there are currently no known diseases states that are associated with type XV collagen. Although COL15 was shown to be upregulated in patients with systemic sclerosis, polymorphisms in the COL15 gene were shown to not be associated with the disease (180). In contrast, the human autosomal recessive Knobloch syndrome has been mapped to the COL18A1 gene, which leads to an early stop codon. Knobloch syndrome manifests with severe myopia, retinal degeneration and retinal detachment leading to blindness. $Col18a1^{-/-}$ mice display impaired blood vessel development in the eye (146), as well as hydrocephalus and structurally altered basement membranes (181). Double knockout mice for $Col18a1$ and $Col15a1$ did not demonstrate any functional compensation between collagens XV and XVIII.

The highest level of sequence homology between type XV and XVIII lies in their NC1 domain. The NC1 domains of collagens XV and XVIII are organized into a trimerization domain, a hinge region and an endostatin domain (117; 119). The trimerization domain of type XVIII collagen has been isolated, characterized and crystallized (177). In this study we present the crystal structure of the trimerization domain of type XV collagen along with biochemical analysis of the protein stability and oliogmeric state. Despite having only 33% sequence identity, the type XV structure is remarkably similar to and displays comparable biochemical properties to the type XVIII trimerization domain. The multiplexin trimerization domain (TD) is therefore a structurally conserved domain that trimerizes at very low concentrations and may thus be a useful tool in the engineering of protein trimers.

Experimental Procedures

Cloning, Expression and Purification

The trimerization domain of human α1(XV) sequence was synthetically constructed by Integrated DNA Technologies. The pIDTSMART-KAN vector contained the coding sequence with an embedded BamHI site and the 5' end and a flanking Sal1 end after the stop codon at the 3' end. Plasmids were expressed and purified by Midi Prep (Qiagen), the insert was excised with BamHI + Sal1 and cloned into the pET23-HisTag-Trx vector. The DNA inserts were verified by Sanger dideoxy DNA sequencing.

The recombinant protein was grown in a *E. coli* BL21(DE3) strain (Novagen) at 37°C. Once the cells reached an $OD_{600} \sim 0.6$, cells were induced with IPTG to a final concentration of 1 mM. The temperature was lowered to 30°C and growth proceeded for

an additional 6 hours prior to harvest. Purification of His and thioredoxin tagged protein from soluble cell lysate was performed using immobilized metal affinity chromatography on a Chelating HP Column (GE Healthcare) charged with $NiCl_2$ following the manufacturer's instructions. Tagged protein was cleaved with 1U/mg thrombin (ICN) or 0.1 mg/mL trypsin in 50 mM Tris (pH 8.0) buffer with 150 mM NaCl. The resulting fragments were separated with a Chelating HP Column as above. The non-tagged flow-through was further purified with a HiTrap Q Column (GE Healthcare) in 50 mM Tris pH 8.0 using a salt gradient from 0 to 0.5M NaCl. The resulting purified proteins were analyzed for sequence by amino acid analysis and correct masses were confirmed by mass spectroscopy.

The trimerization domain (TD) of human collagen XVIII was cloned into the pET23-HisTag-Trx vector and purified as above.

Gel Filtration Chromatography and Analytical Ultracentrifugation

Gel filtration chromatography was performed using a 10/300 Superose 6 column (GE Healthcare) at 0.5 mL/min in phosphate buffered saline. Gel filtration standards (BioRad) were run to calibrate the column prior to injection of the XV(TD). Absorbance was monitored at 280 nm. Sedimentation equilibrium experiments were performed with a Beckman model XLA analytical centrifuge. Runs were carried out using an An60-Ti rotor with 12 mm cells and Epon two-channel centerpieces at 20ºC in phosphate buffered saline. Absorbance was monitored at 280 nm. V-bar values were calculated using UltrascanII (182).

Crystallization and Data Collection

Crystallization was performed using hanging-drop vapor diffusion by mixing 1.5 µl of reservoir solution with 1.5 µl of protein solution at 5 mg/mL. The best crystals were grown at 25ºC using a reservoir solution of 0.1 to 0.2 M $MgCl_2$, 0.1 M Tris (pH 7.0 – 8.0) and 28 – 32% (w/v) polyethylene glycol 3350. The crystals grew to a final size of roughly 0.25 x 0.25 x 0.25 mm after 7-14 days. Immediately prior to X-ray exposure, the crystals were dipped into a cryoprotectant solution containing the reservoir solution supplemented with PEG 400 to a final concentration of 20% (v/v) PEG 400. Data collection was performed using a RIGAKU RU200-H rotating anode generator. Data statistics are summarized in Table 1.

Crystal Structure Determination

The images collected were indexed, integrated and scaled using MOSFLM and SCALA from the CCP4 suite (183). The AMORE (184) program was used within the CCP4 suite (183) to find initial molecular replacement solutions using the type XVIII structure (PDB 3HSH). Iterative cycles of model-building and refinement were performed using COOT (185) and CNS (186).

Circular Dichroism and Fluorescence Analysis

CD Spectra were recorded on an AVIV model 202 spectropolarimeter (AVIV Instruments, Inc.). Quartz cells of 0.05- to 0.1-mm path length were used (Hellma), depending on protein concentration. Spectra were buffer subtracted and normalized for concentration and path length to obtain the mean molar residue ellipticity. Protein

concentrations were determined by amino acid analysis. Fluorescence spectra were recorded on a LM8000C instrument (SLM Instruments) with modified electronics (ISS Corp., Champaign, IL) using 104F-QS cells (Hellma). Emission spectra were obtained by excitation of protein samples at 280 nm at 25 C.

GuHCl Induced CD & Fluorescence Transitions

Folded protein samples were in 50 mM sodium phosphate buffer (pH 8.0) at a stock concentration of 260 μM. Unfolded protein samples were at the same concentration in 50 mM sodium phosphate buffer (pH 8.0) supplemented with 4 M GuHCl. Proteins were diluted to a final concentration of 26 μM into buffers of varying concentrations of GuHCl. Final GuHCl concentrations were determined using refractive indices. Change in secondary structure with varying concentrations of GuHCl was monitored by CD at 224 nm. Change in tertiary structure with varying concentrations of GuHCl was monitored by a peak shift in fluorescence emission spectra from 320 nm to 347nm.

To calculate the standard free energy ($\Delta G°$), unfolding and refolding curves were interpreted using a two state system where unfolded monomers (U) combine into a native trimer (N) in a cooperative mechanism:

$$N \leftrightarrow 3U$$

Each GuHCl concentration can be used to calculate an equilibrium constant (K_{eq}):

$$K_{eq} = [U]^3/[N] = 3f_U^3[M]_0^2/(1-f_U)$$

where $[M]_0$ is the total monomer concentration ($[M]_0=[U] + 3[N]$) and f_u is the fraction of unfolded monomer ($f_u=[U]/[M]_0$). The standard free energy $\Delta G°$ is related to K_{eq} by:

$$\Delta G^0 = -RT \ln K_{eq}$$

The denaturant dependence of ΔG° is linear and can be described by:

$$\Delta G^0(x) = \Delta G^0(H_2O) + mx$$

where $\Delta G^\circ(H_2O)$ is the free energy for unfolding in water.

The calculated concentration at which half of the total chain concentration is incorporated into a trimer $(3[N]_{1/2} = [U]_{1/2})$ can be derived from K_{eq}:

$$K_{eq} = [U]_{1/2}^3 / [N]_{1/2} = 3[U]_{1/2}^2$$

$$[U]_{1/2} = (K_{eq}/3)^{1/2}$$

Protein Data Bank Accession Numbers

The refined atomic model and the observed structure factors were deposited into the RCSB Protein Data Bank with the accession number 3N3F.

Results

Crystallization and Structure Determination

The trimerization domain of human type XV collagen, designated XV(TD), was designed to cover the homologous sequence to the NC1(54) peptide used in the crystal structure of the collagen XVIII trimerization domain (177). The XV(TD) construct covers amino acids 587 through 653 of the full length protein (126) and is numbered such that the first visible residue is labeled as residue 1 (Fig. 1).

The XV(TD) protein crystallized in a hexagonal crystal form. The trimerization domain of type XVIII (PDB 3HSH) was used for molecular replacement (see Experimental Procedures for details). Two single chains were found per asymmetric

unit; two trimers were related by the 3-fold crystal symmetry. Each monomer is comprised of one α-helix and four β-strands in a mixed parallel and antiparallel β-sheet (Fig. 2). The α-helix interacts with all four strands of the β-sheet to shape a hydrophobic core within each monomer. The core is comprised of residues F6, M9, M12, L13, A16, V19, I25, L27, F34 and is stabilized by the hydrophobic parts of residues T4, T23, R36 and W41 (Fig. 3).

The trimer measures approximately 40 Å in length and 20 Å in height. The trimer is stabilized by a central hydrophobic core, interchain hydrophobic interactions between adjacent monomers, and an interchain hydrogen bond network (Fig. 3). The trimer core is comprised of residues V3, F24, F33, I35, V37, L44, L46 and the hydrophobic face of Y26 (Figs. 2 and 3). No solvent molecules are observed in the trimer interior. Residues I50 and I52 interact with the monomeric hydrophobic core of the adjacent subunit. Interchain hydrogen bonds also stabilize trimerization, forming a net of bonds that cover the carboxyl-terminal end of the trimer. The majority of the hydrogen bonds connect two adjacent chains, specifically the hydrogen bonds between the backbones of W4 to T50 and K45 to G47; however, at the core of the hydrogen bond network, the Q45 backbone forms hydrogen bonds with both adjacent Q45 backbones.

Oligomeric State

Gel filtration chromatography was performed to assess the oligomeric state of the XV(TD) protein. The protein eluted as a single peak with a molecular weight of approximately 18 kDa (Fig. 4), which corresponds to a trimer. Chemical crosslinking was performed with glutaraldehyde and analyzed by SDS-PAGE (Fig. 4). The XV(TD)

readily crosslinks to form trimers The XVIII(TD) protein has only two potential sites for crosslinking which results in little formation of crosslinked trimer under similar conditions when compared to the XV(TD) protein. Analytical ultracentrifugation was performed in phosphate-buffered saline, and results independently confirm the trimeric state of purified XV(TD) in solution. The trimerization domain of type XVIII collagen was also analyzed by ultracentrifugation, and the results are also consistent with a trimer in solution (Table 2).

Denaturant Induced Equilibrium Transitions

Biophysical properties of the XV(TD) protein were tested using CD and fluorescence. Figure 5a shows the CD spectrum of the XV(TD) protein in 50 mM sodium phosphate buffer pH 8.0. A positive peak is located at 224 nm and a slight negative peak is located at 236 nm. The maximum positive CD signal at 224 nm was observed for initial monitoring of guanidium chloride (GuHCl) induced transitions. Complete unfolding as assessed by loss of signal at 224 nm was observed at 4 M GuHCl (data not shown).

Fluorescence spectra of the XV(TD) protein in 0 and 4 M GuHCl buffer show a shift in the emission peak from 320 nm to 347 nm (Fig. 5b). Denaturant induced equilibrium transitions were monitored as a function of peak shift at a final protein concentration of 26 μM. Initial refolding samples were prepared at a concentration of 260 μM in 4 M GuHCl, 50 mM phosphate pH 8.0 buffer. Samples were subsequently refolded by dilution to a final concentration of 26 μM into buffers with a final GuHCl concentration of 0.4 to 4 M GuHCl. Samples used for unfolding were prepared in a

buffer with no GuHCl and diluted to a final protein concentration of 26 µM into buffers with a final concentration of GuHCl of 0 to 3.6 M. The agreement between folding and unfolding curves indicates that refolding is reversible.

The standard free energy of trimerization ($\Delta G°$) in water was calculated to be 109.0 (±1.0) kJ/mol. The calculated concentration at which half of the total chain concentration is incorporated into a trimer is equal to 119 pM for the collagen XV(TD). This is similar to the calculated value of 56.6 pM for the midpoint concentration of the minimal trimerization of type XVIII collagen.

Discussion

Collagen molecules are characterized by the presence of one or more triple helical regions formed by a repeating Gly-X-Y sequence.   Much research has been performed to better understand the structure and function of the classic collagen triple helix (8). However, in recent years more attention has been paid to non-collagenous (NC) domains found in all collagen molecules. The NC domains for many collagen families have been shown to mediate trimerization. Here we present the crystal structure and biophysical characterization of the trimerization domain for human type XV collagen. The trimerization domain is structurally and biophysically very similar to the trimerization domain of type XVIII collagen.

Four crystal structures of collagen trimerization domains are known to date: the NC1 domain of network forming collagens IV, VIII and X (91; 94; 92; 93) as well as the only other known multiplexin, collagen XVIII (177). The trimerization domains of the multiplexin family are ~50 residues in length, which are significantly smaller than the

trimerization domains for the network forming collagens that are between 160 and 230 residues in length. The multiplexin trimerization domains appear to be stabilized primarily by their hydrophobic core. In contrast, multiple methods of trimer stabilization are found in the network forming collagens. The crystal structures of type IV collagen are stabilized by intersubunit exchange of β-sheets in addition to extensive hydrophobic and hydrophilic interactions. Solvent molecules and polar residues are found in the trimeric hydrophobic core of collagens VIII and X. Furthermore, the type X trimerization core is further strengthened by a buried cluster of calcium ions.

More significantly, the NC1 domains of network forming collagens form supramolecular assemblies via trimer oligomerization. Collagens VIII and X NC1 structures display strips of solvent exposed aromatic residues on the trimer surface that are believed to facilitate supramolecular assembly. The type IV NC1 domain derived from human placenta forms a stable hexamer with a novel covalent crosslink between a Met and Lys residue on opposite trimers (91). The multiplexin trimerization domains are, in contrast, discrete entities that have no known additional function outside of chain association.

Multiplexins are present in all human tissues examined to date, but are present in very low concentrations (135). Collagen XV and XVIII trimers are capable of forming at picomolar concentrations, and are thus able to successfully associate even at low cellular concentrations. Additionally, the small size of the trimerization domain and the lack of disulphide bonds make trimerization highly efficient.

The multiplexin trimerization domains share relatively low sequence identity of 32%. In contrast, the trimerization domains of collagens VIII and X which have similar

structures but different mechanisms of trimer stabilization share over 60% sequence identity (94). The endostatin and endostatin-like domains of type XV and XVIII collagens also share over 60% identity (119), significantly more than the trimerization domains themselves. Analysis of the trimerization domain of multiplexins from a variety of organisms reveals little sequence homology, with the highest levels of homology involving key residues now known to be in the monomer or trimer hydrophobic core (177). *In silico* analysis of multiplexin trimerization domains predicted that the domain would contain a coiled-coil sequence (90). The crystal structures of both type XV(TD) and XVIII(TD) reveal no such structures, indicating that more accurate details of the multiplexin trimerization domain fold can be identified by structural rather than sequence analysis.

The structures of type XV and XVIII trimerization domains are highly similar, with 1.5 Å RMSD between Cα atoms along the entire backbone of the trimerization domain. The overall topology and organization of both the monomer and trimeric hydrophobic cores are comparable. Key residues in hydrophobic core formation are located in identical positions, although the sequence identity among these selected residues is only 33%. Thus, the relative location of hydrophobic residues plays a more important role in formation of a multiplexin trimerization domain than the absolute sequence identity. This further demonstrates the importance of empirically derived structural analysis, as traditional sequence analysis did not predict the structural similarity observed among these multiplexin trimerization domains

A homologue to multiplexin collagen has been identified in both *C. elegans* and *Drosophila* (159; 187). Types XV/XVIII and IV collagens are the only known collagens

that are conserved from nematode to human, which suggests that multiplexins may play an evolutionarily important role. The NC1 domain of the *C. elegans* multiplexin homologue was shown to trimerize in vitro (159); the multiplexin fold may therefore represent an ancient and well conserved folding motif.

It would be of interest to see how accommodating the multiplexin fold is, specifically through mutational analysis of residues key to either hydrophobic core. It is of note that the multiplexin trimerization domain fold can form trimers at extremely low concentrations, but is unfolded by relatively low concentrations of GuHCl (~2 M can unfold half of the total population). The hydrophobic core may be optimized for fast kinetics, but modifications to the core may also be able to enhance its chemical stability.

The first three residues of the XV(TD) protein were not visible, indicating that they are flexible and thus not resolvable by X-ray crystallography. In contrast, the N-terminus of the XVIII(TD) protein has ordered structure through the analogous residues. These residues were laid out across and stabilized by interactions with the trimer surface such that their ends are more than 20 Å apart. It is unlikely that this conformation would be biologically relevant, since a triple helical domain is immediately adjacent and must place the three chains in a close orientation. The positions of these residues may result from an artifact in the crystallization of type XVIII. The XV(TD) structure here, in contrast, indicates that the first four residues are able to move freely and can thus come in close enough contact to align triple helix. Since the multiplexins are homotrimers, chain selection is not a vital issue. However, the N-terminus may still play a role in chain staggering. The exact nature of the triple helix and trimerization domain interface remains to be determined.

The NC1 domain of both type XV and XVIII collagens contains a trimerization domain, a hinge region and an endostatin (type XVIII) or endostatin-like (XV) C-terminus. NC1 fragments have been isolated from a wide variety of basement membranes (117; 119; 160). The proteins are processed by proteases in the hinge region, releasing the endostatin or endostatin-like domain. Collagen XVIII is readily processed by a variety of proteolytic pathways into endostatin-containing fragments, whereas the NC1 domain of type XV collagen is not substantially converted into smaller fragments (118). Intriguingly, trimeric type XVIII NC1 domain and monomeric endostatin can act antagonistically in certain cellular processes. Thus the trimerization domain of type XVIII collagen can regulate endostatin function.   The biological activities of the full length and processed endostatin-like forms of collagen XV have not been extensively studied. Bacterially expressed endostatin-like protein was capable of inhibiting endothelial cell migration but had no effect on endothelial cell proliferation (138).   Solid state binding assays for full length NC1 and endostatin-like fragments of type XV collagen indicate that both proteins bind to a selection of extracellular matrix proteins with similar affinity.   Sequestration of endostatin-like fragments into obligatory trimers may thus not be an essential function of the type XV trimerization domain (119). However, the regulation of soluble NC1 domains of both type XV and XVIII collagens is likely complex, and the interplay between the trimerization domain and the endostatin-like domain of type XV collagen may play a more important regulatory role in complex cellular systems.

In conclusion, the type XV trimerization domain is remarkably similar to the type XVIII trimerization domain, despite low sequence identity. Its regulatory role in

multiplexin NC1 domain processing remains to be determined, but as a structural fold it represents a new class of small, efficient trimerization domains capable of folding at very low concentrations. The multiplexin trimerization fold may be used in future studies as a tool to create or increase the amount of correctly oligomerized recombinant trimeric proteins.

Footnotes

Table 1

Crystallographic Statistics

Data Collection	
Wavlength (Å)	1.54
Resolution range (Å)	52.5-2.0
Space group	$P6_3$
Cell dimensions	$a = b = 49.12$
	$c = 65.6$
R_{merge} (%)	3.4
$I/\sigma I$	35.4
Completeness (%)	100
Multiplicity	6.6
Wilson plot B	20.8
Matthews Coefficient	1.74
Contents of asymmetric unit	Two monomers
Solvent content (%)	29.5

Refinement	
No. of reflections used in refinement	6079
No. of reflections used for R_{free}	343
R-factor/R_{free} (%)	21.6/26.5
RMSD bonds/angles	0.008/1.4

Table 2

Mass Spectrometry and Sedimentation Equilibrium Data

	Calculated MW (Da)	Mass Spectrometry (Da)	$<v>$ (cm^3g^{-1})	Analytical Ultracentrifuge (kDa)	Oligomeric State
XV(TD)	6552	6552	0.74823	18.7 ± 0.1	2.85
XVIII(TD)	6434	6434	0.73874	19.2 ± 0.1	2.98

Samples were run in phosphate-buffered saline at 20ºC. XV(TD) and XVIII(TD) samples were run at a concentration of 0.375 and 0.235 mg/mL, respectively. $<v>$ were calculated using UltrascanII (182).

```
                1         10        20        30        40        50
   XV(TD)  --GSRNLVTAFSNMDDMLQKAHLVIEGTFIYLRDSTEFFIRVRDGWKKLQLGELIPIPA--
 XVIII(TD)  --GSSG-VRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPR--
   Trimer         * /                    * *       * * *      * *
   Monomer         ●  ●   ●  ●●  ●   ●    ● ● ●       ● ●    ●
```

Figure 1.   Sequence alignments**.** Alignment of type XV and XVIII trimerization domain
sequences used to solve their respective crystal structures. Sequences are derived from
GenBank[TM] accession numbers P39059 and P39060 for human type XV and XVIII
collagens, respectively. Residues common to formation of the trimeric hydrophobic core
are indicated with an asterisk (*) and the residue involved in trimeric hydrophobic core of
type XVIII collagen but not in type XV is indicated with a slash (/). Residues common to
the formation of the monomeric hydrophobic core are marked with dots (●). Italics
indicate residues not visible in the crystal structure of type XV. Underlining represents
non-natural sequence introduced during cloning in type XVIII.   Italics indicate residues
lost with tryptic digestion. The first visible residue in the type XV structure (N) is
labeled as residue 1.
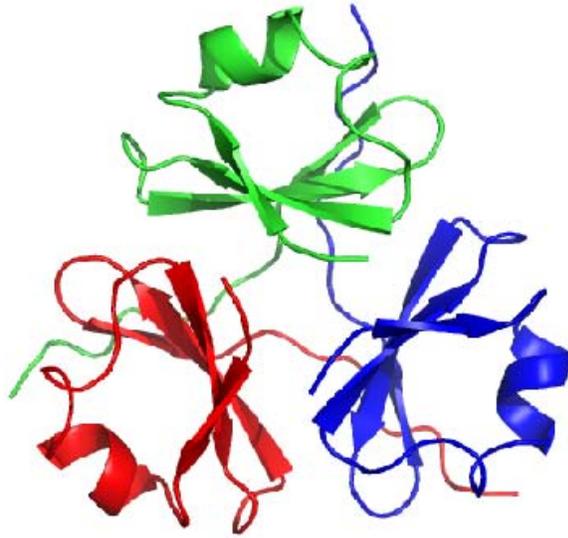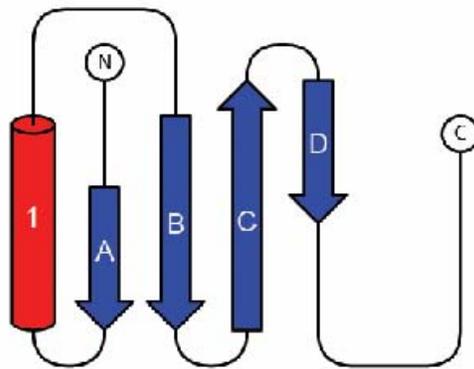
Figure 2. Structure of the type XV trimerization domain. (a) Cartoon representation of the trimeric biological unit. (b) Topology diagram for a single chain: α-helix shown as a red cylinder labeled 1 and beta sheets (labeled A through D) are shown as blue arrows.
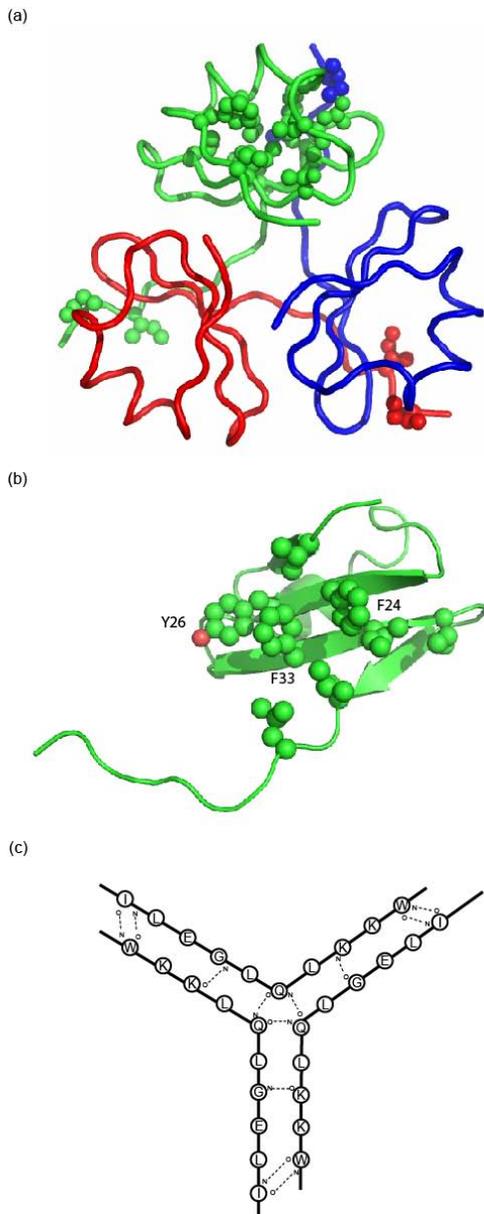
Figure 3. Hydrophobic cores and intersubunit hydrogen bonding. (a) Residues forming the hydrophobic core of a monomeric subunit are shown in green. Additionally, residues I50 and I52 which form hydrophobic contacts between subunits are shown for all three monomers. (b) Residues forming the hydrophobic interior of the trimer are shown for a monomeric subunit. (c) Main chain to main chain intersubunit hydrogen bonding diagram.

(a)
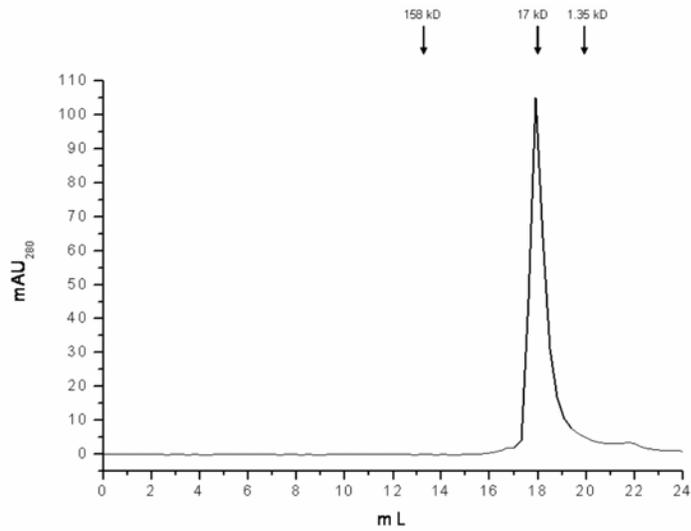
(b)
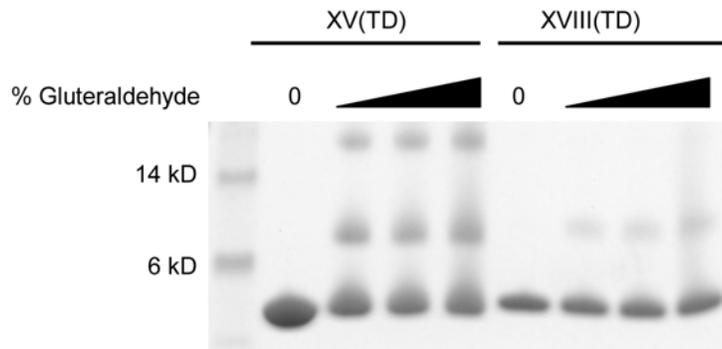
Figure 4. Gel filtration chromatography and chemical crosslinking. (a) XV(TD) protein were run on a Superose 6 column in PBS. Arrows represent elution peaks for the protein standards BSA, myoglobin and vitamin B_{12}. (b) Chemical crosslinking experiments were performed with zero, 0.0625, 0.125 and 0.25% glutaraldehyde for 5 minute with either XV(TD) or XVIII(TD) protein at a final concentration of 0.75 or 0.5 mg/mL. Samples were run on a 12% Bis-Tris gel.
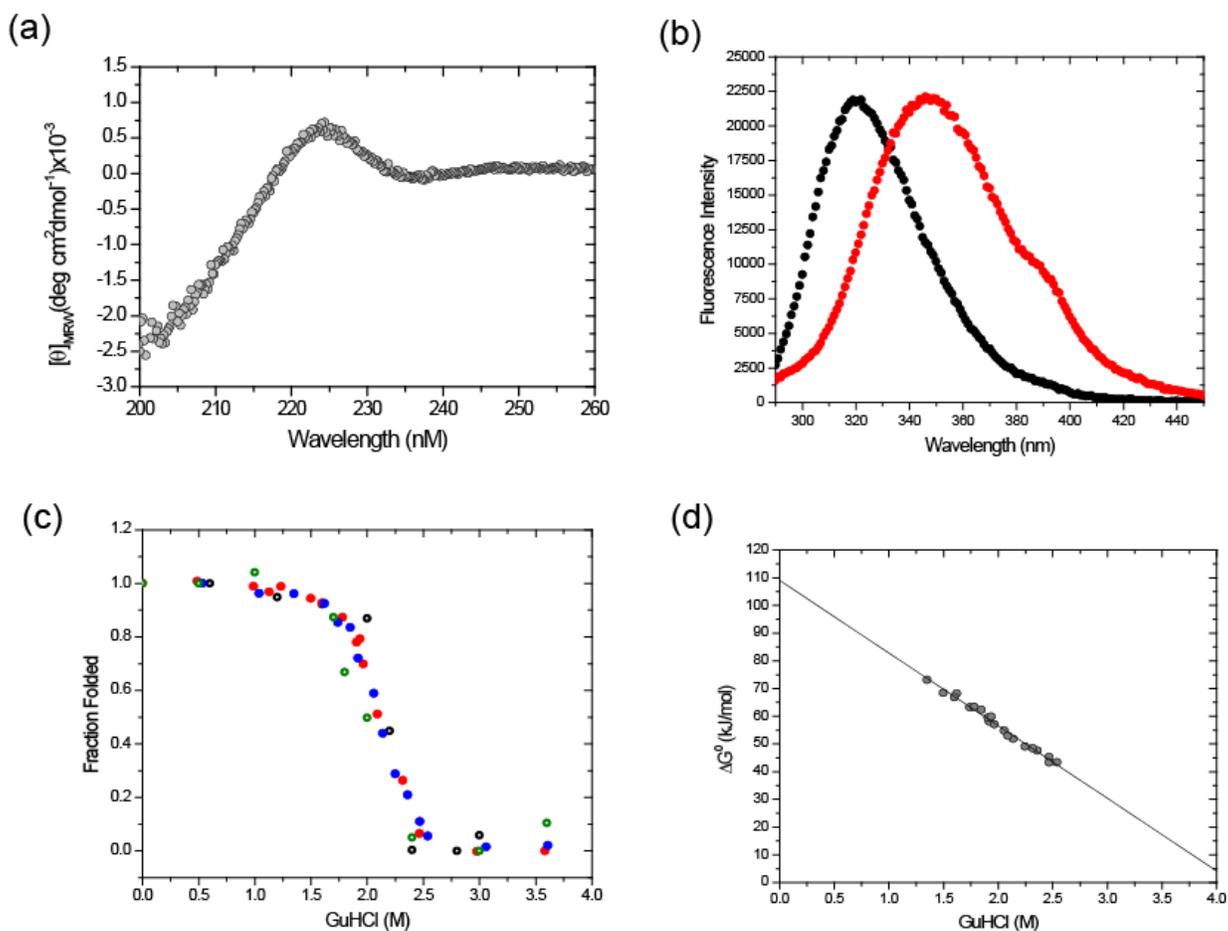
Figure 5. CD and fluorescence spectra and GuHCl-induced equilibrium transitions of the XV(TD). XV(TD was measured at 23.5 μM chain concentration in 50 mM sodium phosphate buffer pH 8.0. (a) CD spectrum of the native protein. (b) Fluorescence spectra of protein in 50 mM sodium phosphate buffer pH 8.0 with 0M GuHCl (●) or 4M GuHCl (●). (c) Fraction of folded protein as a function of GuHCl concentration. Unfolding by CD(○), refolding by CD (○), unfolding by fluorescence (●), refolding by fluorescence (●). (d) ΔG° dependence on GuHCl concentration with a linear fit.

CHAPTER 4

Biochemical Characterization of the

Prolyl 3-hydroxylase 1/CRTAP/Cyclophilin B Complex

Yoshihiro Ishikawa[‡¶], Jackie Wirz[§], Janice A. Vranka[‡], Kazuhiro Nagata[¶], and Hans Peter Bächinger[§‡]

[§]Department of Biochemistry and Molecular Biology, Oregon Health & Science University, and [‡]Shriners Hospital for Children, Research Department, Portland, Oregon 97239, USA; [¶] Department of Molecular and Cellular Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan

Explanation of Contribution

I purified P3H1, P3H1●CRTAP●CypB complex, Type I and Type III collagens used for some of these experiments. I also performed pilot fibril formation assays and pilot Biacore experiments. Additionally, I was involved with troubleshooting for all experiments in this manuscript.

Abstract

The rough endoplasmic reticulum resident protein complex consisting of prolyl 3-hydroxylase 1, CRTAP and cyclophilin B can be isolated from chick embryos on a gelatin Sepharose column, indicating some involvement in the biosynthesis of procollagens. Prolyl 3-hydroxylase 1 modifies a single proline residue in the α-chains of type I, II and III collagens to 3(S)-hydroxyproline. The peptidyl-prolyl *cis-trans* isomerase activity of cyclophilin B was previously shown to catalyze the rate of triple helix formation. Here we show that cyclophilin B in the complex shows peptidyl-prolyl cis-trans isomerase activity and that the P3H1●CRTAP●CypB complex has another important function: It acts as a chaperone molecule when tested with two classical chaperone assays. The P3H1●CRTAP●CypB complex inhibits the thermal aggregation of citrate synthase and is active in the denatured rhodanese refolding and aggregation assay. The chaperone activity of the complex is higher than that of protein disulfide isomerase (PDI), a well characterized chaperone. The P3H1●CRTAP●CypB complex also delays the in vitro fibril formation of type I collagen, indicating that this complex is also able to interact with triple helical collagen, and acts as a collagen chaperone.

Introduction

Procollagen biosynthesis occurs in the rough endoplasmic reticulum of cells and requires a large number of posttranslational modifications (175). While it has been known for a long time that type I collagen contains 3-hydroxyproline residues (188), the enzyme activity catalyzing this reaction was only partially characterized (48; 189; 50) until recently (47). Prolyl 3-hydroxylase 1 (P3H1) modifies a single proline residue in

the Xaa position of the Gly-Xaa-Hyp repeating sequence into 3(*S*)-hydroxyproline (3-Hyp) in the α1-chains of type I, II and III collagens. P3H1 extracted from chick embryos forms a multi-protein complex with CRTAP (previously described as Casp) and cyclophilin B (CypB) (47; 57). Laser light scattering and velocity sedimentation analysis shows that the three proteins form a 1:1:1 complex (57). P3H1 consists of two major domains: a carboxy-terminal dioxygenase domain, which is similar to the α-subunit of prolyl 4-hydroxylase (P4H) and lysyl hydroxylases, and a unique amino-terminal domain. The amino terminal-domain contains four cysteine (CXXXC) repeats and is homologous to CRTAP. This domain is found in five proteins of the human genome: P3H1, P3H2, P3H3, CRTAP and SC65 (47). The function of this domain has not been established but it also contains multiple TPR domains that are known to be important in protein-protein interactions (56; 190).

P3H1 was immuno-localized to cells in tissues that contain fibrillar collagens (47). Cells in tissues rich in basement membrane collagens did not stain with the monoclonal antibody against P3H1 and it is therefore likely that P3H2 or P3H3 3-hydroxylate proline residues in these collagens (65).

In order to understand the potential roles of the P3H1●CRTAP●CypB complex during collagen biosynthesis a brief review of the known steps is given: After the translocation of the growing polypeptide chains of procollagens into the rER, proline residues become 4-hydroxylated by prolyl 4-hydroxlase (P4H).   4-hydroxylation of proline residues increases the stability of the triple helix and is a key element in the folding of the triple helix.   P4H requires an unfolded chain as a substrate. The chain selection and association for triple helix formation is determined by the carboxy-terminal

propeptides in fibrillar collagens. Premature association between procollagen chains is thought to be prevented by chaperones such as PDI, BiP/GRP78, GRP94, HSP47 and FKBP65 and collagen modifying enzymes until the biosynthesis of the individual chain is completed. Additional modifications are the 3-hydroxylation of proline residues by the P3H1●CRTAP●CypB complex, the hydroxylation of lysine residues by lysyl hydroxylases and glycosylation. The chains are then selected and trimers are formed by association of the carboxy-terminal propeptides. Disulfide bonds between the chains are formed and this formation is most likely catalyzed by protein disulfide isomerase (PDI). Triple helix formation proceeds from the carboxy-terminal end towards the amino terminal end in a zipper-like fashion. The rate limiting step in this process is the cis-trans isomerization of peptide bonds. This process is catalyzed by peptidyl-prolyl cis-trans isomerase (CypB). Since procollagen molecules are only marginally stable, it was proposed that folding of procollagen molecules inside cells requires special chaperones (191), with HSP47 and FKBP65 as a potential candidates. Given the complexity of this process, it is not surprising that so many different proteins in the rER are involved and that these proteins interact and function as "folding machines". Additionally, collagen molecules have a strong tendency to aggregate. This process must be inhibited inside the cells and allowed only after secretion into the extracellular space. The functional importance of collagen chaperones is further shown by the embryonic lethality of HSP47 knock-out mice (192) and the accumulation of BiP/GRP78 in cells with mutations in the collagen chains (193).

The importance of the P3H1●CRTAP●CypB complex was recently demonstrated in CRTAP knock-out mice (57). These mice show osteochondrodysplasia characterized

by severe osteoporosis and decreased osteoid production. They lack the 3-Hyp in type I and II collagens. In addition, human mutations in P3H1 and CRTAP lead to a lethal form of recessive Osteogenesis Imperfecta (60; 61; 58). In this report we show that the P3H1●CRTAP●CypB complex is not only responsible for the 3-hydroxylation of proline residues, but is also a potent molecular chaperone.

Experimental Procedures

Extraction and purification of the chicken P3H1●CRTAP●CypB complex

The chicken P3H1●CRTAP●CypB complex was isolated from 15-17 day old chick embryos following the previously published protocol (41; 51) with modifications to improve both yield and purity of the complex. 12 dozen chick embryos were mixed with an equal volume of 10 mM Tris/HCl buffer, pH 7.5, containing 0.25 M sucrose and protease inhibitors (5 mM EDTA, 2 mM PMSF, 2 mM *N*-ethylmaleimide, 1μg/ml pepstatin A and 1μg/ml leupeptin). Homogenization was carried out in a Warring blender at maximum speed for 3 min. This and all subsequent steps were performed at 4°C. The homogenate was centrifuged at 3,000 x *g* for 15 min in a H-6000A rotor (Thermo Fisher Scientific Inc., Waltham, MA). The supernatant was then centrifuged at 125,000 x g for 1.5 h in a 45 Ti rotor (Beckman Coulter, Inc., Fullerton, CA). The pellets from this step were resuspended in the twice volume of 50 mM Tris/HCl buffer, pH 7.5, containing 0.1% Tween 20, 0.2 M NaCl and the same protease inhibitors as described above, treated with 1 μl/ml diisopropyl fluorophosphate and gently stirred for 4 h on ice. The extract was centrifuged at 125,000 x g for 1.5 h, filtered through Miracloth, and run over a gelatin-Sepharose 4B column (2.6 x 30 cm; GE Healthcare Bio-Sciences Corp.,

Piscataway, NJ) equilibrated in buffer A (50 mM Tris/HOAc buffer, pH 7.5, containing 0.2 M NaCl and 0.05% (v/v) Tween 20). The column was washed with at least two bed volumes of buffer A and then with one bed volume of 50 mM Tris/HOAc buffer, pH 7.5, containing 1 M NaCl and 0.05% Tween 20, followed by another bed volume of buffer A. Elution was performed using a pH gradient from 7.5 to 5.0 with buffer A. Peak fractions containing P3H1 complex and collagen related proteins were pooled, dialyzed into PBS at 4°C, and filtrated through a 0.45-μm filter prior to being loaded onto the P3H1 monoclonal antibody affinity column. The column was washed with at least five bed volume of PBS and then eluted by 50 mM glycine/HCl, pH 2.5, 150mM NaCl and 0.1% Triton X-100. P3H1 complex was pooled and dialyzed against each enzyme assay and reaction buffers.

Expression and purification of chicken CypB

The cDNA encoding chick CypB without the signal sequence, was amplified by PCR from chick fibroblast cDNA using primers with the 5' side containing a *Nco*I site and the 3' side containing a *Sal*I site after stop codon. This DNA was inserted between the *Nco*I and *Sal*I restriction sites of a pET30a(+) expression vector (Life Technologies Corp., Carlsbad, CA). The expression vector was transformed into *E. coli* BL21(DE3), grown at 37°C to an optical density of 0.6 at 600 nm, and IPTG to the (final concentration of 1 mM) to induce expression of CypB. After incubation at 30°C overnight, cells were harvested by centrifugation and resuspended in B-PER (Thermo Fisher Scientific Inc., Waltham, MA). Insoluble material was removed by centrifugation and ammonium sulfate (final concentration of 30% (w/v)) was added to the supernatant,

incubated at room temperature for 1 h and then centrifuged to remove the precipitated materials. The supernatant from the ammonium sulfate precipitation was passed through a 0.22 μm filter, applied onto a Co^{2+}-chelating resin column and CypB was eluted by 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl and 0.5 M imidazole after washing with 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl and 0.05 M imidazole for at least 10 column volumes. The fractions containing CypB were dialyzed into enterokinase cleavage buffer (50 mM Tris/HCl, pH 8.0, containing 1 mM $CaCl_2$ and 0.1% Tween 20), and enterokinase (0.1 unit/200 μl reaction volume) (Life Technologies Corp., Carlsbad, CA) was added at 4 °C overnight to cleave the His tag. The reaction mixture was incubated with DEAE sepharose (100 μl/1.5 ml reaction volume) to remove enterokinase at room temperature for 1 h. The reaction mix was spun down and the supernatant applied onto a Co^{2+}-chelating resin column and CypB was eluted by 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl. The purified CypB was dialyzed against each enzyme assay and reaction buffer.

Citrate synthase thermal aggregation assay

The aggregation of citrate synthase upon thermal denaturation(194) was measured by the method of Shao *et al*. (195). Citrate synthase (Sigma-Aldrich, St. Louis, MO) was diluted 200-fold to final concentration 0.15 μM into prewarmed 40 mM Hepes buffer (pH 7.4) at 43ºC. The aggregation of citrate synthase was monitored by absorbance at 500 nm in a Cary4 spectrophotometer (Varian Inc., Palo Alto, CA). All enzyme concentrations were determined by amino acid analysis. A stock solution of 0.5 mM Cyclosporine A (EMD Chemicals, Inc. Gibbstown, NJ) was prepared in DMSO and further diluted to

final concentration 1 μM (0.2% DMSO). Before measurement the P3H1●CRTAP●CypB complex was incubated with Cyclosporine A for 1 h at 4 ºC.

Denatured rhodanese refolding and aggregation assay

Another frequently used assay for chaperone activity is the inhibition of aggregation of chemically denatured rhodanese (196; 197). Bovine rhodanese (Sigma-Aldrich, St. Louis, MO) was denatured in 30 mM Tris/HCl buffer, pH 7.4, containing 6 M guanidine hydrochloride and 1 mM DTT at 25 ^{o}C for 1 h, then diluted 100-fold to a final concentration of 0.3 μM in 30 mM Tris/HCl buffer, pH 7.2, containing 50 mM KCl. The aggregation of denatured rhodanese was monitored at 320 nm with a Cary4 spectrophotometer. All protein concentrations were determined by amino acid analysis.

Enzyme assays and inhibition with cyclosporine A

Measurements of the catalytic efficiency (k_{cat}/K_m) for the isomerization reaction were performed as described (37), based on the α-chymotrypsin assay (198) with the following modifications: stock solutions of substrates were prepared in DMSO at a concentration of 2.5 mM for suc-Ala-Ala-Pro-Phe-MCA. The final DMSO concentration in the assay was 0.352% for k_{cat}/K_m measurements and 0.602% for the inhibition studies. Kinetic measurements were made at 5 °C to minimize the non-enzymatic isomerization reaction in 35 mM Hepes/NaOH buffer, pH 7.8. Final substrate and chymotrypsin concentrations were 8.8 μM and 12.8 μM, respectively. Fluorescence changes were monitored at 380 nm with an HiTech stopped-flow spectrophotometer (TgK Scientific Limited, Bradford-on-Avon, UK). The assay was started by mixture of chymotrypsin

and substrate.   Progression curves were analyzed by fitting to a second-order

exponential-decay function with Origin (OriginLab Corp., Northampton, MA).

Inhibition measurements were performed in a similar manner. A stock solution of

Cyclosporine A (0.1 mM) was prepared in DMSO and further diluted with 35 mM

Hepes/NaOH buffer, pH 7.8. After preincubation of the inhibitor with the P3H1 complex

at 4°C for 1 h, the assay was started. Values for k_{cat}/K_m were calculated according to

$k_{cat}/K_m = (k_{obs} - k_u)$ / [E], where k_u is rate constant for the unanalyzed isomerization

reaction and k_{obs} is the rate constant for the catalyzed reaction in the presence of enzymes

at a given concentration of [E]. k values were calculated using Origin.

Thermal stability of type I and type III collagen and refolding of type III collagen

measured by optical rotatory dispersion

The thermal stability of bovine type I and III collagen was monitored at 365 nm

using a 341MC polarimeter (Perkin-Elmer, Waltham, MA) with a 10 cm path-length

thermostatted cell. The temperature was controlled by a circulating water bath and

programmable temperature controller (RCS, Lauda Division, Brinkmann Instruments,

Inc., Westbury, NY) and measured with a digital thermometer (Omega Engineering, Inc.,

Stamford, CT) and a thermistor inserted into the cell. Both the temperature and the

optical rotatory dispersion signals were recorded and digitized on an HP9070A

measurement and plotting system (Hewlett-Packard, Palo Alto, CA). The temperature

was increased from 25°C to 50°C at a rate of 10°C/h. Stock solutions of type I and type

III collagen (final concentration 0.025 μM) in 10 mM acetic acid were diluted into 50

mM Tris/HCl buffer, pH 7.5, containing 0.4 M NaCl. Refolding of pN type III collagen

(0.125 μM) in 50 mM Tris/HCl buffer, pH 7.5, containing 0.2 M NaCl, was monitored at

365 nm. The sample was denatured for 10 min at 45 $^{\circ}$C and refolded at 25°C by

switching between two water baths.

Type I collagen fibril formation assay

A stock solution of type I collagen in 50 mM acetic acid was diluted to a final

concentration of 0.3 μM in 150 mM sodium phosphate buffer, pH 7.8, containing 150

mM NaCl. The absorbance (light scattering) was monitored at 313 nm as a function of

time at 34°C.   All protein concentrations were determined by amino acid analysis.

Surface plasmon resonance analysis

Surface plasmon resonance experiments were carried out using a BIACore X

instrument (GE Healthcare Bio-Science Corp., Piscataweay, NJ). Purified native bovine

type I collagen was immobilized at a concentration of about 6 ng/mm^2 (6000 RU) on a

CM5 sensor chip by amide coupling. The experiments were carried out at 25°C in 20

mM Hepes buffer, pH 7.4, containing 150 mM HaCl, 5 mM EDTA and 0.05% Tween 20.

Various concentrations of P3H1●CRTAP●CypB complex (from 0.35 µM to 0.075 µM)

were injected at a flow rate of 10 µl/min.

Results

Purification of the P3H1●CRTAP●CypB complex and expression of chicken CypB

The P3H1●CRTAP●CypB complex was isolated from chick embryos and

purified on a gelatin sepharose column followed by an antibody column loaded with a

monoclonal antibody against chicken P3H1. Figure 1 shows a SDS-polyacrylamide gel of the elution of the gelatin sepharose column (Fig. 1A) and the elution from the antibody column (Fig. 1B). To compare the biochemical activities of the P3H1●CRTAP●CypB complex and free CypB, we recombinantly expressed chicken CypB in *E. coli*. A SDS polyacrylamide gel of the purified chicken CypB is shown in Figure 1C. These protein preparations were used for the characterization of the biochemical activities.

Chaperone activity of the P3H1●CRTAP●CypB complex using the thermal aggregation of citrate synthase

The P3H1●CRTAP●CypB complex exhibits potent chaperone activity in this assay. Figure 2A shows the concentration dependent inhibition of the thermal aggregation of citrate synthase. Compared to BSA as a negative control and 0.15 μM PDI, an established chaperone, the P3H1●CRTAP●CypB complex shows much stronger aggregation inhibition activity at only 0.01 μM concentration (Fig. 2B). Chicken CypB or cyclosporine A alone do not show any chaperone activity and the addition of cyclosporine A to the P3H1●CRTAP●CypB complex does not alter its chaperone activity (Fig. 2C).

Chaperone activity of the P3H1●CRTAP●CypB complex using the aggregation and refolding of chemically denatured rhodanese

The concentration dependent inhibition of the aggregation and refolding of chemically denatured rhodanese by the P3H1●CRTAP●CypB complex is shown in Figure 3A. The P3H1●CRTAP●CypB complex at a concentration of 0.08 μM shows a

much higher chaperone activity than PDI at a concentration of 2 μM (Fig. 3B). Free

CypB does not contribute to the chaperone effect in this assay as shown previously (199).

PPIase activity of the P3H1●CRTAP●CypB complex and inhibition by cyclosporine A

The P3H1●CRTAP●CypB complex is active as a PPIase when analyzed with a

synthetic peptide substrate. Figure 4A shows the PPIase activity using suc-Ala-Ala-Pro-

Phe-MCA as a substrate. The catalytic efficiency k_{cat}/K_m 18,800 ± 1000 mM^{-1}s^{-1} of the

complex is a little higher than that of chicken CypB 11,600 ± 1000 mM^{-1}s^{-1} (data not

shown). The PPIase activity of the P3H1●CRTAP●CypB complex can be inhibited by

cyclosporine A in a similar manner as the inhibition of CypB (Fig. 4B). This indicates

that CypB in this complex is fully functional.

The refolding of type III collagen is rate-limited by cis-trans isomerizations of

proline residues and this rate can be catalyzed by CypB. Figure 5 shows the refolding of

thermally denatured pN type III collagen. The P3H1●CRTAP●CypB complex increases

the rate of refolding of pN type III collagen; however an equimolar amount of free CypB

is a better catalyst in this reaction, despite the higher catalytic efficiency of the complex

against peptide substrates.

The P3H1●CRTAP●CypB complex does not stabilize the collagen triple helix

In contrast to the recently described FKBP65, the P3H1●CRTAP●CypB complex

or free CypB does not stabilize the collagen triple helix. Figure 6 shows the melting

curves for type I collagen (Fig. 6A) and type III collagen (Fig. 6B). In both cases the

curves are very close in the presence or absence of a four molar excess of

P3H1●CRTAP●CypB or CypB.

Inhibition of collagen fibril formation by the P3H1●CRTAP●CypB complex

The P3H1●CRTAP●CypB complex also interacts with folded triple helical

collagen as shown by its activity in inhibiting the formation of type I collagen fibrils.

Figure 7 shows the concentration dependent inhibition of type I collagen fibril formation

by the P3H1●CRTAP●CypB complex and compares this activity with the known

inhibitor HSP47. When equimolar amounts of HSP47 and P3H1●CRTAP●CypB are

used Hsp47 shows a lower activity in inhibiting fibril formation. Free CypB does not

show any activity in this assay (data not shown).

Quantitation of binding of the P3H1●CRTAP●CypB complex to native type I collagen

The dissociation constant of the type I collagen P3H1●CRTAP●CypB complex

was determined to be 7.4 µM (Table 1). This value indicates a weaker interaction as

compared to HSP47, which was used as a control ($K_d = 2.3$ µM). The difference in K_d is

the result of a 5 to 10 fold faster dissociation rate of the complex compared to HSP47.

Discussion

P3H1 catalyzes the hydroxylation of the Pro 986 to 3(S)-Hyp of the α1-chains of

type I collagen in vitro (47). However, when P3H1 is extracted from chick embryos,

P3H1 is present in a complex with CRTAP and cyclophilin B. For the in vivo

hydroxylation of the single Pro residue in the α1-chains of type I, II and III collagens the

P3H1●CRTAP●CypB complex seems to be required. The CRTAP knock-out mice and human mutations in CRTAP and P3H1 show a dramatic decrease in the level of 3-hydroxylation (47; 60; 61; 58). Assuming that the hydroxylase activity is in the carboxy-terminal dioxygenase domain of P3H1, why are CRTAP and CypB required for the 3-hydroxylation of proline residues in vivo?

The results shown here indicate that the P3H1●CRTAP●CypB complex is not only responsible for the 3-hydroxylation of specific proline residues in the α-chains of fibrillar collagens, but also acts as a potent molecular chaperone. The complex is active in the inhibition of the thermal aggregation of citrate synthase and in the refolding and aggregation assay of chemically denatured rhodanese. Free CypB has no activity in these assays and inhibition of the peptidyl-prolyl cis-trans isomerase activity of CypB in the complex does not change this activity. The CypB in the complex is active as a PPIase and this activity can be inhibited with cyclosporine A. The catalytic efficiency of CypB in the complex is higher than free CypB against a peptide substrate. The catalytic efficiency of the complex was lower against pN type III collagen, but a slightly higher amount of collagen refolded compared to control or free CypB. The chaperone activity of the complex might help in the recognition of substrates and therefore the complex combines chaperone activity and PPIase activity during the refolding of collagen triple helices. It was recently shown that the insertion of a chaperone domain into FKBP12 (a PPIase) increased the catalytic efficiency of protein refolding (200).

It has been suggested that the chaperones HSP47 and FKBP65 stabilize the collagen triple helix in the rER (191; 41). The P3H1●CRTAP●CypB complex has no effect on the stability of type III collagen. A slight decrease in the stability of type I

collagen can be observed at a four fold molar excess of complex. This indicates that HSP47 and FKBP65 are the likely chaperones which stabilize the triple helix of these collagens in the rER. The P3H1●CRTAP●CypB complex also interacts with folded triple helical collagens, as indicated by the inhibition of fibril formation of type I collagen. The complex inhibits the fibril formation of type I collagen as efficiently as HSP47 and FKBP65 (58). It therefore contributes in the prevention of premature aggregation of collagen chains and molecules in the rER.

The P3H1●CRTAP●CypB complex binds folded type I collagen less efficiently than HSP47. The association rate constants are similar, but the complex shows a higher dissociation rate constant. Binding to unfolded collagen chains is likely to be very complex. The complex will bind to unfolded collagen chains for the 3-hydroxylation activity, the cis-trans isomerase activity and likely for the chaperone activity. One can speculate that the complex preferentially binds to unfolded chains and catalyzes the folding of the triple helix. The complex can then interact at this site with the triple helical structure and temporarily stabilize this junction. The fast dissociation is then used to replace the complex with HSP47 or other helix stabilizing chaperones.

The P3H1●CRTAP●CypB complex therefore has three distinct activities: it is a prolyl 3-hydroxylase, a PPIase and a molecular chaperone. The observed phenotypes in the CRTAP null mice and human mutations in CRTAP and P3H1 may not only result from the lack of the 3-Hyp residues in the type I, II or III collagens but also from a disturbance in the additional function of the P3H1●CRTAP●CypB complex as a chaperone and PPIase. CRTAP null mice show over-modified α-chains on SDS polyacrylamide gels. Over-modified α-chains in collagens of fibroblasts of patients with

CRTAP and P3H1 mutations were also observed and in addition a slower rate of secretion of type I collagen was found in these cells. This indicates that the chaperone function of the P3H1●CRTAP●CypB complex could contribute significantly to the observed phenotypes.

Footnotes
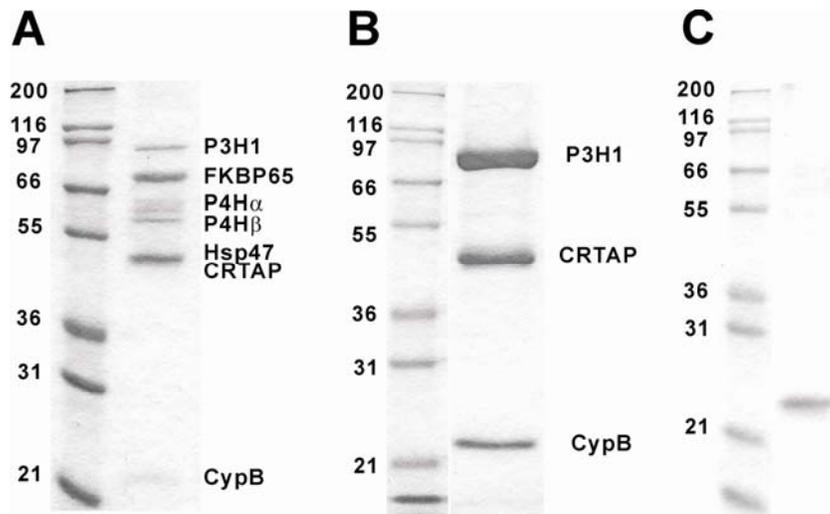
Figure 1. SDS-polyacrylamide gel electrophoresis of purified chicken P3H1●CRTAP●CypB complex and recombinant cyclophilin B. (A) Chicken rER proteins eluted from the gelatin sepharose column by low pH buffer. The proteins were identified by amino acid sequencing. This fraction was applied to a P3H1 antibody column and eluted with a pH 2.5 elution buffer. (B) The fractions containing the purified P3H1 complex were run on SDS-10% polyacrylamide gel under reducing conditions and stained with GelCode Blue Stain Reagent. (C) The purified chick CypB from *E. coli* after cleavage of the His-tag by enterokinase was run on a SDS-10% polyacrylamide gel under reducing conditions and stained with GelCode Blue Stain Reagent. The numbers on the left of each gel indicate the molecular masses of marker proteins in kDa.

Figure 2. Chaperone activity of P3H1●CRTAP●CypB complex using citrate synthase as a substrate. The inhibition of the thermal aggregation of citrate synthase by the P3H1●CRTAP●CypB complex was monitored at 500 nm. A 30 μM citrate synthase solution was diluted 200-fold into pre-warmed 40 mM Hepes buffer, pH 7.5, at 43 $^{\circ}$C. (A) in the absence (black) and presence of 0.001 μM (red), 0.0025 μM (green), 0.005 μM (blue) and 0.01 μM (orange) P3H1●CRTAP●CypB complex. (B) in the absence (black) and presence of 0.01 μM complex (red), 0.15 μM PDI (blue) and 0.15 μM bovine serum albumin (green). (C) in the absence (black) and presence of 1 μM cyclosporine A (red), with 0.01 μM CypB (blue), with 0.005 μM P3H1●CRTAP●CypB complex (green) and with 0.005 μM P3H1●CRTAP●CypB and 1 μM cyclosporine A (orange).
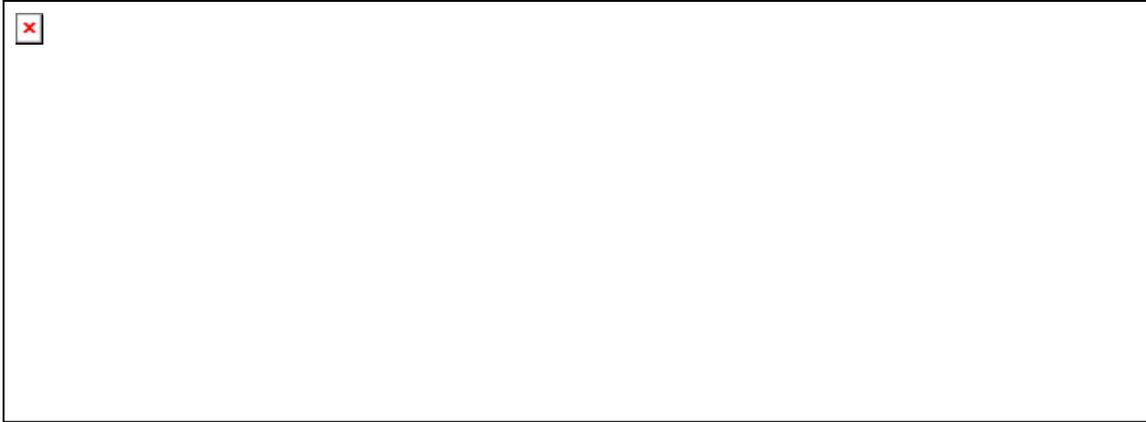
Figure 3. Influence of P3H1●CRTAP●CypB complex on the aggregation and refolding of chemically denatured rhodanese**. Chemically denatured rhodanese was diluted 100-fold (0.3 μM final concentration) into 30mM Tris/HCl buffer, pH 7.2, containing 50 mM KCl. Absorbance (light scattering) was monitored at 320 nm. (A) in the absence (black) and presence of 0.02 μM (red), 0.04 μM (blue) or 0.08 μM (green) P3H1●CRTAP●CypB complex. (B) in the absence (black) and presence of 0.08 μM P3H1●CRTAP●CypB complex (red). 2 μM PDI (green) and bovine serum albumin (blue) were positive and negative controls, respectively.
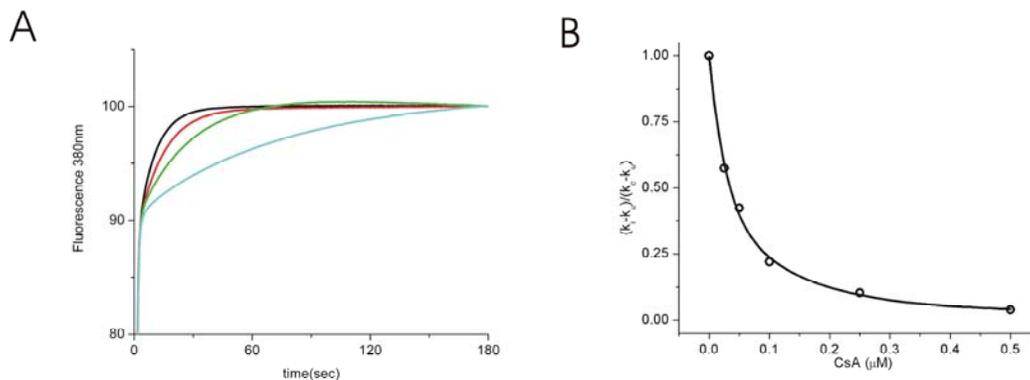
Figure 4. PPIase activity of P3H1●CRTAP●CypB complex and inhibition by cyclosporine A. The PPIase activity of the P3H1●CRTAP●CypB complex was measured in 35mM Hepes/NaOH, pH 7.8, at 5 °C using suc-Ala-Ala-Pro-Phe-MCA as the substrate. (A) The activity was monitored by the fluorescence at 380nm of free MCA that was cleaved by chymotrypsin. In the absence (blue) and presence of 0.0016 μM (green), 0.0032 μM (red) and 0.0064 μM (black) P3H1 complex respectively. (B) The inhibition of the PPIase activity of the P3H1●CRTAP●CypB complex by cyclosporine A was determined by the percentage of activity remaining upon increasing concentrations of cyclosporine A. The activity $(k_i-k_u)/(k_c-k_u)$ was calculated using the rate constant k_i in the presence of inhibitor, k_c uninhibited rate constant and k_u the rate constant of the uncatalyzed reaction. The final P3H1●CRTAP●CypB complex concentration was 0.013μM.
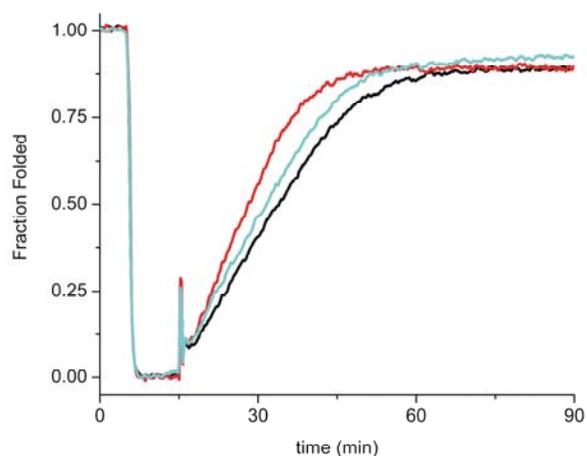
Figure 5. Refolding of pN type III collagen in the presence of P3H1●CRTAP●CypB complex. Refolding was monitored by optical rotatory dispersion at 365 nm. The signal from pN type III collagen was observed for 5 min at 25°C, and then native collagen was denatured for 10 min at 45°C to unfold the triple helix. Refolding was initiated by quickly lowering the temperature back to 25 °C. Refolding curves are shown in the absence (black) and presence of 0.05 μM P3H1●CRTAP●CypB complex (blue) or 0.05 μM chicken CypB (red). The increased slope of the linear refolding phase indicates catalysis of cis-trans isomerizations. The final concentration of pN type III collagen was 0.125 μM.
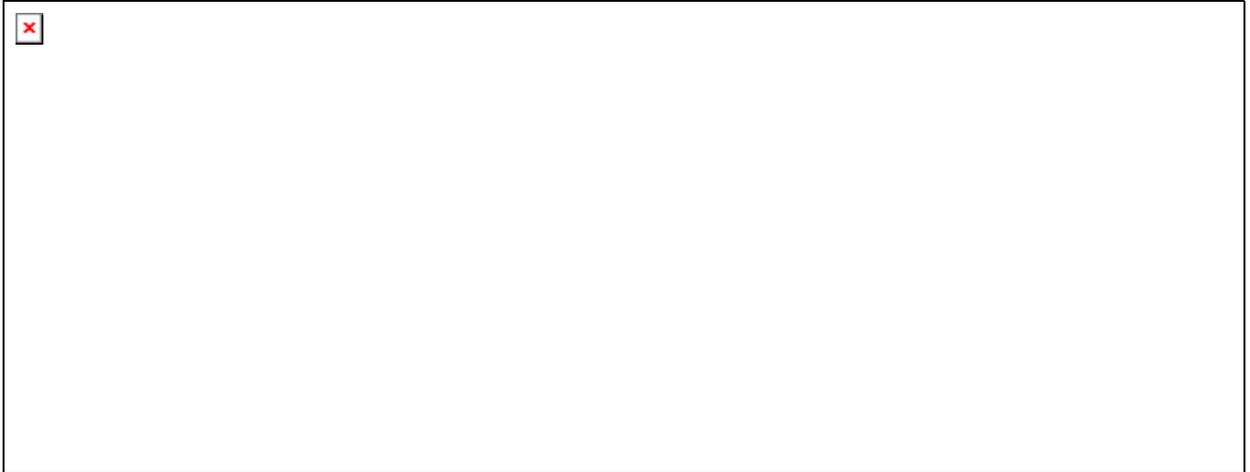
Figure 6. Melting curves of type I and III collagens in the presence of P3H1●CRTAP●CypB complex. Optical rotatory dispersion was measured at 365 nm and the temperature was increased at a rate of 10°C/h. A stock solution of type I collagen and type III collagen (final concentration 0.025 μM) in 50 mM acetic acid was diluted into 50 mM Tris/HCl buffer, pH 7.5, containing 0.4 M NaCl. The final concentration of both P3H1●CRTAP●CypB complex and CypB were 0.1 μM. Denaturation curves in the absence (black) and presence of P3H1●CRTAP●CypB complex (blue) or CypB (red) are shown for type I collagen (A) and type III collagen (B).

Figure 7. Inhibition of fibril formation of type I collagen in the presence of P3H1●CRTAP●CypB complex. A stock solution of type I collagen in 50 mM acetic acid was diluted into 150 mM sodium phosphate buffer, pH 7.8, containing 150mM NaCl and measured at 34°C. The absorbance (light scattering) was monitored at 313 nm. The final concentration of type I collagen was 0.3 μM. The following curves are shown: in the absence (black) and presence of 0.01 μM (red), 0.025 μM (green) or 0.05 μM (orange) P3H1●CRTAP●CypB complex and 0.05 μM Hsp47 (blue) as a positive control.

Table 1

	K_d (10^{-6} M)	k_a (1/Ms)	k_d (1/s)
P3H1●CRTAP●CypB	7.4 ± 2.8	$2.5 \pm 0.4 \cdot 10^4$	$2.0 \pm 0.5 \cdot 10^{-1}$
HSP47	2.3 ± 0.6	$2.8 \pm 0.5 \cdot 10^4$	$6.2 \pm 1.0 \cdot 10^{-2}$
HSP47[a]	1.1	$2.08 \cdot 10^4$	$2.36 \cdot 10^{-2}$

Dissociation and rate constants of binding of native bovine type I collagen to the P3H1●CRTAP●CypB complex and HSP47

[a] taken from reference (27)

CHAPTER 5

Studies on Prolyl 3-Hydroxylase and the

Prolyl 3-Hydroxylase•Cartilage Associated Protein•Cyclophilin B Complex:

Recombinant Protein Expression and Protein Preparation from Whole Chick

Extracts

Jacqueline Alani Wirz[§] and Hans Peter Bächinger[§‡]

[§]Department of Biochemistry and Molecular Biology, Oregon Health & Science

University

[‡]Shriners Hospital for Children, Research Department, Portland, Oregon 97239, USA;

Abstract

Prolyl 3-hydroxylase 1 (P3H1) modifies proline residues in the α-chains of

various collagens to 3-hydroxyproline. P3H1 was isolated from chick embryos on a

gelatin Sepharose column in complex with cartilage associated protein (CRTAP) and

cyclophilin B (CypB). In order to better understand the functions of these proteins,

attempts were made to express P3H1 and CRTAP proteins recombinantly and to enhance

yield of the complex from chick embryos.   Over expression systems were not successful,

as the protein was highly insoluble. Attempts to refold solubilized inclusion bodies were

also unsuccessful, as the protein either rapidly precipitated out of solution or was not

folded into a conformation recognized by a P3H1 antibody. Although some progress was

made enhancing the protein complex yield from chick extracts, preliminary data indicate

that the P3H1●CRTAP●CypB complex may be partially membrane-associated, further

complicating our approach to structural analysis. The complex was shown to also form

large molecular weight aggregates that contain a fragment of P3H1, the α and β subunits

of P4H and CRTAP. The relevance of these additional proteins is unknown, although

they may be involved in aggregate formation. Although the research presented here has

made progress towards generating high concentration protein, production of P3H proteins

for further structural analysis remains a challenge.

Introduction

Collagen undergoes a variety of post-translation modifications that are required

for proper biosynthesis, folding and assembly. 3-hydroxyproline (3Hyp) residues have

been observed in collagens since the early 70s (188). The modification is relatively rare,

occurring roughly 1-15 times per 1000 residues with the highest frequency in type IV collagen (13; 201; 202). Prolyl 3-hydroxylase (P3H) activity was partially purified from rat kidney cortex and initially characterized as a 2-oxoglutarate, ascorbate and Fe(II) iron dependant dehydrogenase (48; 189; 50). The enzyme activities of prolyl 4-hydroxylase (P4H) and P3H were shown to be separate (49). However, the protein responsible for P3H activity was not identified by either amino acid sequence or as a specific gene. In 2004, the chick homologue to a human leucine proline rich proteoglycan (leprecan) was identified as the protein responsible for P3H activity; the gene was subsequently renamed P3H1. Two other genes were identified as P3H1 homologous; myxoid liposarcoma associated protein 4 (*MLAT4*) and gene rich cluster, B (*GRCB*) genes correspond to P3H2 and P3H3, respectively. P3H1 is found in complex with the cartilage-associated protein (CRTAP) and cyclophilin B (CypB) (47). Additionally, the P3H1•CRTAP•CypB complex has been shown to exhibit strong chaperone activity (42).

The primary structure of P3H1 can be divided into two domains: a C-terminal dioxygenase domain and a unique N-terminal domain (Introduction, Figure 2). The dioxygenase domain is similar to both the α-subunit of prolyl 4-hydroxylases and the dioxygenase domain of lysyl hydroxylases, all of which share conserved catalytic residues. The N-terminal domain is unique to the other P3H family members, CRTAP and the synaptonemal protein SC65. This domain has unknown function and contains four cysteines repeats (CXXXC) and several multiple tetratricopeptide (TPR) domains that are known to be important for protein-protein interactions (203). P3H1 was shown to be highly expressed in embryonic tissues, while P3H2 was strongly expressed in kidney. P3H3 gene expression had overlapping expression with P3H1 and P3H2 (64).

CRTAP was also expressed in most tissues, with immunohistochemistry staining highest in cartilage (204).

The importance of the P3H1●CRTAP●CypB complex has been shown through a variety of studies. Mutations in human P3H1, CRTAP or CypB have been associated with severe or lethal forms of recessive osteogenesis imperfecta (OI), a connective-tissue disorder (57-62). OI has varying degrees of severity, and is characterized by bone fragility, low bone mass, short stature, bowing of the long bones and kyphoscoliosis (67; 68). Collagen from patients with mutations in any member of the complex displayed decreased levels of 3-hydroxylation. P3H2 and P3H3 were also down-regulated in certain cancers and may act as a tumor suppressor (75).

The growing body of literature on P3H1 and P3H1●CRTAP●CypB complex indicate that these proteins play an important role in collagen biosynthesis and connective-tissue diseases. Enzyme activities of P3H1 and P3H2 have been analyzed (47; 65) and the complex has been shown to have both peptidyl-prolyl cis-trans isomerase (PPIase) and chaperone activities (42): however, the underlying structural components that facilitate these activities are unknown. Structural studies were limited by the low yield of purified protein from chicken embryos, the source for most of the biochemical studies performed to date. To these ends, a variety of constructs were designed to recombinantly express P3H1, P3H2, P3H3 and CRTAP. Bacterial and insect cell expression systems did not yield soluble protein. Attempts to refold full length protein were partially successful, although the stability of refolded protein was low. In lieu of recombinant expression systems, attempts were made to maximize protein production

from chick embryos. Protein preps were enhanced by the addition of various detergents, indicating possible membrane-association.

Experimental Procedures

Cloning of P3H and CRTAP Constructs into Bacterial Expression Vectors

Chick P3H1, P3H2, P3H3 and mouse CRTAP were cloned into pET30a expression vectors (EMD Biosciences) from existing pCR4 vectors. Constructs were designed to carry a N-terminal His tag, an S-tag and a thrombin cleavage site followed by the sequence of interest and a stop codon. Full length P3H1 and P3H3 were excised from pCR4 and inserted into pET30a using an *EcoRI* restriction site. P3H2 was excised using *NcoI* and *Sma* restriction sites, and cloned into pET30a with *NcoI* and *EcoRV* restriction sites. mCRTAP was excised and inserted using *NcoI* and *EcoRV*.

Chick P3H1 was also cloned into the pET30a vector such that the full length sequence was followed by a C-terminal His tag. P3H1 was amplified using a forward primer containing an *Nde* restriction site (5'-GGAATTCCATATGGCGCTGCTGCTGCGC-3') and a reverse primer containing a *NotI* restriction site with no stop codon (5'-GCAGGGAAGGACGAACTGGCGGCCGCATTCTTAT-3'). Mouse CRTAP was also cloned into a pET30a vector, using a forward primer containing an *Nde* restriction site (5'-GGAATTCCATATGGGGCCCCGCAGCCC-3') and a reverse primer containing and *HindIII* restriction site with no stop codon (5'-CCCAAGCTTGGCAGACTCTTCCGTCTCC-3').

Mouse CRTAP was cloned into the pET15a vector such that a His tag and a thrombin cleavage site were immediately followed by the sequence of interest and a stop codon. A forward primer containing an *Nde* restriction site (see above), and a reverse primer containing a stop codon and a *BamHI* site were used (5'-CG<u>GGATCC</u>TCAGGCAGACTCTTCCGTCTCC-3'). Chick P3H1 was unable to be cloned into the pET15a vector, as the three restrictions sites available to insert the 3' end of the gene were all present in the cLEP sequence. Attempts to mutate the sequences to allow for use of XhoI and BamHI sites were unsuccessful.

All inserts were verified by Sanger dideoxy DNA sequencing.

Expression in E. Coli

Expression vectors were transformed into a variety of *E. coli* expression cell lines. BL21(DE3), Tuner Cells, Origami, Origami2 Cells (all from EMD Biosciences), ArcticExpress and Arctic Express RP (Agilent) cells were all used. BL21(DE3), Tuner, Origami and Origami2 cells were grown at 37ºC to an optical density of 0.6 at 600 nm, and isopropyl 1-thio-β-D-galactopyranoside (IPTG) was added to induce expression with a final concentration varying from 0.1 to 1 mM. Induction was allowed to proceed at 37, 30 and 20ºC for 3, 6 or 12 hours, respectively. Arctic Express and Arctic Express RP cells were grown at 37°C to an optical density of 0.8 at 600 nm, after which the cells were cooled to 15 °C. After equilibration, cells were induced with 1 mM IPTG for 24 hours. Cells were harvested by centrifugation. Pellets were frozen at -80ºC.

Protein was initially purified using a HiTrap Chelating column (GE Healthcare), which was charged with Ni^{2+} ions as per manufacturer's instructions. Purifications were

run in a variety of buffer systems, with differing buffering ions, pH, denaturants and detergent concentrations.

Inclusion bodies were extracted using a variety of buffers containing guanidine hydrochloride (GuHCl) or Urea. Initial buffers were made of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.1% TX-100 and either 0, 2, 4, 6 or 8M Urea or 0, 2, 4 or 6M GuHCl. Detergent screens were performed with a variety of detergents (Table 1) in 50 mM Tris-HCl pH 8.0 and 500 mM NaCl.

Soluble chick CypB was purified as previously described (42). The pET30a(CypB) expression vector was transformed into BL21(DE3) cells, grown at 37ºC to an optical density of 0.6 at 600 nm. IPTG was added to a final concentration of 1 mM to induce expression of CypB. After incubation at 30ºC overnight, cells were harvested by centrifugation and resuspended in B-PER (Thermo Fisher Scientific Inc). Insoluble material was removed by centrifugation after ammonium sulfate was added to a final concentration of 30% (w/v). The supernatant from the ammonium sulfate precipitation was passed through a 0.22 μm filter, applied onto a Ni^{2+} chelating column (GE Healthcare) and CypB was eluted with 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl and 0.5 M imidazole after washing with 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl and 0.05 M imidazole. The fractions containing CypB were cleaved by enterokinase (Life Technologies Corp.) following manufacturer's instructions. After cleavage, the reaction mixture was incubated at room temperature for 1 h with DEAE sepharose to remove enterokinase. The reaction mix was spun down and the supernatant applied onto a Ni^{2+} chelating column and CypB was

eluted with 20 mM sodium phosphate buffer, pH 7.5, containing 0.5 M NaCl. Purified

CypB was dialyzed against phosphate buffered saline (PBS).

Protein Refolding

Refolding experiments were performed using the QuickFold Protein Refolding

Kit (AthenaES) or iFOLD (EMD Biosciences) as per manufacturer's instructions with

minor modifications to maximize extraction of protein from inclusion bodies.

Briefly, the QuickFold kit uses solubilized inclusion bodies extracted with 50 mM

Tris-HCl pH 8.0, 500 mM NaCl and 0.4 mM Z3-14 diluted to 1 mg/mL. 50 µl of protein

solution was added to 950 µl of refolding buffer while vortexing the solution gently. The

samples were incubated at room temperature for 1 hour and then centrifuged at 14,000 x

g for 10 minutes. Soluble protein was analyzed by SDS-PAGE.

The iFOLD system purifies and washes inclusion bodies in a N-Lauroylsarcosine

buffer system. Frozen pellets were resuspended in 10 mL 1X IB Wash Buffer (50 mM

Tris-HCl pH 8.0, 50 mM NaCl, 1 mM TCEP, 0.5 mM EDTA, 5% glycerol and 1% Triton

X-100) per gram of wet cell paste. Lysozyme was added to 1 mg/mL final concentration,

and the solution was gently stirred at room temperature for 15 minutes. Cells were lysed

by sonication while on ice. Triton X-100 was then added to a final concentration of 1%

(v/v). The solution was stirred at room temperature for 15 minutes, then spun at 8000 x *g*

for 15 minutes at 4°C. The inclusion bodies were washed 3x in 1X IB wash buffer and

the resulting purified inclusion bodies were frozen at -80°C.

Inclusion bodies were denatured in 10 mL of 1X IB Denaturation Buffer (50 mM

Tris-HCl 8.0, 50 mM NaCl, 5 mM TCEP, 0.5 mM EDTA, 5% glycerol which was

supplemented with 0.4 mM Z3-14) per 0.5 g of purified inclusion body pellet. Pellets

were homogenized using a Dounce homogenizer then sonicated briefly on ice. 1.75 mL

of 30% *N*-Lauroylsarcosine was added per 0.5g inclusion body pellet. The solution was

stirred at room temperature until the solution became translucent (~15 minutes). The

solution was subsequently spun at 25,000 x *g* for 20 minutes at 4°C. The supernatant was

passed through a 0.45 µM filter and dialyzed against 10 mM Tris-HCl pH 8.0, 0.05 mM

EDTA, 0.1 mM TCEP and 0.06% (w/v) *N*-Lauroylsarcosine.

Solubilized inclusion bodies were assessed for concentration by Bradford assay.

The protein was then diluted to 1 mg/mL. 50 µl of protein solution was added to each

well of the refolding matrix and rapidly mixed by pipetting. The refolding reaction was

incubated while shaking at room temperature overnight. Folded protein was assessed by

measuring absorbance at 340 nm. Conditions with lowest absorbance at 340 nm were

chosen for further analysis.

For protein folding experiments performed with P3H1, CRTAP and CypB

combined, proteins were at 1 mg/mL. CypB was added as a soluble protein in PBS.

Adjusting protein concentrations to equimolar ratios did not alter the refolding results.

Cloning of P3H1 and CRTAP into Insect Cell Expression Vectors

Full length cP3H1 and mCRTAP were cloned into the pMIB/V5-HisA vector

(Invitrogen) such that the Honeybee melittin secretion signal and a melittin cleavage site

was immediately followed by the sequence of interest, followed by a V5 epitope and a

His tag. cP3H1 was cloned using a forward primer containing a *HindIII* site (5'-

CGC<u>AAGCTT</u>CATGGCGCTGCTG-3') and a reverse primer containing an *Xba* site (5'

CGCTCTAGACAGTTCGTCCTTCC-3’). mCRTAP was cloned using a *HindIII*

forward primer (5’-CGCAAGCTTCATGGGGCCCC-3’) and a *Xba* reverse primer (5’-

CCGTCTAGAGGCAGACTCTTCC-3’).

Expression of P3H1 and CRTAP in Insect Cells

Recombinant vectors were transfected into *Spodoptera frugiperda* (Sf9) insect

cells using calcium phosphate, Insect GeneJuice (EMD Biosciences) or Cellfectin

(Invitrogen) transfection methods following manufacturer’s instructions. Supernatant

was collected 3, 4 and 5 days post-transfection. Whole cell lysates were prepared by

suspension in cell lysis buffer (50 mM Tris, pH 7.8, 150 mM NaCl, 1% NP-40) after

washing the cells twice with 1X PBS.

Protein production was monitored by Coomassie staining of SDS-PAGE gels or

by western blot. Western blots were probed with V5 (Invitrogen), His-tag (GE

Healthcare) or P3H1 (IC10) antibodies.

Extraction and Purification of the P3H1●CRTAP●CypB complex from Chick Embryos

Chick P3H1●CRTAP●CypB complex was isolated from 15–17-day-old chick

embryos following the previously published protocol with modifications to improve both

yield and purity of the complex as described in the results section (42; 47). 16 dozen

chick embryos were mixed with an equal volume of 10 mM Tris/HCl buffer, pH 7.5,

containing 0.25 M sucrose, protease inhibitors (5mM EDTA, 2 mM

phenylmethylsulfonyl fluoride, 2 mM N-ethylmaleimide, 1μg/ml pepstatin A, and 1

μg/ml leupeptin), and 0.1% Tween 20. Homogenization was carried out in a Waring

blender at maximum speed for two cycles of two minutes of blending followed by a 5

minute rest period to prevent the sample from overheating. This and all subsequent steps

were performed at 4 °C. The homogenate was centrifuged at 3,000 x g for 15 min to

remove tissue debris. The supernatant was filtered through cheesecloth and then

centrifuged at 125,000 x g for 1.5 h in a 45 Ti rotor (Beckman Coulter). At this point

pellets may have been frozen for later use. Pellets from up to three preps were combined

for extraction and purification. Fresh or frozen pellets were resuspended in 50 mM

Tris/HCl buffer, pH 7.5, 0.2 M NaCl, the same protease inhibitors as described above and

0.1% Tween 20 at twice the pellet volume. Different detergents were also used as

outlined in the results section. Resuspended pellets were treated with 1 μl/ml diisopropyl

fluorophosphate, and gently stirred for a minimum of 4 h on ice. The extract was

centrifuged at 125,000 x g for 1.5 h, filtered through Miracloth, and run over a 200 mL

gelatin-Sepharose 4B column (GE Healthcare) equilibrated in buffer A

(50mMTris/HOAc buffer, pH 7.5, containing 0.2 M NaCl and 0.1% (v/v) Tween 20), or

other detergent as outlined in the results section. The column was washed with at least 5

column volumes of buffer A, then 1 column volume of high salt buffer B (50mM

Tris/HOAc buffer, pH 7.5, containing 1 M NaCl and 0.1% Tween 20), followed by re-

equilibration with 5 column volume of buffer A. Elution was performed using a pH

gradient from 7.5 to 5.0 with buffer A. Peak fractions containing P3H1complex and

collagen-related proteins were pooled and dialyzed into phosphate-buffered saline at 4°C.

Dialyzed fractions were passed through a 0.22 μm filter prior to being loaded onto the

P3H1 monoclonal antibody affinity column. The column was washed with at least 5

column volumes of phosphate-buffered saline and then eluted with 50 mM glycine/HCl,

pH 2.5, 150 mM NaCl, and 0.1% Triton X-100. Fractions were neutralized with

saturated Tris Base to prevent acid-induced degradation. Protein fractions were pooled

and dialyzed into PBS.

For protein stability studies, complex was dialyzed into PBS overnight, solubility

of complex that had been stored at 4ºC for 2 and 5 days was first checked. 100 μL of

protein was spun at 14,000 x g for 10 minutes at 4ºC. Immediately following the spin,

soluble protein was removed. The pellet was brought up in an equal volume of 1X SDS-

PAGE gel sample buffer. Aliquots of soluble and insoluble protein prepped in sample

buffer were run on a SDS-PAGE under non-reducing and reducing conditions. The gel

was stained with GelCode Blue.

For analysis of high molecular weight bands, several lanes of

P3H1●CRTAP●CypB complex were run under non-reducing conditions on a SDS-

PAGE gel. Bands of interest were excised with a razor blade, combined and minced into

pieces. The gel pieces were loaded onto a SDS-PAGE gel and run under reducing

conditions and stained with Coomassie blue. Bands of interest were transferred to PVDF

and analyzed by N-terminal sequencing.

For studies on detergent concentrations, 10g of crude microsomal pellets were

used for each condition and prepped as above, save variances in detergent concentration.

Extracted pellets were run over a gelatin Sepharose column as previously described.

Eluted gelatin Sepharose proteins were analyzed on SDS-PAGE gels.

Laser Light Scattering

P3H1●CRTAP●CypB molecular masses were observed using a multi-angle light scattering (MALS) instrument (DAWN Eos, Wyatt Technology). The instrument was placed in line with a UV detector, refractive index detector (RI, Wyatt), and a quasi-elastic light scattering detector (QELS, Wyatt Technology) during size-exclusion chromatography (SEC). SEC was performed with a Superose 12 10/300 GL column (GE Healthcare) equilibrated in phosphate-buffered saline with a flow rate of 0.5 ml/min. Signals from the UV, RI, and MALS detectors were normalized using bovine serum albumin. Monodisperse regions under the peak were analyzed. The weight-averaged molar masses (MW) were reported based on the average protein concentration for the peak area analyzed

Results

Cloning and Expression Studies

Chick P3H1, P3H2, P3H3 and mouse CRTAP constructs were cloned into pET30 vectors such that the constructs carried an N-terminal His-tag, an S-tag, a thrombin cleavage site and the sequence of interest. Test inductions were initially performed using BL21(DE3) cells. CRTAP was well expressed, whereas P3H1 was weakly expressed as analyzed by SDS-PAGE (Figure 1a). P3H2 and P3H3 exhibited almost non-existent expression levels as measured by both SDS-PAGE and anti-His western blots. Attempts to optimize expression for P3H2 and P3H3 were unsuccessful: these constructs will not be discussed further. Chick P3H1 and mouse CRTAP were also cloned into a pET30a vector such that the full length sequence was followed by a C-terminal His tag. Mouse

CRTAP was also cloned into a pET15a vector such that the construct had a N-terminal His tag followed by a thrombin cleavage site. Test inductions of pET30a(cP3H1) and pET15(mCRTAP) exhibited stronger expression of both proteins (Figure 1b). pET30a(mCRTAP) however, did not express significant amounts of protein. To optimize P3H1 induction, a variety of cell lines were tested: Bl21(DE3), Tuner Cells, Origami, Origami 2, Arctic Express (DE3) and Arctic Express (DE3)RP cell lines all produced similar results (Figure 2).

P3H1 was expressed primarily in inclusion bodies (IBs) (Figure 3). Attempts to purify soluble protein were unsuccessful. To harvest protein from the IBs, a variety of Urea and GuHCl concentrations were tested in extraction buffers (Figure 3). An extraction buffer comprised of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.1% TX-100 and 8M Urea exhibited the highest level of protein extraction. However, a significant portion of protein remained in the extracted pellet.

Detergents are often used to enhance protein solubility. A screen of detergents was therefore used to maximize IB extraction yield (Figure 4). Compounds were chosen to represent three major classes: non-ionic, ionic and zwitterionic detergents (Table 1). Purified IBs were extracted with 25 mM Tris-HCl pH 8.0, 250 mM NaCl and 8M Urea containing detergents at optimal critical micelle concentrations or in 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 6M GuHCl supplemented with detergents as above (Figure 5). The optimal condition for IB purification was found to be 25 mM Tris-HCl pH 8.0, 250 mM NaCl and 8M Urea containing 0.4 mM Z3-14. This buffer (solubilization buffer) was used for all subsequent IB extractions.

Extracted protein was further purified under denaturing conditions using a HiTrap Chelating column charged with Ni^{2+} ions (Figure 5a). Solubilization buffer containing 500 mM imidazole was used for elution. Purification was not absolute, as a single band was not achieved. Attempts to optimize denaturing protein purification by varying pH and salt concentration did not further improve the quality of the protein. It is of interest that a significant portion of the unfolded protein did not bind successfully to the chelating column. This was not an artifact of overloading the column, as lower concentrations of protein yielded similar ratios of bound/weakly bound/unbound protein populations.

mCRTAP was also expressed primarily in inclusion bodies and behaved in a manner similar to P3H1. The optimal extraction buffer is the same as for P3H1, and IBs were also partially purified on a chelating column (Figure 5b). Again, a significant portion of the extracted protein did not bind efficiently to a chelating column.

Refolding Studies

No reliable method is known for predicting conditions that allow for proper protein refolding. Identification of optimal parameters, therefore, remains an empirical science. There are many factors involved in protein refolding, including ionic strength, pH and oxidation state. Additives such as chaotropic chemicals and other proteins can also influence protein refolding. To best approach the identification of proper refolding conditions, a variety of commercially available refolding screens were used. The QuickFold protein refolding kit from AthenaES contains 15 different refolding conditions. Solubilized IBs were rapidly diluted into buffer, incubated for one hour and then analyzed for successful refolding by analysis of the soluble fraction on a SDS-PAGE

gel. Experiments performed with IBs extracted without Z3-14 detergent did not produce

any soluble protein in any of the 15 conditions. IBs extracted in the presence of Z3-14

detergent were more successful (Figure 6). Conditions 11 (50 mM Tris-Cl pH 8.5, 9.6

mM NaCl, 0.4 mM KCl, 1 mM EDTA, 0.5% Triton X-100, 1 mM DTT) and 12 (50 mM

Tris-Cl pH 8.5, 240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.05% polyethylene glycol

3,550, 1 mM GSH, 0.1 mM GSSH) produced a significant amount of refolded, soluble

protein. Attempts to refold protein while bound to a chelating column were unsuccessful,

as the protein quickly precipitated on the column and was only eluted with high

concentrations of denaturant. Further investigation revealed that the refolded protein

precipitated out of solution within a few hours after rapid dilution.

To examine additional refolding buffers that may allow for increased stability, the

iFOLD 1 screen was used (EMD Biosciences). The iFOLD matrix covers 92 different

buffers based on an extensive literature review of successful refolding experiments (more

information can be found at the REFOLD database http://refold.med.monash.edu.au).

The iFOLD refolding assay was performed in a 96 well plate, and successful refolding

reactions were screened by measuring A_{340}. Smaller A_{340} values correlate with decreased

precipitation. Following cell lysis, a series of detergent and buffer washes were

performed to remove membrane components and contaminating proteins from the IB

pellet. Purified IBs were denatured with the reducing agent tris(2-carboxyethyl)

phosphine HCl (TCEP) and the detergent N-Lauroylsarcosine in the recommended lysis

buffer supplemented with 0.4 mM Z3-14. The IBs prepped under this buffer system

contained a high level of contaminants, which were present despite denaturing

purification with a chelating column. IBs extracted using the Athena solubilization

buffer and applied to the iFOLD matrix were not as successfully refolded as those
extracted with the iFOLD buffers. The three best conditions are shown in Figure 7:
Condition 1E was composed of 50 mM Tris pH 8.0, 1 mM TCEP and 20% w/v glycerol.
Condition3E was composed of 50 mM Tris pH 8.0, 100 mM NaCl and 1 mM TCEP.
Condition 3F was composed of 50 mM Tris pH 8.5, 100 mM NaCl, 3.8 mM GSH, 1.2
mM GSSG and 0.1% PEG6000 w/v. Different refolding buffers produced soluble
proteins with slightly different migration patterns. Addition of DTT normalized protein
migration. However all soluble protein precipitated out of solution over the course of 48
hours.

Since P3H1 is found in complex with CRTAP and CypB (47), P3H1 may require
these proteins to confer stability. To these ends, refolding experiments were conducted
with P3H1, CRTAP and CypB. Denatured extracted P3H1 and CRTAP were added at 1
mg/mL to refolding reactions. Soluble CypB was also added at 1mg/mL. The two best
conditions from the QuickFold protein screen were tested with P3H1 only, CRTAP only,
and a combination of P3H1, CRTAP and CypB (Figure 8). Although refolding of the
complex greatly enhanced the solubility of CRTAP in condition 11, neither of the
refolding reactions were stable over the course of two days. All protein except for CypB
precipitated out of solution. A similar experiment was performed with the three best
iFOLD conditions (Figure 9). The refolded protein was much more stable, with most of
the protein staying in solution over the course of a week at 4ºC. However, the refolded
proteins were not recognized by the 1C10 antibody used to purify P3H1 from chick
embryos. KD90, an antibody that recognizes reduced P3H1, reacted with the refolded
protein (Figure 10). The refolded proteins may, therefore, not be in the correct biological

conformation.   Attempts to purify the refolded trio of proteins on size exclusion or chelating columns were unsuccessful.

Expression in Insect Cells

Both chick P3H1 and mouse CRTAP were cloned into the pMIB/V5-His A vector. This vector contains a Honeybee melittin secretion signal followed by a melittin cleavage site and the sequence of interest followed by V5 and His tags. Constructs were transfected into *Spodoptera frugiperda* (Sf9) insect cells using calcium phosphate, Insect GeneJuice, or Cellfectin transfection reagents. Protein was assayed from secreted media as well as whole cell lysates 3, 4 and 5 days post-transfection. Transfection of pMIB(cP3H1) with Insect GeneJuice followed by harvest three days-post transfection produced the highest yield of recombinant protein. P3H1 was not visible in Coomassie stained SDS-PAGE gels, although protein was detectable with a V5 antibody (Figure 11). Western blots using the 1C10 antibody produced barely detectable signal, suggesting that the recombinantly expressed protein was not correctly folded. The pMIB(mCRTAP) construct failed to exhibit any protein expression by either SDS-PAGE or western blot. Due to the low (or absent) yield of protein, further experiments with insect cell expression systems were not pursued.

Extraction of P3H1●CRTAP●CypB complex from Chick Embryos

Soluble P3H1●CRTAP●CypB complex can be purified out of chick embryos. However, the purified complex was relatively low in concentration. Moreover, the complex of proteins precipitated out of solution over time (Figure 12). Attempts to

concentrate the protein resulted in precipitation with no overall concentration of the sample. Addition of glycerol did not stabilize the complex.

When run under non-reducing conditions, several additional bands of larger apparent molecular weight were visible (Figure 12, 14). Occasionally these bands were visible under reducing conditions, depending on the prep. Higher molecular weight aggregates are consistent with laser light scattering experiments, which found a large molecular weight aggregate in addition to the 1:1:1 complex of ~150 kDa (Figure 13). Fresh preps displayed lower ratios of aggregate to 1:1:1 complex, consistent with precipitation of protein over time.

To further examine the composition of the higher molecular weight bands, several lanes of P3H1●CRTAP●CypB complex were run under non-reducing conditions on a 12% SDS-PAGE gel. Following staining with GelCode Blue, bands of similar size were excised from the gel. Three large molecular weight bands were consistently present, and were numbered 4, 3 and 2 in order of decreasing MW. P3H1 was run as sample 1 as a control (Figure 14a). The samples were subsequently minced into small pieces using a razor blade, loaded onto a SDS-PAGE gel, run under reducing conditions and stained (Figure 14b). Band one, as expected, contained only P3H1. Bands 2, 3 and 4 contained a mixture of a P3H1 fragment, P4H, PDI and a blocked protein which is most likely CRTAP. No evidence of CypB was present.

The effect of detergents on the extraction of complex was examined. Chick preps in the presence or absence of 0.1% Tween-20 in the extraction and gelatin Sepharose running buffers were performed. The proteins eluted from the Gelatin Sepharose column were analyzed on an SDS-PAGE gel. P3H1 protein was further analyzed by western blot

(Figure 15a). The presence of detergent in the gelatin Sepharose running buffers greatly

enhanced the yield of P3H1 and CRTAP, although the yield of remaining proteins was

not significantly altered by the presence or absence of detergent in the gelatin Sepharose

buffers.

Recent studies on collagen-associated proteins in the rER have used Tween-20 in

both the extraction and gelatin Sepharose buffers (47; 42; 41). However, previous

protocols have used up to 1% Triton X-100 or 1% NP-40 during extraction of crude

microsomal extracts (40; 51). To determine the most efficient extraction conditions,

microsomal pellets were extracted using 1% Triton X-100, 1% NP-40, 0.1% Tween 80 or

0.1% Tween 20. Following extraction, collagen-associated proteins were isolated using a

gelatin Sepharose column (Figure 15b). Protein yield was analyzed on a non-reducing

SDS-PAGE gel and by western blot using the P3H1 antibody 1C10. 1% Triton X-100

and 1% NP-40 effectively increased the yield of extracted proteins.

To better explore the effect of detergent on gelatin Sepharose chromatography, 1

g of microsomal pellet was extracted and run on a gelatin Sepharose column under two

different conditions: the first condition was extracted and run with 0.1% Tween 20 in all

buffers, while the second condition extracted pellets with 1% Triton X-100 and included

0.1% Triton X-100 in the gelatin Sepharose running buffers. Following extraction, the

proteins were bound to a gelatin Sepharose column, washed with a high salt buffer and

then eluted with a low pH buffer. Fractions from each step were analyzed by western

blot with the 1C10 antibody. When extracted with 1% Triton X-100 and run with 0.1%

Triton X-100 in the gelatin Sepharose buffers, P3H1 elutes in the high salt fraction in

addition to the low pH fraction (Figure 16a).

A large scale prep using microsomal pellets harvested from 16 dozen chick embryos was extracted with 1% Triton X-100 and run with gelatin Sepharose buffers containing 0.1% Triton X-100. A normal compliment of rER proteins were found within the low pH elution. However, P3H1 also eluted in the high salt wash. P3H1 appears to elute with CypB in the early portion of the elution peak, and with HSP47 and CRTAP towards the end of the elution peak (Figure 16b). The low pH pool and the high salt pool were dialyzed into PBS and run on a 1C10 antibody column. The low pH pool, when applied to the antibody column, eluted the P3H1●CRTAP●CypB complex. The high salt pool, when applied to the antibody column, contained primarily P3H1.

Discussion

The P3H1●CRTAP●CypB complex has been shown to be a prolyl 3-hydroxylase, a PPIase and a molecular chaperone (47; 42). Defects in any component of the complex are associated with severe and lethal recessive osteogenesis imperfecta (57-59; 62; 73; 74; 77; 205). Despite numerous reports demonstrating the importance of the complex, little is known about the structure of P3H1, CRTAP or the complex as a whole. Sufficient complex for functional assays can be isolated from chick embryos and by recombinant expression in insect cells (47; 42; 65). However, structural analysis has been difficult due to low protein yield and concentration. Here we attempted to harvest soluble, high concentration P3H1 and CRTAP in *E. coli,* insect cells and out of chick embryos.

Recombinant P3H1 expression proved to be unsuccessful in a variety of expression systems. *E. coli* produced protein was almost exclusively found in IBs

(Figure 3). Extraction studies were performed to maximize denatured protein yield out of them. High concentrations of Urea and addition of the zwitterionic detergent Z3-14 were necessary to achieve sufficient extraction (Figure 4). Denatured protein was subsequently purified and used in refolding assays (Figure 5). P3H1 was capable of refolding under a variety of conditions; however, it quickly precipitated out of solution (Figures 6 & 7).

Since P3H1 is found in complex with CRTAP and CypB, refolding experiments with all proteins in the complex were performed (Figures 8 & 9). Although refolding was successful and stable under some conditions, the overall quality of extracted protein was questionable as the 1C10 antibody raised against native protein extracted from chick embryos did not recognized the refolded product (Figure 10). The 1C10 antibody has been used to purify P3H1 and P3H1●CRTAP●CypB complex that was used in enzyme, chaperone and PPIase assays. 1C10, therefore, recognizes biologically functioning protein. KD90, an antibody that recognizes reduced protein, readily detected the refolded product, further strengthening the argument that the refolded product was not in the biologically relevant conformation.

Expression in Sf9 cells produced soluble protein in minute amounts (Figure 11). A previous study from the Myllyharju laboratory expressed human P3H1, in addition to human P3H2 and P3H3, in insect cells using a recombinant bacculovirus system(65). Roughly 40% of the total human P3H1 produced was extracted from whole cell lysates using 0.1% Triton X-100, with the remaining protein only solubilized by 1% SDS. Our constructs carried a honeybee melittin signal sequence designed to excrete the target protein outside of the cell. It has been shown that the addition of this signal enhances

protein folding and stability (206). Chick P3H1 was expressed in low amounts, detectable only by western blot against the V5 epitope and was produced in insufficient quantities for preliminary purification studies. However, it is of interest that our pMIB/V5(P3H1) construct produced soluble protein with no detectable protein in the whole cell lysate. This is an improvement over the BaculoGold system employed by the Myllyharju laboratory for expression, where more than half of the produced protein was insoluble.

Although not covered in this work, extensive studies involving expression of P3H1, P3H2, P3H3 and CRTAP in mammalian cell lines were pursed by myself and others in the laboratory. These studies were marginally successful as chick and human P3H constructs were expressed, albeit at low levels. P3H1 was produced at the highest concentrations (detectable by Coomassie staining of an SDS-PAGE gel). However, the protein precipitated out of solution during the course of purification. CRTAP was never significantly produced in mammalian cell lines. Co-expression of P3H1 and CRTAP or triple-expression of P3H1, CRTAP and CypB did not yield an improvement in protein production or stability in our experiments.

Thus far, the only successful recombinant system for P3H proteins was produced by the Myllyharju laboratory using a BaculoGold viral system to infect and produce human P3H proteins in Sf9 cells (65). Recombinant P3H2 protein was barely detectable by western blot against a human P3H2 peptide. Luckily, enzyme assays are exquisitely sensitive: extracts of P3H2 cells were shown to be enzymatically active. Since P3H proteins are the only proteins known to 3-hydroxylate collagen substrates, it is unlikely that proteins found in Sf9 extracts would exhibit additional P3H activity. In fact, control

experiments using extracts of insect cells expressing human C-P4H-I did not exhibit 3-hydroxylase activity. However, cells contain multiple proteins that exhibit chaperone and disulfide isomerase activity. Therefore, whole cell lysates cannot be used as a protein source for chaperone or disulfide isomerase activities. Until P3H2 can be produced in sufficient quantities for purification, the full functionality of the recombinant protein expressed in Sf9 cells can not be assessed. Human P3H1, despite being produced at higher concentrations, did not demonstrate any enzymatic activity. The human P3H3 construct was entirely insoluble. Thus recombinant P3H2 is the only protein to be over expressed as a partially functional protein, although the chaperone and disulfide isomerase activity remains to be determined. Recombinant P3H1 that demonstrates any known functional activity has yet to be produced.

P3H1●CRTAP●CypB complex has been purified from chick embryos using a gelatin Sepharose column and an 1C10 antibody column. The former has previously been used to identify other proteins and complexes of proteins that bind to denatured collagen (47; 40; 51; 207). The 1C10 antibody column has been used to purify biologically active P3H1 which is found in complex with CRTAP and CypB. P3H1 can be further purified from the complex by reapplying the complex to the column and eluting with variable pH to select for P3H1 enriched fractions. The quality and concentration of protein is variable from prep to prep, due in part to the erratic quality of chick embryos. Crude microsomal pellets can be frozen from multiple batches of embryos in order to increase yield (up to three batches of pellets which represent 48 dozen chick embryos have been used at one time), although the protein was seldom

purified at a concentration above a few micromolar. Additionally, the complex readily

precipitated out of solution (Figure 12).

When run under non-reducing conditions, large molecular weight complexes were

observed (Figure 13, 14). Reduction of the protein prior to electrophoresis generally

eliminates these bands, indicating that they are formed of aggregates of P3H1, CRTAP

and CypB. Occasionally, even under stringent reducing conditions, these higher

molecular weight bands still appeared. Further investigation of these large molecular

weight complexes revealed that they contain a fragment of P3H1, CRTAP and the α and

β subunits of P4H. These proteins were relatively low in concentration, although the

presence of P4H subunits have been occasionally observed in P3H1●CRTAP●CypB

preps. The P3H1 fragment was identified by N-terminal sequencing, although the

cleavage point was not determined as there was insufficient protein for the tryptic

digestion necessary to complete the sequencing of the fragment. The higher molecular

weight bands appear in increasing intensity as a sample ages (Figure 14). Additionally,

the higher molecular weight bands appeared in three discrete sizes that were stable

enough to resist boiling and denaturation by SDS, rather than being found as a large

protein aggregate retained in the well of a SDS-PAGE gel. However, laser light

scattering did not indicate the presence of discrete complexes other than the 1:1:1 155

kDa complex and a multi-million Dalton aggregate (Figure 13). The heat and SDS-

resistant aggregates likely assemble into larger aggregates in solution. The consistent

nature of the SDS-resistant aggregates is intriguing, perhaps indicating an intermediate

and controlled aggregate. Possibly, the P4H and P3H1 fragment assemble in complex

with full length P3H1 and CRTAP in a regulatory mechanism designed to pull

malfunctioning proteins out of solution. Consistent with this theory, as the 1C10 antibody columns degrade, increasing levels of P4H are retained on the column (data not shown).

Detergents appear to play an important role in the efficient extraction of P3H1 and CRTAP out of microsomal pellets. The presence of 0.1% Triton X-100 is important for extraction of all collagen-associated proteins; however, P3H1 and CRTAP appear to be particularly sensitive to the presence of 0.1% Triton X-100 in the gelatin Sepharose chromatography buffers (Figure 15). Previous studies used different detergents during initial extraction of pellets, and high concentrations of Triton X-100 were indeed more effective in extracting P3H1 from microsomal pellets. This is of particular interest as many membrane-associated proteins require high levels of detergent for solubilization. Collagen has been shown to associate with the rough endoplasmic reticulum membrane during folding (33). It is likely that proteins involved with collagen biosynthesis are also membrane associated. Proteins extracted with 1% Triton X-100 and run on a gelatin Sepharose column with buffers containing 0.1% Triton X-100 exhibited slightly different elution profiles (Figure 16a). Traditional extraction and purification of collagen-associated proteins with Tween 20 as a detergent elute P3H1 exclusively in the low pH wash. Extraction and purification performed with Triton X-100 detergent also elute P3H1 in the high salt wash (Figure 16b). This fraction, when pooled, dialyzed against PBS and run on a 1C10 antibody column, does not co-purify with CRTAP or CypB (Figure 16c). This suggests that a free form of P3H1 which does not strongly associate with CRTAP or CypB does not efficiently bind to denatured collagen. However, the presence of P3H1 in a high salt wash may also indicate that this fraction of protein is

incorrectly folded and unable to bind to gelatin Sepharose. Aggregates of this protein, on the other hand, were not observed under non-reducing conditions on a SDS-PAGE gel. Further studies are necessary to determine the role of these forms of P3H1.

Structural studies on P3H1 and the P3H1●CRTAP●CypB complex remain elusive. Biophysical analysis of P3H1 is hampered by low concentrations. Although the P3H1●CRTAP●CypB complex can be purified in sufficient quantities for enzyme or chaperone assays, structural studies remain difficult. Since the solubility of the complex is enhanced by detergents, the complex may be membrane associated. Although the complex is tightly associated when isolated from a gelatin Sepharose column, additional proteins may be necessary to enhance the long term stability of the complex. Proteins found in the gelatin Sepharose elutant include the chaperone proteins FKBP65 and HSP47, both of which may stabilize P3H1 in addition to their known functions as collagen chaperones.

Further investigations into the stability of the complex derived from chick embryos may yield information that will enhance over expression of recombinant constructs. Once sufficient yield of protein has been achieved, questions about the structure and assembly of P3H1 and the P3H1●CRTAP●CypB complex can be pursued: What is the structure of the common N-terminal portion of the P3H proteins and CRTAP? Do the N-terminal portions of P3H1 and CRTAP interact? What structural differences differentiate the enzymatic portion of P3H from P4H? What residues are important for substrate recognition? What are the binding constants between the members of the complex, and what is the affinity of the complex for known sites of 3-hydroxylation? We are not currently able to address these questions; however, when

sufficient protein concentrations and quantities have been produced, structural and

biochemical analysis of the individual proteins as well as the complex will significantly

advance our understanding of collagen biosynthesis.

Condition Number	Detergent	Type	MW (ave)	CMC mM	mM in buffer
1	Triton X-100	Non-ionic	650	0.2-0.9	0.5
2	Tween 80	Non-ionic	1310	0.012	0.015
3	NP-40	Non-ionic	603	0.05-0.3	0.025
4	Deoxycholic Acid	Ionic	432.6	1.5	1.5
5	CHAPS	Zwitterionic	614.9	9	9
6	Z3-08	Zwitterionic	279.6	330	330
7	Z3-10	Zwitterionic	307.6	25-40	40
8	**Z3-12**	**Zwitterionic**	**335.6**	**2-4**	**4**
9	**Z3-14**	**Zwitterionic**	**363.6**	**0.1-0.4**	**0.4**
10	Z3-16	Zwitterionic	391.6	0.01-0.06	0.06

Table 1. Buffers used for inclusion body screening. Final detergent concentrations used in the solubilization buffer are shown in the right most column. Detergents most effective in inclusion body extraction are shown in bold.
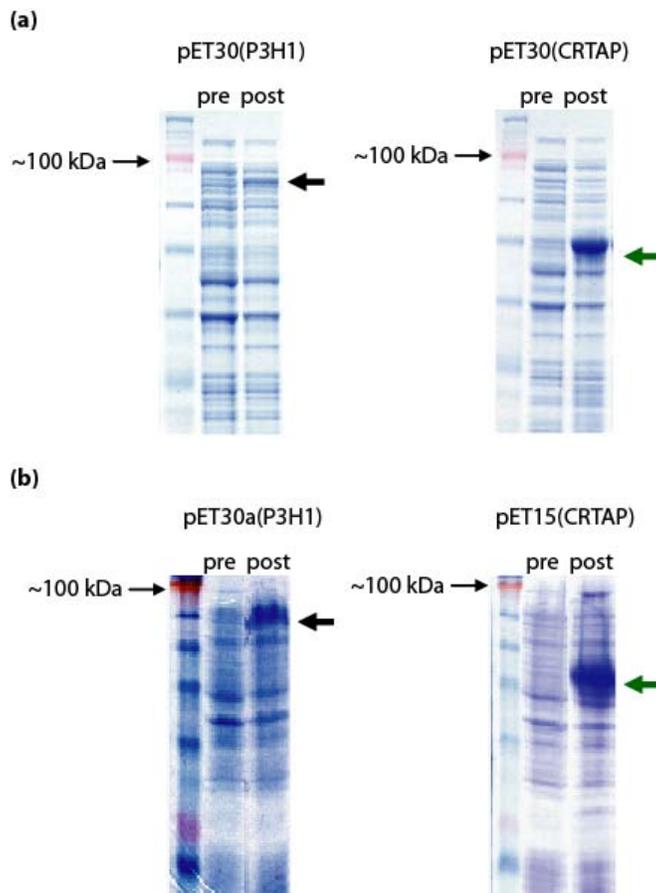
Figure 1. Test inductions of P3H1 and CRTAP constructs. (a) Coomassie Blue stained 12% Bis-Tris gels of pET30(cP3H1) and pET30(mCRTAP) test inductions. An arrow marks the induced protein band. (b) Coomassie Blue stained 12% SDS-PAGE gels of pET30a(cP3H1) and pET15(mCRTAP) test inductions. A black arrow marks P3H1 while a green arrow marks CRTAP.
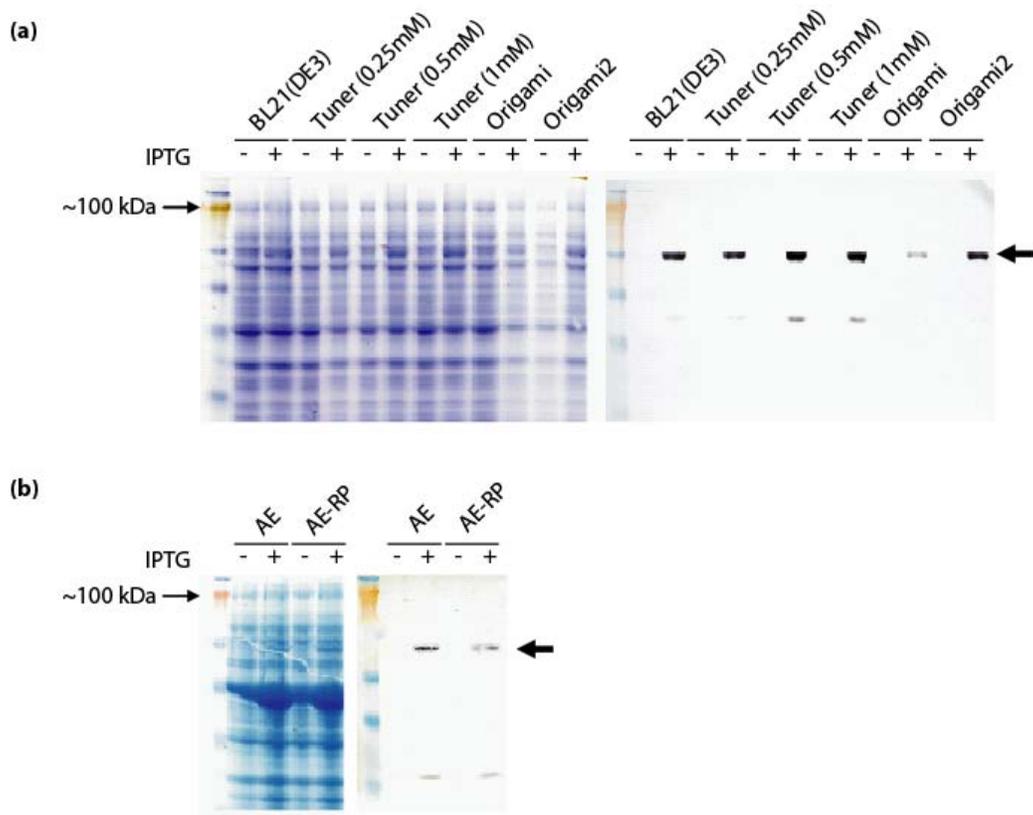
Figure 2. Test inductions of full length cP3H1 in a variety of cell lines. (a) Coomassie Blue stained 12% SDS-PAGE and anti-His western blot of test inductions at 1 mM IPTG for all cell lines, except for Tuner cells where the IPTG concentration is listed in parenthesis. Uninduced cells are indicated with a - sign, induced cells with a + sign. An arrow marks the induced protein. (b) Coomassie Blue stained 12% SDS-PAGE and anti-His western blot of Arctic Express (AE) or Arctic Express RP (AE-RP) cell line test inductions labeled as above.
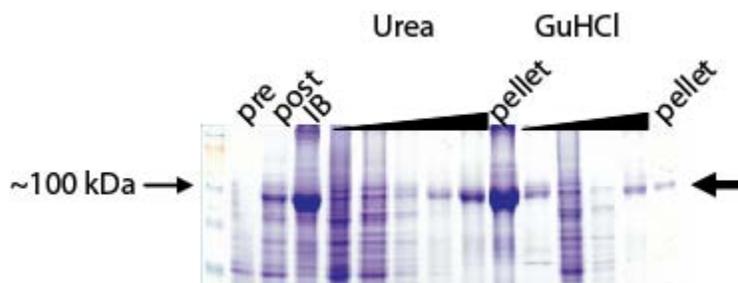
Figure 3. Full length P3H1 is insoluble and high levels of denaturant are required for efficient extraction. Test induction of pET30a(cP3H1) clearly showed induction of protein production, although the majority of the induced protein was present in inclusion bodies (labeled IB). Inclusion bodies were purified in a buffer of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.1% TX-100. Purified inclusion bodies were subsequently extracted with either 0, 2, 4, 6 or 8M Urea or 0, 2, 4 or 6M GuHCl. Samples were precipitated using trichloracetic acid to remove denaturant, resuspended in 1X SDS-PAGE sample buffer and analyzed on a 10% SDS-PAGE gel.
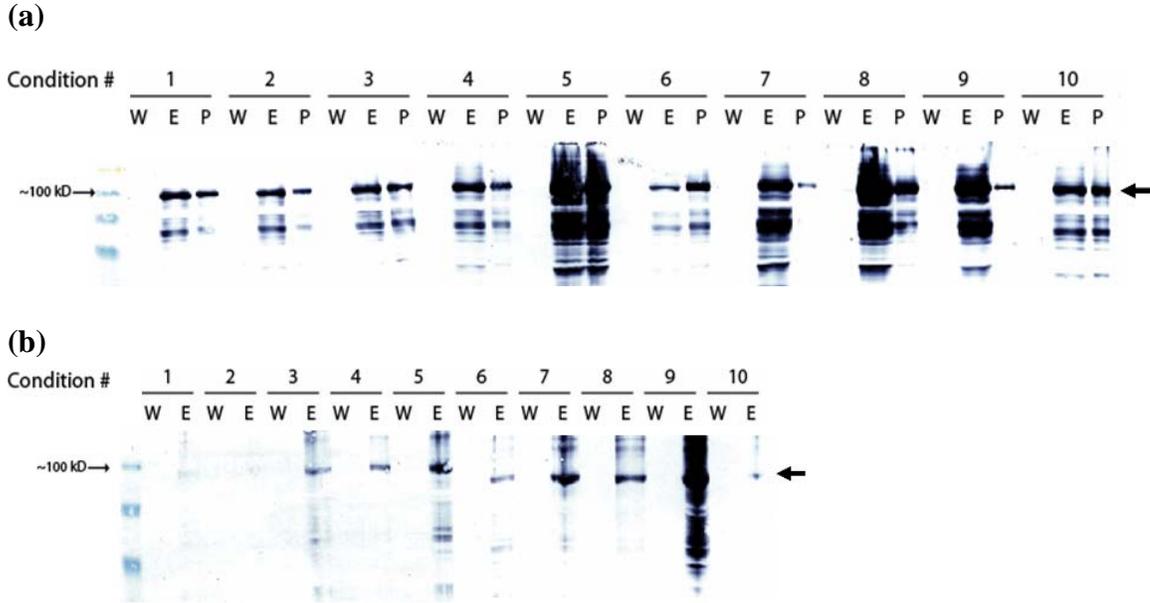
(a)

(b)

Figure 4. Inclusion body extractions performed with a variety of detergents. (a) Western blot against anti-His$_6$ tag protein showing P3H1 protein prepared from Urea buffers containing a variety of detergents labeled 1-10 according to Table 1. Wash fractions are labeled with a W, extracted protein with an E and insoluble pelleted protein with a P. (b) Western blot against anti-His$_6$ tag showing P3H1 protein prepared from GuHCl buffers containing a variety of detergents labeled 1-10 according to Table 1. Wash fractions are labeled with a W and extracted protein with an E.
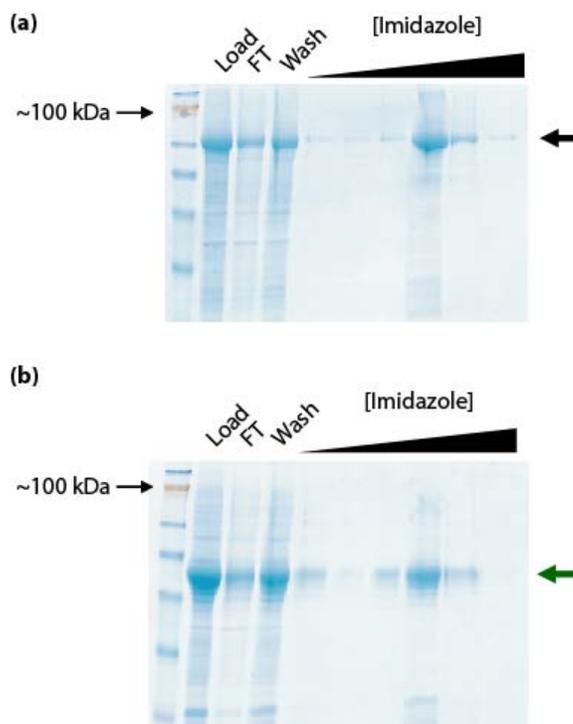
Figure 5. Inclusion body purification. (a) Extracted P3H1 purification under denaturing conditions on a chelating column. Inclusion bodies were resuspended in solubilization buffer and loaded onto a HiTrap Chelating column charged with Ni^{2+} ions and eluted with a gradient of solubilization buffer containing 0 to 500 mM imidazole. Samples of the load, flow through, wash and gradient fractions were analyzed on a 12% SDS-PAGE gel under non-reducing conditions. The black arrow indicates the P3H1 protein band. (b) Extracted mCRTAP purification under denaturing conditions on a chelating column. Inclusion bodies were resuspended in solubilization buffer and purified as above Samples of the Load, flow through, wash and gradient fractions were analyzed on a 12% SDS-PAGE gel under non-reducing conditions. The green arrow indicates the CRTAP protein band.
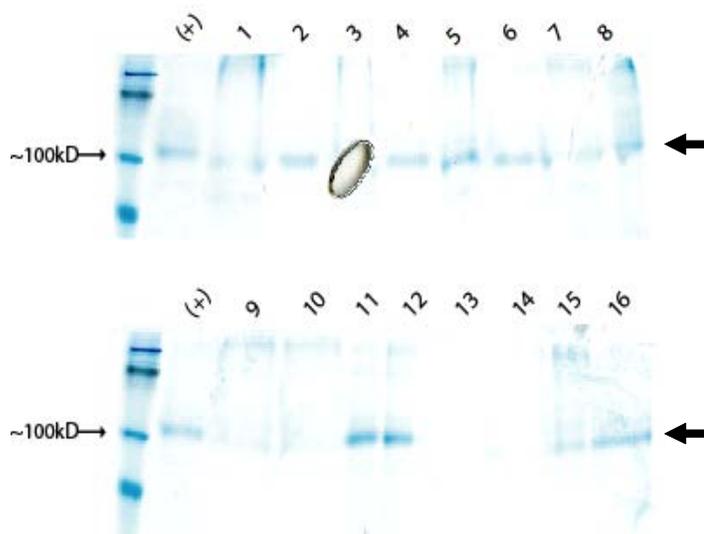
Figure 6. QuickFold protein refolding assay. Soluble P3H1 protein samples were analyzed on a 12% SDS-PAGE gel. Refolding conditions are labeled 1 -16. A positive control of 30% input is marked with a (+).
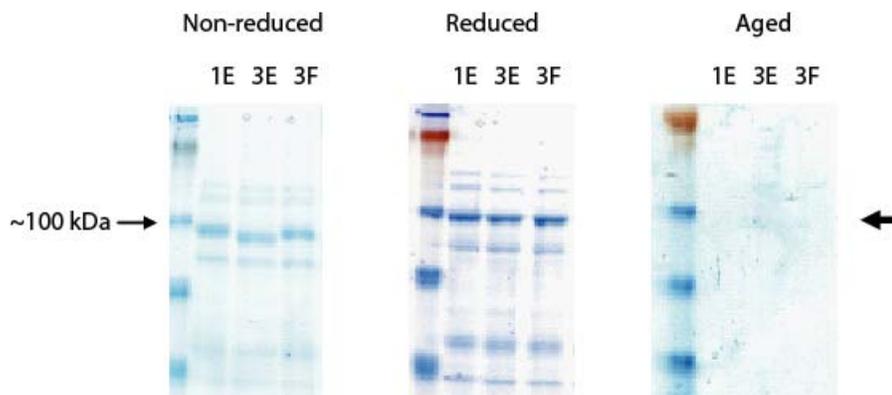
Figure 7. Best conditions of P3H1 refolding from the iFOLD screen. Refolded P3H1 appeard in two different primary bands, which resolved to a single band when reduced with 1mM DTT when run on a 12% SDS-PAGE gel. P3H1 precipitated out of solution over the course of several days. The three best conditions are shown above. 1E was composed of 50 mM Tris pH 8.0, 1 mM TCEP and 20% w/v glycerol. 3E was composed of 50 mM Tris pH 8.0, 100 mM NaCl and 1 mM TCEP. 3F was composed of 50 mM Tris pH 8.5, 100 mM NaCl, 3.8 mM GSH, 1.2 mM GSSG and 0.1% PEG6000 w/v.
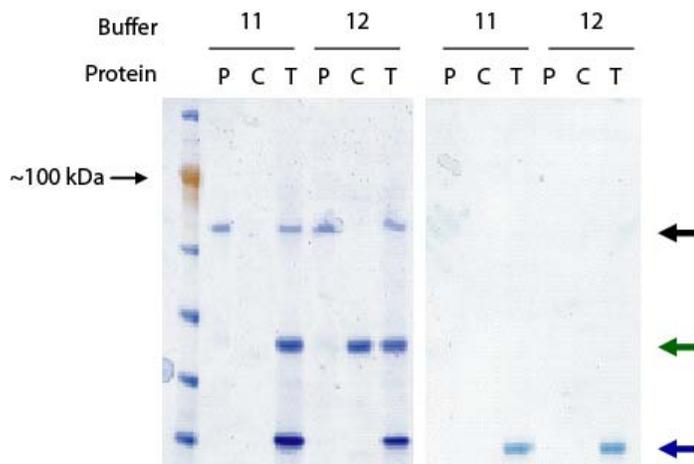
Figure 8. P3H1 and CRTAP protein refolding in the presence of CypB using the best QuickFold kit conditions.   Condition 11 was composed of 50 mM Tris-Cl pH 8.5, 9.6 mM NaCl, 0.4 mM KCl, 1 mM EDTA, 0.5% Triton X-100, 1 mM DTT; condition 12 was composed of 50 mM Tris-Cl pH 8.5, 240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.05% polyethylene glycol 3,550, 1 mM GSH, 0.1 mM GSSH. P, C and T represent protein folding reactions containing P3H1 only, CRTAP only and P3H1, CRTAP and CypB mixed in equal concentrations. P3H1 protein bands are marked with a black arrow, CRTAP protein bands are marked with a green arrow, and CypB protein bands are marked with a blue arrow. Soluble proteins following protein refolding were analyzed on a 4-12% SDS-PAGE gel. The left panel shows soluble samples taken immediately after refolding. The right panel shows soluble samples 48 hours after refolding.
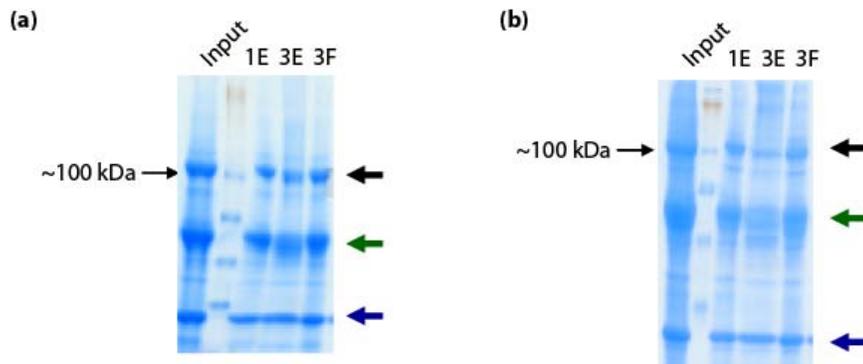
Figure 9. P3H1 and CRTAP protein refolding in the presence of CypB using the best iFOLD conditions.   1E was composed of 50 mM Tris pH 8.0, 1 mM TCEP and 20% w/v glycerol. 3E was composed of 50 mM Tris pH 8.0, 100 mM NaCl and 1 mM TCEP. 3F was composed of 50 mM Tris pH 8.5, 100 mM NaCl, 3.8 mM GSH, 1.2 mM GSSG and 0.1% PEG6000 w/v. (a) Soluble portions of refolding reactions were analyzed by 12% SDS-PAGE gels. P3H1 protein bands are marked with a black arrow, CRTAP protein bands are marked with a green arrow, and CypB protein bands are marked with a blue arrow. (b) Soluble portions of refolding reactions after 7 days of storage at 4ºC. Protein bands are labeled as above.
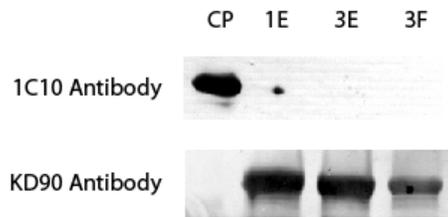
Figure 10. Western blot analysis of refolded P3H1. Chick prep derived P3H1●CRTAP●CypB and refolded P3H1, CRTAP and CypB samples were run on an non-reduced SDS-PAGE gel and probed with antibodies. The 1C10 antibody recognizes non-reduced P3H1 purified from chick embryos. The KD90 antibody recognizes reduced P3H1 purified from chick embryos.
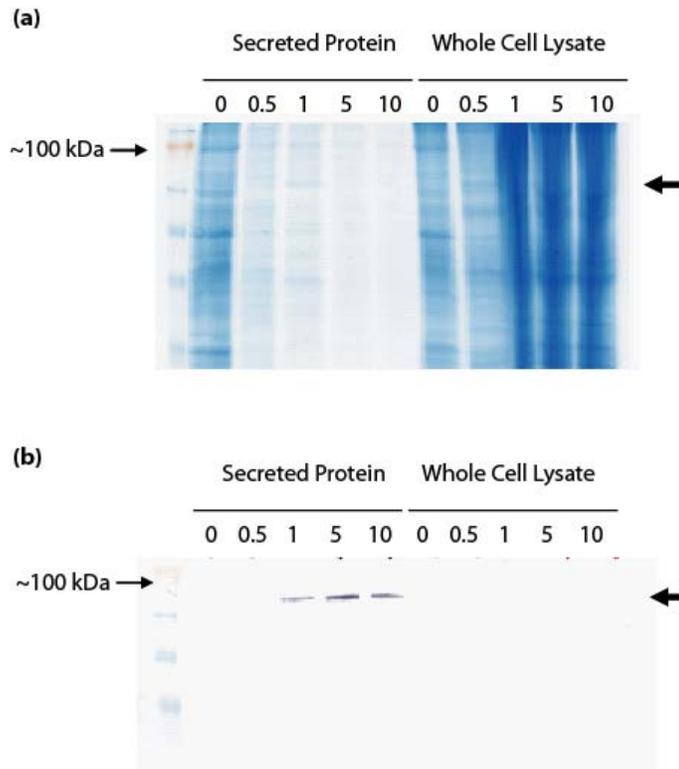
Figure 11. Analysis of recombinant P3H1 expression in Sf9 cells. Sf9 cells were transfected using 0, 0.5, 1, 5 or 10 μg of pMIB(cP3H1) DNA using Insect GeneJuice. Three days post-transfection, media and whole cell lysates were collected for analysis by (a) Coomassie Blue stained 12% SDS-PAGE gel or (b) western blot using a V5 epitope.
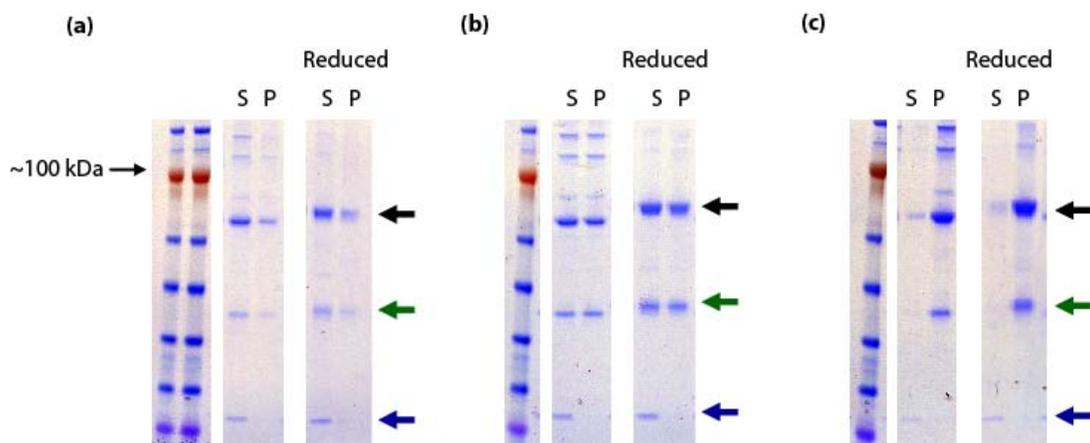
Figure 12. Analysis of P3H1●CRTAP●CypB complex stability as a function of time under non-reducing (NR) and reducing (R) conditions on a 4-12% SDS-PAGE gel stained with GelCode Blue. Relative migration of P3H1, CRTAP and CypB proteins are marked with black, green and blue arrows, respectively. Lanes labeled with an S represent soluble protein, lanes labeled with a P represent precipitated protein. Gels are shown under non-reducing and reducing conditions. (a) Samples taken immediately following dialysis into PBS. (b) Samples taken 48 hours after dialysis into PBS. (c) Samples taken five days after dialysis into PBS.

(a)

(b)
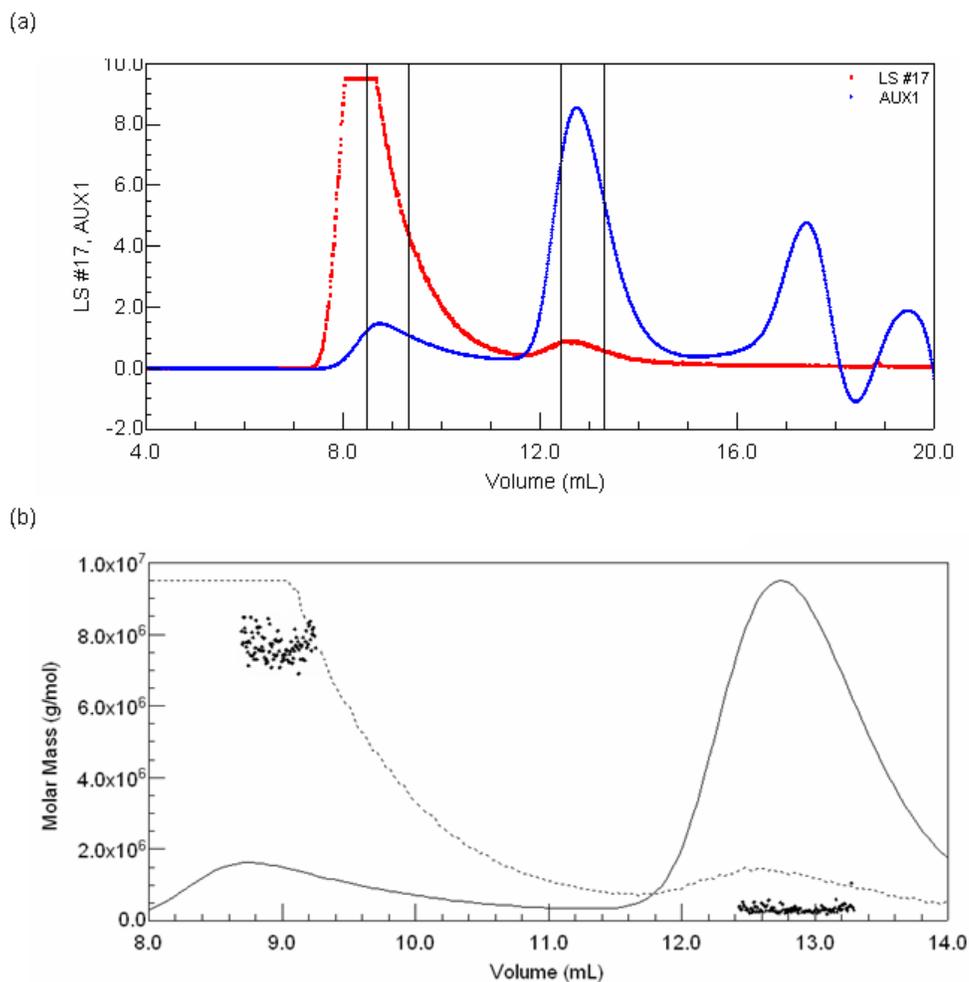
Figure 13. P3H1●CRTAP●CypB exist as a 1:1:1 complex as well as a high molecular mass aggregate. (a) Elution profile of P3H1●CRTAP●CypB complex run on a molecular sieve column attached to a laser-light scattering instrument. The refractive index signal is shown in blue and the light-scattering signal is shown in red. (b) The two major peaks, contained by the vertical bars in (a) were analyzed and the molar mass assessed. The peak at ~ 9 mL was a large aggregate with a mass of several million Daltons. The peak eluting between 12.5 and 13.25 mL had a mass of ~150 kDa.
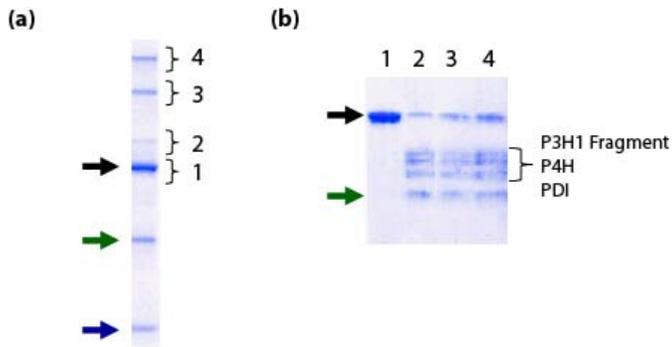
Figure 14. P3H1 complex is present in higher order aggregates that contain a P3H1 fragment, P4H, PDI and CRTAP. (a) 4-12% SDS-PAGE gel stained with GelCode Blue with P3H1, CRTAP and CypB marked with black, green and blue arrows respectively. Under non-reducing conditions, several higher molecular weight bands were identified (labeled 2, 3 and 4 with P3H1 identified as band 1). (b) Following excision from the gel, bands were minced and re-run on a reducing 12% SDS-PAGE gel and stained with GelCode Blue. Samples were submitted for protein sequencing: the band marked with a black arrow was identified as P3H1. The band marked with the green arrow was blocked and most likely corresponds to CRTAP. The three remaining bands were identified as a P3H1 fragment, P4H and PDI.
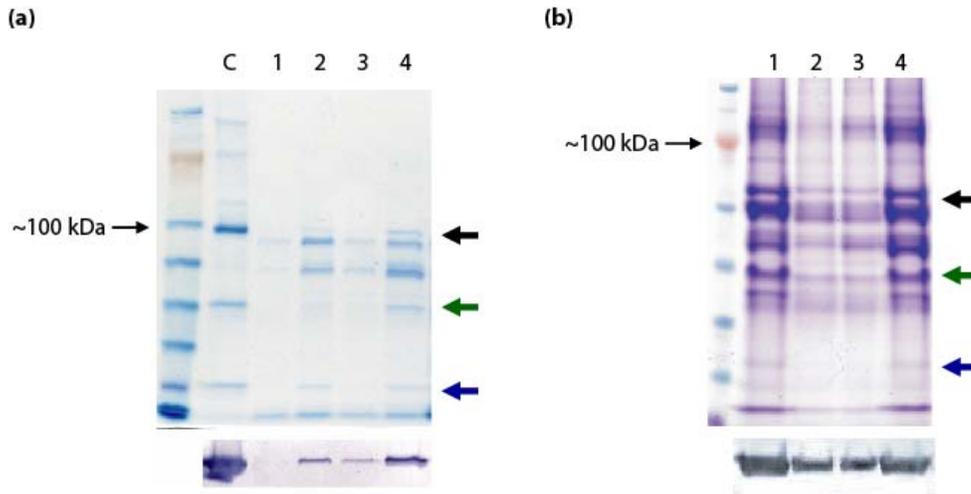
Figure 15. The presence of absence of detergent affect the yield of protein from chick embryos. (a) 0.1% Tween-20 affects the yield of P3H1 from chick embryos. Eluted proteins from a gelatin Sepharose column were run on a 12% SDS-PAGE gel and stained with GelCode Blue. C represents a sample of purified P3H1●CRTAP●CypB complex. Lane 1 did not have Triton X-100 in either the extraction or gelatin Sepharose buffers. Lane 2 had detergent in the extraction buffer only. Lane 3 had detergent in the gelatin Sepharose buffer only. Lane 4 had detergent in both the extraction and gelatin Sepharose buffers. P3H1, CRTAP and CypB are marked with black, green and blue arrows, respectively. Below the gel is a western blot against P3H1 using the 1C10 antibody for the same conditions. (b) Detergents affect the yield of P3H1 from chick embryos. Eluted proteins from a gelatin Sepharose column were run on a 4-12% SDS-PAGE gel and stained with GelCode Blue. P3H1, CRTAP and CypB are marked with black, green and blue arrows, respectively. Lane 1 was extracted in the presence of 1% Triton X-100, lane 2 with 0.1% Tween 80, lane 3 with 0.1% Tween 20 and lane 4 with 1% NP-40. Below the gel a western blot against the P3H1 antibody 1C10 for the same conditions is shown.
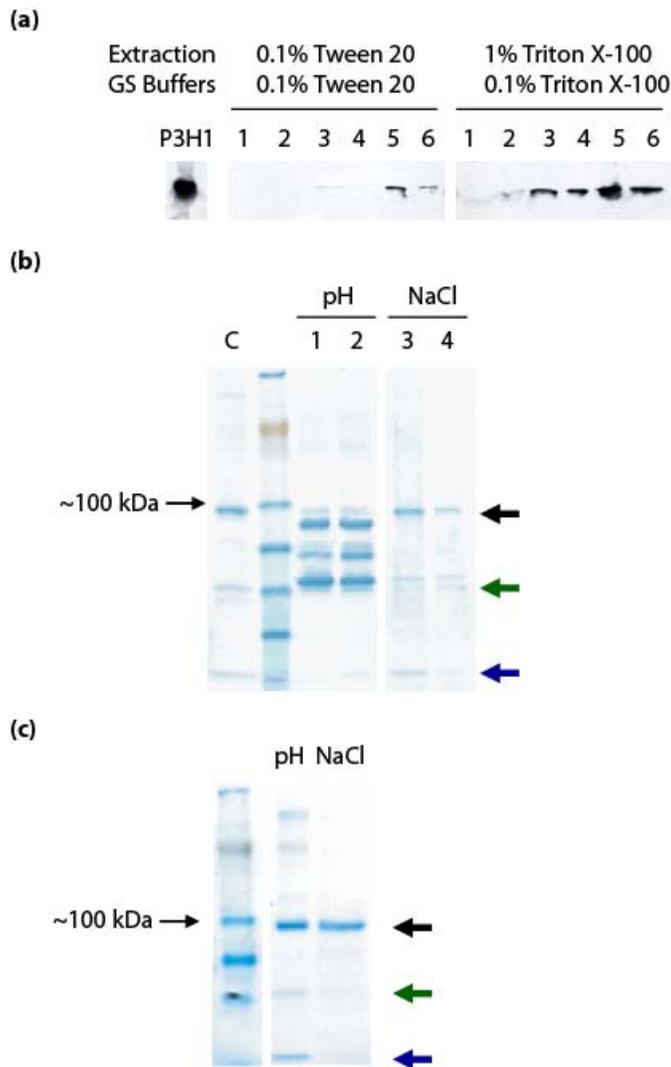
Figure 16. Proteins extracted and run with buffers containing 0.1% Tween 20 display different binding properties than proteins extracted with 1% Triton X-100 and run with 0.1% Triton X-100 containing buffers. (a) Western blot of a small protein prep. Purified P3H1 is shown as a control in the left most lane. Microsomal pellets were extracted with either 0.1% Tween 20 or 1% Triton X-100. The extracts were run on a gelatin Sepharose column in the presence of either 0.1% Tween 20 or 0.1% Triton X-100. Lane 1, column flow through. Lane 2, wash fraction. Lane 3, high salt elution. Lane 4, re-equilibration in running buffer. Lane 5, elution at low pH. Lane 6, re-equilibration in running buffer. (b) A large scale prep of microsomal pellets extracted with 1% Triton X-100 and run on a gelatin Sepharose column with buffers containing 0.1% Triton X-100. C marks a reference lane run with purified complex. Fractions representing the front or back of elution peaks were run on a 10% SDS-PAGE gel under non-reducing conditions. Lane 1, low pH elution front of peak. Lane 2, low pH elution back of peak. Lane 3, high salt elution front of peak. Lane 4, high salt elution end of peak. P3H1, CRTAP and CypB are marked with black, green and blue arrows, respectively. (c) Pooled peaks eluted with

either low pH or high salt from the gelatin Sepharose column were dialyzed into PBS and further purified on a 1C10 antibody column. P3H1, CRTAP and CypB are marked as above.

CHAPTER 6

Summary, Future Directions and Conclusions

Summary

The type XV trimerization domain is a stable trimer that forms a trimerization fold common to multiplexin collagens

In Chapter 3 of this thesis, I solved the crystal structure of the type XV trimerization domain. The domain is composed of three monomers, each of which contains four β-sheets and an α-helix that form a monomeric hydrophobic core. Three monomers combine to form a trimer that is stabilized by an internal hydrophobic core. Unlike the crystal structures of the type IV, VIII and X trimerization domains, the trimerization domain of type XV collagen was not stabilized by intersubunit exchange or hydrophilic interactions. This is consistent with the only other member of the multiplexin collagen family, type XVIII collagen, whose structure is known. The crystal structures of type XV and XVIII collagen trimerization domains are remarkably similar, despite having only 33% sequence identity. In addition to using crystallographic techniques to solve the structure using molecular replacement, I also performed a variety of biophysical experiments. These have shown that type XV trimerization domain forms a trimer in solution and is extremely stable, with a standard free energy of trimerization in water of 109.0 (±1.0) kJ/mol. These experiments demonstrate that the trimerization domain of multiplexin collagens are extremely stable and form a common three dimensional fold despite low sequence similarity.

The P3H1●CRTAP●CypB complex is a molecular chaperone, a peptidyl-prolyl cis-trans *isomerase and associates with folded type I collagen*

The P3H1:CRTAP:CypB complex was shown to be a potent molecular chaperone by citrate synthase and rhodanese aggregation assays. However, the complex did not stabilize the collagen triple helices of type I or type III collagens. My studies demonstrated that P3H1●CRTAP●CypB does in fact interact with triple helical type I collagen by fibril formation assays. Biacore studies were able to quantitate complex binding to native type I collagen, with a dissociation constant of 7.4 μM. Interestingly, the on-rate of association was similar to HSP47, but the dissociation constant was larger due to an increased off-rate. Given that the enzymatic activity of P3H1 and the P3H1●CRTAP●CypB complex require a denatured chain, it is likely that the complex binds to an unfolded α chain, hydroxylates proline in the X position in a Gly-X-Y sequence and temporarily stabilizes the folded/unfolded junction until triple-helix-stabilizing chaperones are able to bind. These studies demonstrate additional functions of P3H1 outside of 3-hydroxylation, which may contribute to the severe osteogenesis imperfecta phenotypes found in patients with defects in P3H1, CRTAP or CypB.

Recombinant expression of P3H1 does not yield soluble protein

In Chapter 4, I performed a variety of experiments in an attempt to overexpress soluble, recombinant P3H1. Initial experiments did not yield high levels of induction in *E. coli*, although construct modification increased the overall yield. The protein was almost exclusively found in inclusion bodies. Experimentation with detergents and denaturants enhanced extraction of P3H1 from inclusion bodies, although subsequent

purification did not yield adequately clean protein. Refolding experiments were marginally successful: although protein refolded in several conditions, the soluble protein quickly precipitated out of solution over the course of several hours or days. Since P3H1 purified from chick embryos is found in complex with CRTAP and CypB, attempts were made to refold P3H1 and CRTAP simultaneously in the presence of soluble CypB. Several conditions yielded soluble complex which was significantly more stable than refolded P3H1 alone. Unfortunately, the refolded protein was not recognized by a chick antibody raised against folded protein while it readily reacted with an antibody that recognized reduced protein. Thus, the refolded trio of proteins, although more stable, was most likely not in the biologically relevant conformation. Additional experimentation not covered in this thesis has also shown that P3H1 expression in mammalian cells was largely unsuccessful, as the protein yields were extremely low and the protein was difficult to purify and concentrate. In summary, all attempts to overexpress P3H1 failed to yield significant amounts of protein, and the limited protein produced was unstable and precipitated out of solution quickly.

P3H1●CRTAP●CypB complex is marginally stable and extraction is enhanced by detergents.

Since recombinant expression was unsuccessful, I attempted improve the yield of the protein complex from chick embryos. The P3H1●CRTAP●CypB complex precipitates out of solution and, over the course of several days, soluble protein completely disappears. Laser light scattering clearly showed two populations of purified complex: a 1:1:1 complex with a MW of roughly 155 kDa, and a large mass aggregate

was also found. Aggregates of the P3H1●CRTAP●CypB complex were found to contain

a truncated form of P3H1, P4H and PDI. The importance of detergents in chick preps

was studied next. Detergents were necessary for effective extraction of rER proteins

associated with collagens, although P3H1 and CRTAP appear to be especially sensitive

to the presence of detergents in the gelatin Sepharose running buffer. Additionally,

extraction of P3H1●CRTAP●CypB or P3H1 alone was greatly enhanced by high levels

of detergent.

Future Directions

Multiplexin Trimerization Domains

Analyses of multiplexins from other organisms show a trimerization domain of

defined length with no deletions or insertions (Figure 1). Only three residues are

identical throughout all of the sequences although the distribution of hydrophobic

residues important to hydrophobic core formation is relatively well conserved. It appears

that the multiplexin trimerization domain can accommodate significant sequence

variability provided that a specific hydrophobic pattern is maintained. Given the

flexibility in sequence, is it possible for type XV and XVIII trimerization domains to

form heterotrimers?

Although their expression patterns and tissue distributions vary, type XV and

XVIII collagens are often present in the same tissues (132) and can be produced by the

same cell types (105). Multiplexins are also produced at low levels: for example, 153

grams of human umbilical cords yielded less than 0.004 mg of type XV collagen (135).

To compensate for low levels of expression, the multiplexin trimerization domain has

likely evolved as a fast-folding domain designed to efficiently locate and bind complementary chains in the dense protein milieu of the rER. Prior to my work on the crystal structure of the type XV trimerization domain, Boudko and colleagues presented the crystal structure of the type XVIII trimerization domain (177). In this thesis, the type XVIII model was used as an initial model for three dimensional homology modeling of the collagen XV trimerization domain. The final model was energy minimized and did not reveal any gross stereochemical discrepancies. The most notable difference between the XV and XVIII models was an increased volume of the trimeric hydrophobic core. L24 and L33 of every chain of the type XVIII trimerization domain were replaced by two phenylalanine residues. Additionally, V35 and V44 were substituted by the bulkier I35 and L44. The change in the trimeric hydrophobic core volume was hypothesized to prevent heterotrimerization of type XV and XVIII chains. However, the overall size of the type XV trimerization domain as solved in this thesis was very similar in size. Specifically, the bulkier side chains did not appear to alter the size of the trimeric hydrophobic core. Analysis of a heterotrimer model did not yield steric clashes that would preclude heterotrimer formation. Based on the current molecular models, a heterotrimeric molecule could possibly form.

To address this issue, initial experiments were performed to determine if type XV and XVIII heterotrimers can form *in vitro*. Type XV and XVIII collagens have different isoelectric points and elute from a Q column under slightly different conditions. Type XV and type XVIII proteins were denatured, mixed, refolded and run on a Q column. Theoretically, four different species could appear: a XV homotrimer, a XV:XV:XVIII heterotrimer, a XV:XVIII:XVIII heterotrimer and a XVIII homotrimer. Regrettably, the

experiment was not easily interpretable, since the elution peaks were broad and displayed significant overlap. A more elegant experiment could be performed using type XVIII collagen that has not been cleaved from the His-thrombin tag used for expression and purification. The tagged XVIII could be denatured and mixed with denatured type XV trimerization domain, refolded and run on a Ni^{2+} chelating column. In this scenario, a type XV trimer would flow through the column while heterotrimers and a His-Thrombin type XVIII homotrimer would bind specifically to the column. Analysis of the eluted proteins by SDS-PAGE should reveal distinct bands between type XV and His-Thrombin type XVIII proteins due to the disparate sizes of the molecules. This experiment has the disadvantage of being run under relatively high protein concentrations – recall that type XV and XVIII multiplexins form trimers at picomolar concentrations. A more biologically relevant experiment may require lower protein concentrations. To these ends, a fluorescence polarization assay could examine potential heterotrimerization. Denatured type XV trimerization domain could be coupled to a fluorophore, mixed with either unlabeled type XV trimerization domain or denatured His-Trombin type XVIII and then refolded. The His-Thrombin tag greatly increases the mass of the type XVIII chain, and incorporation into a heterotrimer would tumble at much slower speeds than a type XV homotrimer.

The low degree of sequence homology among the multiplexin trimerization domains of a variety of organisms is perhaps the most striking feature of this association domain. With such variability, the conservation of key hydrophobic residues appears to dictate formation of the monomer and the trimeric hydrophobic core. The greatest deviation from the pattern of hydrophobic residues is observed in non-mammalian

organisms. It would be of interest to solve the crystal structures of these domains to investigate the evolution of the structure as a whole and to examine the extent of sequence flexibility allowed for the formation of stable trimers. Additionally, point mutations of key residues, specifically the three conserved residues, could be made to examine their effect on trimerization. Systematic analysis of hydrophobic residues may vary the overall stability of the trimer. Optimization of hydrophobic residues may create an even more potent trimerization domain.

Both the type XV and XVIII trimerization domains are capable of refolding quickly and completely. Additionally, there is a noticeable change in fluorescence signal between unfolded and folded protein. This makes the multiplexin family of proteins an excellent choice for studies on the kinetics of refolding using a stopped-flow. The newest graduate student in the lab will pursue these studies as a portion of his dissertation work, which will provide a satisfying extension of the experiments presented here.

Further studies on the P3H1●CRTAP●CypB complex.

Extractions of P3H1 are enhanced by relatively high levels of Triton X-100, which may be a sign of possible membrane association. The nature and strength of membrane association can be analyzed by extraction studies. Preliminary experiments have provided some evidence of membrane association, although further experimentation will be necessary (Figure 2). Microsomal pellets were extracted with 0.1% Tween-20, at which point a fraction of the soluble protein and the pellet were saved. The pellet was then re-extracted using PBS supplemented with 1M NaCl, 100 mM $NaCO_3$ or 4M Urea. Following extraction, soluble protein fractions were saved. Protein samples were

analyzed by western blot using the 1C10 antibody. High salt, carbonate and Urea extractions have been shown to release peripheral membrane proteins (208). Extraction under normal conditions produced roughly equal amounts of protein in the soluble and insoluble fractions (lanes 4 and 5), indicating that a portion of P3H1 protein remains in the pellet. Re-extraction of the insoluble pellet did in fact yield additional P3H1 under conditions designed to release membrane associated proteins: high salt, lane 1; high pH, lane 2; denaturant, lane 3.

Membrane extraction experiments do not necessarily differentiate between membrane associated proteins and aggregated material. To address this issue, membrane flotation experiments using a discontinuous sucrose gradient could be performed. Due to their intrinsic buoyant density, membranes migrate out of 60% sucrose into 35% sucrose when subjected to ultracentrifugation. Large protein aggregates that may naturally pellet under extraction experiments will separate out into the high sucrose fraction.

Further analysis of P3H1 enzymatic activity and substrate specificity

P3H1 has been isolated and used in full length chick procollagen substrate enzyme assays (47). In these assays, it was shown that 1) P3H1 activity was not altered by the presence of CRTAP and CypB, 2) P3H1 activity requires prior 4-hydroxylation of the collagen substrate and 3) the overall K_M value for the full length substrate was similar to that of a previous report using partially purified enzyme (189). As more research has been performed on the location of 3Hyp sites in collagens, it has become apparent that there are multiple sites that occur in type I, II, III, IV and V collagen (13). A peptide assay has been modified to examine 3-hydroxylation, which allows for determination of

substrate specificity (65). However, only P3H2 was used in this assay as enzymatically active P3H1 was not obtained with the insect cell expression systems used by the Myllyharju laboratory. A similar peptide assay could be developed using peptides designed to carry known sites of 3-hydroxylation present in chick collagens. This would allow for a more detailed analysis of P3H1 substrate specificity and kinetic parameters.

Additionally, substrate affinity can be examined by use of surface plasmon resonance (SPR) experiments, in which the peptides developed above are coupled to biosensor chips and real time peptide:P3H1 interactions are monitored as a function of refractive index change. SPR experiments have the advantage of determining kinetics of association and dissociation as well as a the use of label free proteins. In this thesis, SPR has been used to determine the affinity of P3H1●CRTAP●CypB for full length folded collagen substrates. Analysis of P3H1●CRTAP●CypB binding affinity to full length, denatured collagen substrates is difficult since denatured collagen may fold on the chip surface over time. The peptide experiment proposed above has the advantage of providing a single chain substrate for binding experiments.

Analysis of CRTAP and CypB interactions with P3H1 on enzymatic activity and substrate specificity

It is a matter of great interest in our lab to determine the exact affinities of the individual complex members for each other. These experiments are difficult since recombinant, biologically active P3H1 was not produced in significant amounts. CRTAP has yet to be isolated as an individual protein from chick embryos. Attempts to separate CRTAP from the P3H1 complex under a variety of conditions have resulted in complete

precipitation and loss of protein. As with P3H1, CRTAP was not successfully over

expressed in a variety of experiments. Due to these difficulties, examination of the

binding affinities between members of the complex is not possible at this time.

Identification of the interaction domains, on the other hand, may be possible.

Complex purified from chick embryos could be tryptically digested in the presence or

absence of crosslinkers. The change in tryptic fragment population would result from

crosslinks that are formed from close contact between proteins. Crosslinkers of different

link size would be optimized for these experiments to best capture protein:protein

interactions. Mass spectrometry analysis of the crosslinked fragments could identify the

domains involved in protein:protein interactions (209). This would provide insight into

the location of interactions between members of the complex. Recombinant GST-CypB

binding to collagen has been shown to be greatly dependant on P3H1 or CRTAP,

although addition of cyclosporine A did not significantly affect the ability of GST-CypB

to bind collagen (79). This indicates that the active site of CypB is not involved in

binding interactions with P3H1 or CRTAP.

P3H homologues

The N-terminal domain of P3H1 is found in only five proteins: P3H1, P3H2,

P3H3, CRTAP and SC65. The studies presented here focus primarily on P3H1 – much

work remains to be performed on the remaining homologues. Unfortunately, the

difficulties encountered in these studies on P3H1 appear to be consistent between all

family members. As with P3H1, both P3H2 and P3H3 were expressed in inclusion

bodies in *E. coli*. Others had marginal success expressing P3H2 in insect cells, although

P3H3 was entirely insoluble (65). CRTAP was not expressed in either Sf9 or mammalian cells in our laboratory. Attempts to clone SC65 were also unsuccessful.

A variety of studies using P3H1, CRTAP or CypB null or knock-down cell lines have demonstrated that a) P3H1 and CRTAP are mutually stabilizing, b) CRTAP maintains cellular levels of P3H1 and c) P3H1 and CRTAP facilitate CypB binding to type I collagen (77; 79). Therefore it is logical to hypothesize that either P3H1 and CRTAP co-expression or P3H1, CRTAP and CypB triple-expression would facilitate the expression of stable protein. Consistent with other reports (65), our attempts at co- or triple-expression in both bacterial and mammalian cell systems did not increase protein solubility or yield. The aforementioned studies have demonstrated the dependence of P3H1 and CRTAP stability on cell systems that maintained a normal complement of other proteins found in the rER. Under these conditions, P3H1 and CRTAP levels were found to be constant over 24 hours when protein translation was inhibited with cycloheximide, indicating that P3H1 and CRTAP are significantly more stable in the cellular milieu than as isolated proteins in solution. The stability of the complex may, in part, depend on other proteins found in the rER. If numerous proteins are necessary for recombinant P3H1 stability, this further complicates the ability of over expression systems to successfully produce protein. Even if significant amounts of soluble protein could be produced recombinantly in a system simultaneously expressing many proteins found in the rER, separation of the individual components may very well result in swift precipitation of P3H1.

Since over expression is not currently a viable option, investigations of these proteins will have to focus on extraction from biological tissues. P3H1 and CRTAP can,

as demonstrated by work in the Bächinger lab and in this thesis, be successfully purified from chick embryos. Although P3H2 and P3H3 RNA transcripts and protein have been detected in a variety of cells and tissues, endogenous P3H2 and P3H3 have yet to be purified from these systems. P3H2 is found primarily associated with basement membranes (65) and has been shown to preferentially hydroxylate a collagen peptide derived from type IV collagen (65). With this in mind, preliminary attempts to purify P3H2 and/or a P3H2 complex were performed using bovine kidney cortex as a tissue source. Kidney is rich in type IV collagen, and was hypothesized to selectively enrich P3H2. Sadly, only bovine P3H1 was found in the gelatin Sepharose elution fractions, with no trace of P3H2. Others in the lab have also tried to purify P3H2 from chick embryos using a type IV Sepharose column. In this instance, only chick P3H1 was found. P3H3 is expressed in the brain at higher levels than P3H1 or P3H2. Extraction from chick heads or bovine brains did not produce either chick or bovine P3H3.

The experiments presented above were performed on a small scale, and it may be necessary to increase the volume of extracted tissue to locate P3H2 and P3H3. Antibodies against P3H2 and P3H3 have been developed: columns generated with these antibodies could be used against proteins eluted from a gelatin Sepharose column in attempts to concentrate trace P3H2 or P3H3 that may be found in chick embryos. If P3H1 contamination remains high, the proteins eluted form the gelatin Sepharose column could be run over a P3H1 column prior to P3H2 or P3H3 antibody columns.

Conclusions

The work presented here has demonstrated the following main points:

1) The trimerization domain of type XV collagen displays strong three-dimensional homology to the type XVIII trimerization domain and is extremely stable. The multiplexin trimerization domain fold may be an evolutionarily conserved motif where extreme sequence variability can be accommodated provided a pattern of hydrophobic residues is conserved.

2) The P3H1•CRTAP•CypB complex is a molecular chaperone and PPIase as well as a 3-hydroxylase. In addition to interacting with denatured collagen, the complex is capable of binding to folded type I collagen with an affinity of ~7.5 μM.

3) P3H1 and the P3H1•CRTAP•CypB complex proteins are difficult to over express. Recombinant protein is unstable and significant quantities are difficult to obtain. Extraction from tissues remains a more consistent source of protein, although large molecular weight aggregates still form in solution. Further work on protein purification may lead to increases in protein concentration that would allow for more biochemical analysis of P3H1 and the P3H1•CRTAP•CypB complex.

Figure 1. Sequence alignment of potential trimerization domains of different multiplexins. Multiplexins from human, mouse , chicken, zebrafish, drosophila, nematode, sea squirt and sea urchin were aligned using Vector NTI. Residue numbering is given for the human collagen type XV trimerization domain. Identical residues from all sequences are shown in white with black shading, conservative residues in black with gray shading, blocks of similar residues are shown in red, and weakly similar residues in blue. Residue numbering is given for the human collagen XV trimerization domain. Residues common to the formation of the hydrophobic core of the type XV and type XVIII trimerization domains are indicated with an asterisk (*) and the residue involved in the trimeric hydrophobic core of type XVIII collagen but not in type XV is indicated with a forwardslash (/). Residues forming the hydrophobic core of the type XV or type XVIII monomer are marked with a dot (●). GenBank™ accession numbers for the sequences are as follows: human collagen XV P39059, human collagen XVIII P39060, mouse collagen XV O35206, mouse collagen XVIII P39061, chicken collagen XV XP_418896, chicken collagen XVIII NP_989495, zebrafish collagen XVIII Q2PBM7, zebrafish collagen XV Q05H57, drosophila multiplexin ACD03747, nematode CLE-1 protein Q7JL30, sea squirt collagen XVIII homologue Q86SC8, sea urchin collagen XVIII homologue XP_791637.
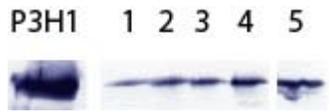
Figure 2. P3H1 may be membrane associated. Microsomal pellets were extracted with extraction buffer containing 0.1% Triton X-100. A fraction of the soluble portion was run in lane 4. A sample of the remaining pellet was loaded in lane 5. The remaining pellet was further extracted with extraction buffer supplemented with 100 mM $NaCO_3$, 1M NaCl or 4M Urea in lanes 1, 2 and 3, respectively.

References

1. Ricard-Blum S, Ruggiero F, van der Rest M. The Collagen Superfamily. In: Collagen. Topics in Current Chemistry, Springer Berlin. 2005 p.35-84.

2. Söderhäll C, Marenholz I, Kerscher T, Rüschendorf F, Esparza-Gordillo J, Worm M, Gruber C, Mayr G, Albrecht M, Rohde K, Schulz H, Wahn U, Hubner N, Lee Y. Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. PLoS Biol. 2007 Sep;5(9)

3. Veit G, Kobbe B, Keene DR, Paulsson M, Koch M, Wagener R. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J Biol Chem. 2006 Feb;281(6):3494-3504.

4. Pauling L, Corey RB. The structure of fibrous proteins of the collagen-gelatin group. Proc Natl Acad Sci USA.1951 May;37(5):272-281.

5. Ramachandran GN, Kartha G. Structure of collagen. Nature. 1954 Aug;174(4423):269-270.

6. Rich A, Crick FH. The structure of collagen. Nature. 1955 Nov;176(4489):915-916.

7. Engel J, Bächinger HP. Structure, Stability and Folding of the Collagen Triple Helix. In: Collagen. Topics in Current Chemistry, Springer Berlin. 2005 p. 7-33.

8. Shoulders MD, Raines RT. Collagen structure and stability. Annu Rev Biochem. 2009;78:929-958.

9. Hodges JA, Raines RT. Stereoelectronic effects on collagen stability: the dichotomy of 4-fluoroproline diastereomers. J Am Chem Soc. 2003;125(31):9262-9263.

10. Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet. 2004 Jan;20(1):33-43.

11. Koide T, Nagata K. Collagen Biosynthesis. In: Collagen. Topics in Current Chemistry, Springer Berlin. 2005 p. 85-114.

12. Myllyharju J. Intracellular Post-Translational Modifications of Collagens. In: Collagen. Topics in Current Chemistry, Springer Berlin. 2005 p. 115-147.

13. Weis MA, Hudson DM, Kim L, Scott M, Wu J, Eyre DR. Location of 3-Hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. J Biol Chem. 2010 Jan;285(4):2580-2590.

14. Fietzek PP, Wendt P, Kell I, Kühn K. The covalent structure of collagen: amino acid sequence of 1-CB3 from calf skin collagen. FEBS Lett. 1972 Oct;26(1):74-76.

15. Fietzek PP, Rexrodt FW, Hopper KE, Kühn K. The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 from calf-skin collagen. Eur J Biochem. 1973 Oct;38(2):396-400.

16. Rexrodt FW, Fietzek PP, Kühn K. The covalent structure of collagen. The chymotrypsin, trypsin and hydroxylamine peptides derived from alpha2-CB4 of calf-skin collagen. Eur J Biochem. 1975 Nov;59(1):105-112.

17. Miller EJ, Martin GR, Piez KA, Powers MJ. Characterization of chick bone collagen and compositional changes associated with maturation. J Biol Chem. 1967 Dec;242(23):5481-5489.

18. Miller EJ. Isolation and characterization of a collagen from chick cartilage containing three identical alpha chains. Biochemistry. 1971 Apr;10(9):1652-1659.

19. Rhodes RK, Miller EJ. The isolation and characterization of the cyanogen bromide peptides from the B chain of human collagen. J Biol Chem. 1979 Dec;254(23):12084-12087.

20. Gryder RM, Lamon M, Adams E. Sequence position of 3-hydroxyproline in basement membrane collagen. Isolation of glycyl-3-hydroxyprolyl-4-hydroxyproline from swine kidney. J Biol Chem. 1975 Apr;250(7):2470-2474.

21. Kresina TF, Miller EJ. Isolation and characterization of basement membrane collagen from human placental tissue. Evidence for the presence of two genetically distinct collagen chains. Biochemistry. 1979 Jul;18(14):3089-3097.

22. Gay S, Miller EJ. Characterization of lens capsule collagen: evidence for the presence of two unique chains in molecules derived from major basement membrane structures. Arch Biochem Biophys. 1979 Dec;198(2):370-378.

23. Kleinman HK, McGarvey ML, Liotta LA, Robey PG, Tryggvason K, Martin GR. Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. Biochemistry. 1982 Nov;21(24):6188-6193.

24. Kefalides NA. Isolation and characterization of cyanogen bromide peptides from basement membrane collagen. Biochem Biophys Res Commun. 1972 Jun;47(5):1151-1158.

25. Sage H, Woodbury RG, Bornstein P. Structural studies on human type IV collagen. J Biol Chem. 1979 Oct;254(19):9893-9900.

26. Schumacher MA, Mizuno K, Bächinger HP. The Crystal Structure of a collagen-like polypeptide with 3(S)-Hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. J Biol Chem. 2006;281(37):27566-27574.

27. Mizuno K, Peyton DH, Hayashi T, Engel J, Bächinger HP. Effect of the -Gly-3(S)-hydroxyprolyl-4(R)-hydroxyprolyl- tripeptide unit on the stability of collagen model peptides. FEBS J. 2008 Dec;275(23):5830-5840.

28. Saito J, Imamura Y, Itoh J, Matsuyama S, Maruta A, Hayashi T, Sato A, Wada N, Kashiwazaki K, Inagaki Y, Watanabe T, Kitagawa Y, Okazaki I. ELISA measurement for urinary 3-hydroxyproline-containing peptides and its preliminary application to healthy persons and cancer patients. Anticancer Res. 2010 Mar;30(3):1007-1014.

29. Kivirikko KI, Myllyharju J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. Matrix Biol. 1998 Feb;16(7):357-368.

30. Kukkola L, Hieta R, Kivirikko KI, Myllyharju J. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-Hydroxylase isoenzyme. J Biol Chem. 2003 Nov;278(48):47685-47693.

31. Myllyharju J. HIF prolyl 4-hydroxylases and their potential as drug targets. Curr Pharm Des. 2009;15(33):3878-3885.

32. Bulleid NJ, Wilson R, Lees JF. Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation. Biochem J. 1996 Jul;317(Pt 1):195-202.

33. Beck K, Boswell BA, Ridgway CC, Bächinger HP. Triple Helix Formation of Procollagen Type I Can Occur at the Rough Endoplasmic Reticulum Membrane. J Biol Chem. 1996;271(35):21566-21573.

34. Engel J, Prockop DJ. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. Annu Rev Biophys Biophys Chem. 1991;20:137-152.

35. Bächinger HP, Bruckner P, Timpl R, Engel J. The role of cis-trans isomerization of peptide bonds in the coil leads to and comes from triple helix conversion of collagen. Eur J Biochem. 1978 Oct;90(3):605-613.

36. Bächinger HP, Bruckner P, Timpl R, Prockop DJ, Engel J. Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. Eur. J Biochem. 1980 May;106(2):619-632.

37. Harrison RK, Stein RL. Mechanistic studies of peptidyl prolyl cis-trans isomerase: evidence for catalysis by distortion. Biochemistry. 1990 Feb;29(7):1684-1689.

38. Coss MC, Winterstein D, Raymond CS, Simek SL. Molecular Cloning, DNA Sequence Analysis, and Biochemical Characterization of a Novel 65-kDa FK506-binding Protein (FKBP65). J Biol Chem. 1995 Dec;270(49):29336-29341.

39. Steinmann B, Bruckner P, Superti-Furga A. Cyclosporin A slows collagen triple-helix formation in vivo: indirect evidence for a physiologic role of peptidyl-prolyl cis-trans-isomerase. J Biol Chem. 1991 Jan;266(2):1299-1303.

40. Macdonald JR, Bächinger HP. HSP47 binds cooperatively to triple helical type I collagen but has little effect on the thermal stability or rate of refolding. J Biol Chem. 2001 Jul;276(27):25399-25403.

41. Ishikawa Y, Vranka J, Wirz J, Nagata K, Bächinger HP. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 is a molecular chaperone that interacts with collagens. J Biol Chem. 2008 Nov;283(46):31584-31590.

42. Ishikawa Y, Wirz J, Vranka JA, Nagata K, Bächinger HP. Biochemical characterization of the prolyl 3-hydroxylase 1.cartilage-associated protein.cyclophilin B complex. J Biol Chem. 2009 Jun;284(26):17641-17647.

43. Bonfanti L, Mironov AA, J A Martínez-Menárguez, Martella O, Fusella A, Baldassarre M, Buccione R, Geuze HJ, Mironov AA, Luini A. Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. Cell. 1998 Dec;95(7):993-1003.

44. Greenspan DS. Biosynthetic Processing of Collagen Molecules. In: Collagen. Topics in Current Chemistry. 2005 p. 149-183.

45. Canty EG, Kadler KE. Procollagen trafficking, processing and fibrillogenesis. J Cell Sci. 2005 Apr;118(7):1341-1353.

46. Stephens DJ, Pepperkok R. Imaging of procollagen transport reveals COPI-dependent cargo sorting during ER-to-Golgi transport in mammalian cells. J Cell Sci. 2002 Mar;115(6):1149-1160.

47. Vranka JA, Sakai LY, Bächinger HP. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. J Biol Chem. 2004 May;279(22):23615-23621.

48. Risteli J, Tryggvason K, Kivirikko KI. Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. Eur J Biochem. 1977 Mar;73(2):485-492.

49. Tryggvason K, Risteli J, Kivirikko KI. Separation of prolyl 3-hydroxylase and 4-hydroxylase activities and the 4-hydroxyproline requirement for synthesis of 3-hydroxyproline. Biochem Biophys Res Commun. 1976 May;76(2):275-281.

50. Tryggvason K, Majamaa K, Risteli J, Kivirikko KI. Partial purification and characterization of chick-embryo prolyl 3-hydroxylase. Biochem J. 1979 Nov;183(2):303-307.

51. Zeng B, MacDonald JR, Bann JG, Beck K, Gambee JE, Boswell BA, Bächinger HP. Chicken FK506-binding protein, FKBP65, a member of the FKBP family of peptidylprolyl cis-trans isomerases, is only partially inhibited by FK506. Biochem J. 1998 Feb;330 (Pt 1):109-114.

52. Wassenhove-McCarthy DJ, McCarthy KJ. Molecular characterization of a novel basement membrane-associated proteoglycan, leprecan. J Biol Chem. 1999;274(35):25004-25017.

53. Kaul SC, Sugihara T, Yoshida A, Nomura H, Wadhwa R. Gros1, a potential growth suppressor on chromosome 1: its identity to basement membrane-associated proteoglycan, leprecan. Oncogene. 2000 Jul;19(32):3576-3583.

54. Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol. 2001;2(3):Research0007.1-Research0007.8.

55. Loenarz C, Schofield CJ. Expanding chemical biology of 2-oxoglutarate oxygenases. Nat Chem Biol. 2008 Mar;4(3):152-156.

56. Blatch GL, Lässle M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays. 1999 Nov;21(11):932-939.

57. Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M, Castagnola P, Rauch F, Glorieux FH, Vranka J, Bächinger HP, Pace JM, Schwarze U, Byers PH, Weis M, Fernandes RJ, Eyre DR, Yao Z, Boyce BF, Lee B. CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. Cell. 2006 Oct;127(2):291-304.

58. Baldridge D, Schwarze U, Morello R, Lennington J, Bertin TK, Pace JM, Pepin MG, Weis M, Eyre DR, Walsh J, Lambert D, Green A, Robinson H, Michelson M, Houge G, Lindman C, Martin J, Ward J, Lemyre E, Mitchell JJ, Krakow D, Rimoin DL, Cohn DH, Byers PH, Lee B. CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. Hum Mutat. 2008 Dec;29(12):1435-1442.

59. Willaert A, Malfait F, Symoens S, Gevaert K, Kayserili H, Megarbane A, Mortier G, Leroy JG, Coucke PJ, De Paepe A. Recessive osteogenesis imperfecta caused by LEPRE1 mutations: clinical documentation and identification of the splice form responsible for prolyl 3-hydroxylation. J Med Genet. 2009 Apr;46(4):233-241.

60. Barnes AM, Chang W, Morello R, Cabral WA, Weis M, Eyre DR, Leikin S, Makareeva E, Kuznetsova N, Uveges TE, Ashok A, Flor AW, Mulvihill JJ, Wilson PL, Sundaram UT, Lee B, Marini JC. Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. N Engl J Med. 2006 Dec;355(26):2757-2764.

61. Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, Leikin S, Makareeva E, Kuznetsova NV, Rosenbaum KN, Tifft CJ, Bulas DI, Kozma C, Smith PA, Eyre DR, Marini JC. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet. 2007 Mar;39(3):359-365.

62. van Dijk FS, Nesbitt IM, Zwikstra EH, Nikkels PGJ, Piersma SR, Fratantoni SA, Jimenez CR, Huizer M, Morsman AC, Cobben JM, van Roij MHH, Elting MW, Verbeke JIML, Wijnaendts LCD, Shaw NJ, Högler W, McKeown C, Sistermans EA, Dalton A, Meijers-Heijboer H, Pals G. PPIB mutations cause severe osteogenesis imperfecta. Am J Hum Genet. 2009 Oct;85(4):521-527.

63. Järnum S, Kjellman C, Darabi A, Nilsson I, Edvardsen K, Åman P. LEPREL1, a novel ER and Golgi resident member of the Leprecan family. Biochem Biophys Res Commun. 2004 Apr;317(2):342-351.

64. Vranka J, Stadler HS, Bächinger HP. Expression of prolyl 3-hydroxylase genes in embryonic and adult mouse tissues. Cell Struct Funct. 2009;34(2):97-104.

65. Tiainen P, Pasanen A, Sormunen R, Myllyharju J. Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. J Biol Chem. 2008 Jul;283(28):19432-19439.

66. Vranka JA, Pokidysheva E, Hayashi L, Zientek K, Mizuno K, Ishikawa Y, Maddox K, Tufa S, Keene DR, Klein R, Bachinger HP. Prolyl 3-hydroxylase 1 null mice display abnormalities in fibrillar collagen-rich tissues such as tendons, skin and bones. J Biol Chem 2010 May 28;285(22):17253-62.

67. Sillence DO, Senn A, Danks DM. Genetic heterogeneity in osteogenesis imperfecta. J Med Genet. 1979 Apr;16(2):101-116.

68. Rauch F, Glorieux FH. Osteogenesis imperfecta. The Lancet. 2004 Apr;363(9418):1377-1385.

69. Pope FM, Nicholls AC, McPheat J, Talmud P, Owen R. Collagen genes and proteins in osteogenesis imperfecta. J Med Genet. 1985 Dec;22(6):466-478.

70. Cheung MS, Glorieux FH. Osteogenesis Imperfecta: update on presentation and management. Rev Endocr Metab Disord. 2008 Jun;9(2):153-160.

71. Johanna Myllyharju, Kari I Kivirikko. Collagens and collagen-related diseases. Ann Med. 2001 Feb;33(1):7-21.

72. Van Dijk FS, Nesbitt IM, Nikkels PGJ, Dalton A, Bongers EMHF, van de Kamp JM, Hilhorst-Hofstee Y, Den Hollander NS, Lachmeijer AMA, Marcelis CL, Tan-Sindhunata GMB, van Rijn RR, Meijers-Heijboer H, Cobben JM, Pals G. CRTAP mutations in lethal and severe osteogenesis imperfecta: the importance of combining biochemical and molecular genetic analysis. Eur J Hum Genet. 2009 Dec;17(12):1560-9.

73. Barnes A, Chang W, Morello R, Cabral W, Weis M, Eyre D, Leikin S, Mulvihill J, Lee B, Marini J. Recessive lethal form of OI caused by null mutations in CRTAP. Matrix Biol. 2006 Nov;25:S61-S62.

74. Cabral W, Chang W, Barnes A, Eyre D, Weis M, Leikin S, Makareeva E, Kuznetsova N, Bulas D, Marini J. Null mutations of P3H1 cause recessive OI-like bone dysplasia. Matrix Biol. 2006 Nov;25:S62.

75. Shah R, Smith P, Purdie C, Quinlan P, Baker L, Aman P, Thompson AM, Crook T. The prolyl 3-hydroxylases P3H2 and P3H3 are novel targets for epigenetic silencing in breast cancer. Br J Cancer. 2009 May;100(10):1687-1696.

76. Baldridge D, Lennington J, Weis M, Homan EP, Jiang M, Munivez E, Keene DR, Hogue WR, Pyott S, Byers PH, Krakow D, Cohn DH, Eyre DR, Lee B, Morello R. Generalized connective tissue disease in Crtap-/- Mouse. PLoS ONE. 2010 May;5(5):e10560.

77. Choi JW, Sutor SL, Lindquist L, Evans GL, Madden BJ, Bergen HR, Hefferan TE, Yaszemski MJ, Bram RJ. Severe osteogenesis imperfecta in cyclophilin B-deficient mice. PLoS Genet. 2009 Dec;5(12):e1000750.

78. Barnes AM, Carter EM, Cabral WA, Weis M, Chang W, Makareeva E, Leikin S, Rotimi CN, Eyre DR, Raggio CL, Marini JC. Lack of cyclophilin B in osteogenesis imperfecta with normal collagen folding. N Engl J Med 2010 Feb;362(6):521-528.

79. Chang W, Barnes AM, Cabral WA, Bodurtha JN, Marini JC. Prolyl 3-hydroxylase 1 and CRTAP are mutually stabilizing in the endoplasmic reticulum collagen prolyl 3-hydroxylation complex. Hum Mol Genet. 2010 Jan;19(2):223-234.

80. Khoshnoodi J, Cartailler J, Alvares K, Veis A, Hudson B. Molecular recognition in the assembly of collagens: Terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. J Biol Chem. 2006;281(50):38117-38121.

81. McLaughlin SH, Bulleid NJ. Molecular recognition in procollagen chain assembly. Matrix Biol. 1998 Feb;16(7):369-377.

82. Doege KJ, Fessler JH. Folding of carboxyl domain and assembly of procollagen I. J Biol Chem. 1986 Jul;261(19):8924-8935.

83. Dion AS, Myers JC. COOH-terminal propeptides of the major human procollagens. Structural, functional and genetic comparisons. J Mol Biol. 1987 Jan;193(1):127-143.

84. Koivu J. Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen. FEBS Lett. 1987 Feb;212(2):229-232.

85. Risteli J, Niemi S, Kauppila S, Melkko J, Risteli L. Collagen propeptides as indicators of collagen assembly. Acta Orthop Scand Suppl. 1995 Oct;266:183-8.

86. Lim A, Doyle SA, Balian G, Smith BD. Role of the pro-α2(I) COOH-terminal region in assembly of type I collagen: Truncation of the last 10 amino acid residues of pro- α2(I) chain prevents assembly of type I collagen heterotrimer. J Cell Biochem. 1998;71(2):216-232.

87. Bernocco S, Finet S, Ebel C, Eichenberger D, Mazzorana M, Farjanel J, Hulmes DJS. Biophysical characterization of the C-propeptide trimer from human procollagen III reveals a tri-lobed structure. J Biol Chem. 2001 Dec;276(52):48930-48936.

88. Hulmes DJS. Building collagen molecules, fibrils, and suprafibrillar structures. J Struct Biol. 2002 Feb;137(1-2):2-10.

89. Malone JP, Alvares K, Veis A. Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of type I collagen: Significance of the α2(I) chain. Biochemistry. 2005 Nov;44(46):15269-15279.

90. McAlinden A, Smith T, Sandell L, Ficheux D, Parry D, Hulmes D. α-helical coiled-coil oligomerization domains are almost ubiquitous in the collagen superfamily. J Biol Chem. 2003;278(43):42200-42207.

91. Than ME, Henrich S, Huber R, Ries A, Mann K, Kühn K, Timpl R, Bourenkov GP, Bartunik HD, Bode W. The 1.9-Å crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. Proc Natl Acad Sci USA. 2002 May;99(10):6607-6612.

92. Sundaramoorthy M, Meiyappan M, Todd P, Hudson BG. Crystal Structure of NC1 Domains. J Biol Chem. 2002;277(34):31142-31153.

93. Bogin O, Kvansakul M, Rom E, Singer J, Yayon A, Hohenester E. Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. Structure. 2002;10(2):165-173.

94. Kvansakul M, Bogin O, Hohenester E, Yayon A. Crystal structure of the collagen alpha1(VIII) NC1 trimer. Matrix Biol. 2003 Apr;22(2):145-152.

95. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG. Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med. 2003 Jun;348(25):2543-2556.

96. Bailey AJ, Sims TJ, Light N. Cross-linking in type IV collagen. Biochem. J 1984 Mar;218(3):713-723.

97. Brass A, Kadler KE, Thomas JT, Grant ME, Boot-Handford RP. The fibrillar collagens, collagen VIII, collagen X and the C1q complement proteins share a similar domain in their C-terminal non-collagenous regions. FEBS Lett. 1992 Jun;303(2-3):126-128.

98. Frischholz S, Beier F, Girkontaite I, Wagner K, Pöschl E, Turnay J, Mayer U, von der Mark K. Characterization of human type X procollagen and its NC-1 domain expressed as recombinant proteins in HEK293 cells. J Biol Chem. 1998 Feb;273(8):4547-4555.

99. Zhang Y, Chen Q. The noncollagenous domain 1 of type X collagen. J Biol Chem. 1999;274(32):22409-22413.

100. Boudko SP, Zientek KD, Vance JM, Hacker JL, Engel J, Bachinger HP. The NC2 domain of collagen IX provides chain selection and heterotrimerization. J Biol Chem 2010 Jul 30;285(31):23721-31

101. Boudko SP, Engel J, Bächinger HP. Trimerization and Triple Helix Stabilization of the Collagen XIX NC2 Domain. J Biol Chem. 2008 Dec;283(49):34345-34351.

102. Shaw LM, Olsen BR. FACIT collagens: diverse molecular bridges in extracellular matrices. Trends Biochem Sci. 1991 May;16(5):191-194.

103. Jäälinoja J, Ylöstalo J, Beckett W, Hulmes D, Ala-Kokko L. Trimerization of collagen IX α-chains does not require the presence of the COL1 and NC1 domains. Biochem. J. 2008;409(2):545.

104. Snellman A, Tu H, Väisänen T, Kvist AP, Huhtala P, Pihlajaniemi T. A short sequence in the N-terminal region is required for the trimerization of type XIII collagen and is conserved in other collagenous transmembrane proteins. EMBO J. 2000 Oct;19(19):5051-5059.

105. Rehn M, Pihlajaniemi T. Type XV and XVIII collagens, a new subgroup within the family of collagens. Semin Cell Dev Biol. 1996 Oct;7(5):673-679.

106. Muragaki Y, Abe N, Ninomiya Y, Olsen BR, Ooshima A. The human alpha 1(XV) collagen chain contains a large amino-terminal non-triple helical domain with a tandem repeat structure and homology to alpha 1(XVIII) collagen. J Biol Chem. 1994 Feb;269(6):4042-4046.

107. Kivirikko S, Heinämäki P, Rehn M, Honkanen N, Myers JC, Pihlajaniemi T. Primary structure of the alpha 1 chain of human type XV collagen and exon-intron organization in the 3' region of the corresponding gene. J Biol Chem. 1994 Feb;269(7):4773-4779.

108. Oh SP, Warman ML, Seldin MF, Cheng S, Knoll JHM, Timmons S, Olsen BR. Cloning of cDNA and Genomic DNA Encoding Human Type XVIII Collagen and Localization of the [alpha]1(XVIII) Collagen Gene to Mouse Chromosome 10 and Human Chromosome 21. Genomics. 1994 Feb;19(3):494-499.

109. Saarela J, Ylikärppä R, Rehn M, Purmonen S, Pihlajaniemi T. Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. Matrix Biol. 1998 Jan;16(6):319-328.

110. Oh SP, Kamagata Y, Muragaki Y, Timmons S, Ooshima A, Olsen BR. Isolation and sequencing of

cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins. Proc Natl Acad Sci USA. 1994 May;91(10):4229-4233.

111. Pihlajaniemi T, Rehn M. Two new collagen subgroups: membrane-associated collagens and types XV and XVII. Prog Nucleic Acid Res. Mol. Biol 1995;50:225-262.

112. Rehn M, Hintikka E, Pihlajaniemi T. Primary structure of the α1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain. J Biol Chem. 1994 May;269(19):13929-13935.

113. Rehn M, Hintikka E, Pihlajaniemi T. Characterization of the mouse gene for the α1 chain of type XVIII collagen (Col18a1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters. Genomics. 1996 Mar;32(3):436-446.

114. Hägg PM, Muona A, Liétard J, Kivirikko S, Pihlajaniemi T. Complete exon-intron organization of the human gene for the α1 chain of type XV collagen (COL15A1) and comparison with the homologous COL18A1 gene. J Biol Chem. 1998 Jul;273(28):17824-17831.

115. Boot-Handford RP, Tuckwell DS, Plumb DA, Rock CF, Poulsom R. A novel and highly conserved collagen (proα1(XXVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. J Biol Chem. 2003;278(33):31067-31077.

116. Bork P. The modular architecture of vertebrate collagens. FEBS Lett. 1992 Jul;307(1):49-54.

117. Sasaki T, Fukai N, Mann K, Göhring W, Olsen B, Timpl R. Structure, function and tissue forms of the C-terminal globular domain of collagen XVIII containing the angiogenesis inhibitor endostatin. EMBO J. 1998;17(15):4249-4256.

118. Sasaki T, Hohenester E, Timpl R. Structure and function of collagen-derived endostatin inhibitors of angiogenesis. IUBMB Life. 2002 Feb;53(2):77-84.

119. Sasaki T, Larsson H, Tisi D, Claesson-Welsh L, Hohenester E, Timpl R. Endostatins derived from collagens XV and XVIII differ in structural and binding properties, tissue distribution and anti-angiogenic activity. J Mol Biol. 2000 Sep;301(5):1179-1190.

120. Boehm T, Folkman J, Browder T, O'Reilly MS. Antiangiogenic therapy of experimental cancer does not induce acquired drug resistance. Nature. 1997 Nov;390(6658):404-407.

121. Folkman J. Antiangiogenesis in cancer therapy - endostatin and its mechanisms of action. Exp Cell Res. 2006 Mar;312(5):594-607.

122. Ferreras M, Felbor U, Lenhard T, Olsen BR, Delaissé J. Generation and degradation of human endostatin proteins by various proteinases. FEBS Lett. 2000 Dec;486(3):247-251.

123. Digtyar AV, Pozdnyakova NV, Feldman NB, Lutsenko SV, Severin SE. Endostatin: current concepts about its biological role and mechanisms of action. Biochemistry Mosc. 2007 Mar;72(3):235-246.

124. Hohenester E, Sasaki T, Olsen BR, Timpl R. Crystal structure of the angiogenesis inhibitor endostatin at 1.5 A resolution. EMBO J. 1998 Mar;17(6):1656-1664.

125. Ding Y, Javaherian K, Lo K, Chopra R, Boehm T, Lanciotti J, Harris BA, Li Y, Shapiro R, Hohenester E, Timpl R, Folkman J, Wiley DC. Zinc-dependent dimers observed in crystals of human endostatin. Proc Natl Acad Sci USA. 1998 Sep;95(18):10443-10448.

126. Myers JC, Kivirikko S, Gordon MK, Dion AS, Pihlajaniemi T. Identification of a previously

unknown human collagen chain, α1(XV), characterized by extensive interruptions in the triple-helical region. Proc Natl Acad Sci USA. 1992 Nov;89(21):10144-10148.

127. Li D, Clark CC, Myers JC. Basement membrane zone type XV collagen is a disulfide-bonded chondroitin sulfate proteoglycan in human tissues and cultured cells. J Biol Chem. 2000 Jul;275(29):22339-22347.

128. Amenta PS, Scivoletti NA, Newman MD, Sciancalepore JP, Li D, Myers JC. Proteoglycan-collagen XV in human tissues is seen linking banded collagen fibers subjacent to the basement membrane. J. Histochem. Cytochem. 2005 Feb;53(2):165-176.

129. Kivirikko S, Saarela J, Myers JC, Autio-Harmainen H, Pihlajaniemi T. Distribution of type XV collagen transcripts in human tissue and their production by muscle cells and fibroblasts. Am J Pathol. 1995 Nov;147(5):1500-1509.

130. Myers JC, Dion AS, Abraham V, Amenta PS. Type XV collagen exhibits a widespread distribution in human tissues but a distinct localization in basement membrane zones. Cell Tissue Res. 1996 Dec;286(3):493-505.

131. Hägg PM, Hägg PO, Peltonen S, Autio-Harmainen H, Pihlajaniemi T. Location of type XV collagen in human tissues and its accumulation in the interstitial matrix of the fibrotic kidney. Am J Pathol. 1997 Jun;150(6):2075-2086.

132. Tomono Y, Naito I, Ando K, Yonezawa T, Sado Y, Hirakawa S, Arata J, Okigaki T, Ninomiya Y. Epitope-defined monoclonal antibodies against multiplexin collagens demonstrate that type XV and XVIII collagens are expressed in specialized basement membranes. Cell Struct. Funct. 2002 Feb;27(1):9-20.

133. Amenta PS, Briggs K, Xu K, Gamboa E, Jukkola AF, Li D, Myers JC. Type XV collagen in human colonic adenocarcinomas has a different distribution than other basement membrane zone proteins. Hum Pathol. 2000 Mar;31(3):359-366.

134. Quarmby S, West C, Magee B, Stewart A, Hunter R, Kumar S. Differential expression of cytokine genes in fibroblasts derived from skin biopsies of patients who developed minimal or severe normal tissue damage after radiotherapy. Radiat Res. 2002 Mar;157(3):243-248.

135. Myers J, Amenta P, Dion A, Sciancalepore J, Nagaswami C, Weisel J, Yurchenco P. The molecular structure of human tissue type XV presents a unique conformation among the collagens. Biochem J. 2007;404(3):535-544.

136. Eklund L, Piuhola J, Komulainen J, Sormunen R, Ongvarrasopone C, Fássler R, Muona A, Ilves M, Ruskoaho H, Takala TE, Pihlajaniemi T. Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. Proc Natl Acad Sci USA. 2001 Jan;98(3):1194-1199.

137. John H, Preissner KT, Forssmann W, Standker L. Novel glycosylated forms of human plasma endostatin and circulating endostatin-related fragments of collagen XV. Biochemistry. 1999;38(32):10217-10224.

138. Ramchandran R, Dhanabal M, Volk R, Waterman MJ, Segal M, Lu H, Knebelmann B, Sukhatme VP. Antiangiogenic activity of restin, NC10 domain of human collagen XV: comparison to endostatin. Biochem Biophys Res Commun. 1999 Feb;255(3):735-739.

139. Rehn M, Pihlajaniemi T. Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. Proc Natl Acad Sci USA. 1994 May;91(10):4234-4238.

140. Elamaa H, Snellman A, Rehn M, Autio-Harmainen H, Pihlajaniemi T. Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. Matrix Biol. 2003 Sep;22(5):427-442.

141. Dong S, Cole GJ, Halfter W. Expression of Collagen XVIII and localization of its glycosaminoglycan attachment sites. J Biol Chem. 2003 Jan;278(3):1700-1707.

142. Saarela J, Rehn M, Oikarinen A, Autio-Harmainen H, Pihlajaniemi T. The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol. 1998 Aug;153(2):611-626.

143. Sertié AL, Sossi V, Camargo AA, Zatz M, Brahe C, Passos-Bueno MR. Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). Hum Mol Genet. 2000 Aug;9(13):2051-2058.

144. Hurskainen M, Eklund L, Hägg PO, Fruttiger M, Sormunen R, Ilves M, Pihlajaniemi T. Abnormal maturation of the retinal vasculature in type XVIII collagen/endostatin deficient mice and changes in retinal glial cells due to lack of collagen types XV and XVIII. FASEB J. 2005 Sep;19(11):1564-1566.

145. Muragaki Y, Timmons S, Griffith CM, Oh SP, Fadel B, Quertermous T, Olsen BR. Mouse Col18a1 is expressed in a tissue-specific manner as three alternative variants and is localized in basement membrane zones. Proc Natl Acad Sci USA. 1995 Sep;92(19):8763-8767.

146. Fukai N, Eklund L, Marneros AG, Oh SP, Keene DR, Tamarkin L, Niemelä M, Ilves M, Li E, Pihlajaniemi T, Olsen BR. Lack of collagen XVIII/endostatin results in eye abnormalities. EMBO J. 2002 Apr;21(7):1535-1544.

147. Musso O, Rehn M, Saarela J, Théret N, Liétard J, Hintikka, Lotrian D, Campion JP, Pihlajaniemi T, Clément B. Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver. Hepatology. 1998 Jul;28(1):98-107.

148. Ylikärppä R, Eklund L, Sormunen R, Kontiola AI, Utriainen A, Määttä M, Fukai N, Olsen BR, Pihlajaniemi T. Lack of type XVIII collagen results in anterior ocular defects. FASEB J. 2003 Dec;17(15):2257-2259.

149. Marneros AG, Olsen BR. Physiological role of collagen XVIII and endostatin. FASEB J. 2005 May;19(7):716-728.

150. Suzuki OT, Sertié AL, Der Kaloustian VM, Kok F, Carpenter M, Murray J, Czeizel AE, Kliemann SE, Rosemberg S, Monteiro M, Olsen BR, Passos-Bueno MR. Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in knobloch syndrome. Am J Hum Genet. 2002 Dec;71(6):1320-1329.

151. Suzuki O, Kague E, Bagatini K, Tu H, Heljasvaara R, Carvalhaes L, Gava E, de Oliveira G, Godoi P, Oliva G, Kitten G, Pihlajaniemi T, Passos-Bueno M. Novel pathogenic mutations and skin biopsy analysis in Knobloch syndrome. Mol Vis. 2009;15:801-9.

152. O'Reilly MS. Angiostatin: an endogenous inhibitor of angiogenesis and of tumor growth. EXS. 1997;79:273-294.

153. Ständker L, Schrader M, Kanse SM, Jürgens M, Forssmann W, Preissner KT. Isolation and characterization of the circulating form of human endostatin. FEBS Lett. 1997 Dec;420(2-3):129-133.

154. Dhanabal M, Ramchandran R, Waterman MJF, Lu H, Knebelmann B, Segal M, Sukhatme VP. Endostatin induces endothelial cell apoptosis. J Biol Chem. 1999 Apr;274(17):11721-11726.

155. Yamaguchi N, Anand-Apte B, Lee M, Sasaki T, Fukai N, Shapiro R, Que I, Lowik C, Timpl R, Olsen BR. Endostatin inhibits VEGF-induced endothelial cell migration and tumor growth independently of zinc binding. EMBO J. 1999 Aug;18(16):4414-4423.

156. Pawliuk R, Bachelot T, Zurkiya O, Eriksson A, Cao Y, Leboulch P. Continuous intravascular secretion of endostatin in mice from transduced hematopoietic stem cells. Mol Ther. 2002 Apr;5(4):345-351.

157. Eisterer W, Jiang X, Bachelot T, Pawliuk R, Abramovich C, Leboulch P, Hogge D, Eaves C. Unfulfilled promise of endostatin in a gene therapy-xenotransplant model of human acute lymphocytic leukemia. Mol Ther. 2002 Apr;5(4):352-359.

158. Seppinen L, Sormunen R, Soini Y, Elamaa H, Heljasvaara R, Pihlajaniemi T. Lack of collagen XVIII accelerates cutaneous wound healing, while overexpression of its endostatin domain leads to delayed healing. Matrix Biol. 2008 Jul;27(6):535-546.

159. Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ, Kramer JM. The NC1/endostatin domain of caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance. J Cell Biol. 2001 Mar;152(6):1219-1232.

160. Wen W, Moses MA, Wiederschain D, Arbiser JL, Folkman J. The generation of endostatin is mediated by elastase. Cancer Res. 1999 Dec;59(24):6052-6056.

161. Felbor U, Dreier L, Bryant RA, Ploegh HL, Olsen BR, Mothes W. Secreted cathepsin L generates endostatin from collagen XVIII. EMBO J. 2000 Mar;19(6):1187-1194.

162. Lin H, Chang J, Jain S, Gabison EE, Kure T, Kato T, Fukai N, Azar DT. Matrilysin cleavage of corneal collagen type XVIII NC1 domain and generation of a 28-kDa fragment. Invest Ophthalmol Vis Sci. 2001 Oct;42(11):2517-2524.

163. Kuo CJ, LaMontagne KR, Garcia-Cardeña G, Ackley BD, Kalman D, Park S, Christofferson R, Kamihara J, Ding Y, Lo K, Gillies S, Folkman J, Mulligan RC, Javaherian K. Oligomerization-dependent regulation of motility and morphogenesis by the collagen XVIII NC1/endostatin domain. J Cell Biol. 2001 Mar;152(6):1233-1246.

164. Meyer F, Moussian B. Drosophila multiplexin (Dmp) modulates motor axon pathfinding accuracy. Dev. Growth Differ. 2009 Jun;51(5):483-498.

165. Rhodes G. Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models. Academic Press; 2006.

166. Jr CWC. Macromolecular Crystallography, Part C, Volume 368. 1st ed. Academic Press; 2003.

167. Drenth J. Principles of Protein X-Ray Crystallography. 3rd ed. Springer; 2006.

168. McRee DE. Practical Protein Crystallography, Second Edition. 2nd ed. Academic Press; 1999.

169. Bragg L. Nobel Lectures, Physics 1901 - 1921. Elsevier Publishing Company (out of print). Available from: http://nobelprize.org/obel_prizes/physics/laureates/1915/wl-bragg-lecture.html

170. Cruickshank D, Juretschke H, Kato N. P.P. Ewald and his Dynamical Theory of X-ray Diffraction. International Union of Crystallographers (out of print). Available from:

http://ukcatalogue.oup.com/product/9780198553793.do

171. Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968 Apr;33(2):491-497.

172. Rossmann M, Blow D. The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst. 1962;15:24-31.

173. Trapani S, Navaza J. AMoRe: classical and modern. Acta Crystallogr. D Biol. Crystallogr 2008 Jan;64(Pt 1):11-16.

174. Rossmann MG. Molecular replacement--historical background. Acta Crystallogr. D Biol. Crystallogr 2001 Oct;57(Pt 10):1360-1366.

175. Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends in Genetics 2004 Jan;20(1):33-43.

176. Boudko S, Zintek K, Vance J, Hacker J, Engel J, Bachinger HP. The NC2 domain of collagen IX provides chain selection and heterotrimerization. J Biol Chem. In Press;

177. Boudko SP, Sasaki T, Engel J, Lerch TF, Nix J, Chapman MS, Bächinger HP. Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. J Mol Biol. 2009 Sep;392(3):787-802.

178. Halfter W, Dong S, Schurer B, Cole G. Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem. 1998;273(39):25404-25412.

179. Iozzo RV. The biology of the small leucine-rich proteoglycans. J Biol Chem. 1999 Jul;274(27):18843-18846.

180. Pushpakom SP, Herrick AL, Kumar S, Worthington J. Polymorphisms in COL15 gene are not associated with systemic sclerosis. J Rheumatol. 2008 Feb;35(2):251-253.

181. Utriainen A, Sormunen R, Kettunen M, Carvalhaes LS, Sajanti E, Eklund L, Kauppinen R, Kitten GT, Pihlajaniemi T. Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. Hum Mol Genet. 2004 Sep;13(18):2089-2099.

182. Demeler B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Curr Protoc Protein Sci. 2010 Apr;Chapter 7:Unit 7.13.

183. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr 1994 Sep;50(Pt 5):760-763.

184. Trapani S, Navaza J. AMoRe: classical and modern. Acta Crystallogr. D Biol. Crystallogr 2008 Jan;64(Pt 1):11-16.

185. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr 2010 Apr;66(Pt 4):486-501.

186. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr 1998 Sep;54(Pt 5):905-921.

187. Momota R, Naito I, Ninomiya Y, Ohtsuka A. Characterization of drole, drosophila type XV/XVIII collagen. Matrix Biol. 2008 Dec;27(Supplement 1):23-24.

188. Ogle JD, Arlinghaus RB, Lgan MA. 3-Hydroxyproline, a new amino acid of collagen. J Biol Chem. 1962 Dec;237:3667-3673.

189. Risteli J, Tryggvason K, Kivirikko KI. A rapid assay for prolyl 3-hydroxylase activity. Anal Biochem. 1978 Feb;84(2):423-431.

190. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci. 2003 Dec;28(12):655-662.

191. Makareeva E, Leikin S. Procollagen triple helix assembly: an unconventional chaperone-assisted folding paradigm. PLoS ONE. 2007;2(10):e1029.

192. Nagai N, Hosokawa M, Itohara S, Adachi E, Matsushita T, Hosokawa N, Nagata K. Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. J Cell Biol. 2000 Sep;150(6):1499-1506.

193. Chessler SD, Byers PH. BiP binds type I procollagen pro-α chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. J Biol Chem. 1993;268(24):18226-18233.

194. Buchner J, Schmidt M, Fuchs M, Jaenicke R, Rudolph R, Schmid FX, Kiefhaber T. GroE facilitates refolding of citrate synthase by suppressing aggregation. Biochemistry. 1991 Feb;30(6):1586-1591.

195. Shao F, Bader MW, Jakob U, Bardwell JC. DsbG, a protein disulfide isomerase with chaperone activity. J Biol Chem. 2000 May;275(18):13349-13352.

196. Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. Nature. 1991 Jul;352(6330):36-42.

197. Mendoza JA, Rogers E, Lorimer GH, Horowitz PM. Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. J Biol Chem. 1991 Jul;266(20):13044-13049.

198. Fischer G, Bang H, Mech C. Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides. Biomed Biochim Acta. 1984;43(10):1101-1111.

199. Horibe T, Yosho C, Okada S, Tsukamoto M, Nagai H, Hagiwara Y, Tujimoto Y, Kikuchi M. The chaperone activity of protein disulfide isomerase is affected by cyclophilin B and cyclosporin A in vitro. J Biochem. 2002 Sep;132(3):401-407.

200. Knappe TA, Eckert B, Schaarschmidt P, Scholz C, Schmid FX. Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. J Mol Biol. 2007 May;368(5):1458-1468.

201. Rhodes RK, Miller EJ. Physicochemical characterization and molecular organization of the collagen A and B chains. Biochemistry. 1978 Aug;17(17):3442-3448.

202. Kefalides NA. Structure and biosynthesis of basement membranes. Int Rev Connect Tissue Res. 1973;6:63-104.

203. Pekkala M, Hieta R, Bergmann U, Kivirikko KI, Wierenga RK, Myllyharju J. The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. J Biol Chem. 2004 Dec;279(50):52255-52261.

204. Morello R, Tonachini L, Monticone M, Viggiano L, Rocchi M, Cancedda R, Castagnola P. cDNA cloning, characterization and chromosome mapping of Crtap encoding the mouse Cartilage

Associated Protein. Matrix Biol. 1999 Jun;18(3):319-324.

205. Marini JC, Cabral WA, Barnes AM. Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. Cell Tissue Res. 2010 Jan;339(1):59-70.

206. Tessier DC, Thomas DY, Khouri HE, Laliberté F, Vernet T. Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. Gene. 1991 Feb;98(2):177-183.

207. Saga S, Nagata K, Chen WT, Yamada KM. pH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein. J Cell Biol. 1987 Jul;105(1):517-527.

208. OkamotoT, Schwab RB, Scherer PE, Lisanti MP. Analysis of the Association of Proteins with Membranes | Current Protocols. Unit 5.4. Available from: http://www.currentprotocols.com/protocol/cb0504

209. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom Rev. 2006;25(4):663-682.