Designed triple-helical peptides as tools for collagen biochemistry and matrix engineering

Takaki Koide*†
Faculty of Pharmaceutical Sciences, Niigata University of Pharmacy and Applied Life Sciences, Niigata 956-8603, Japan

Collagens, characterized by a unique triple-helical structure, are the predominant component of extracellular matrices (ECMs) existing in all multicellular animals. Collagens not only maintain structural integrity of tissues and organs, but also regulate a number of biological events, including cell attachment, migration and differentiation, tissue regeneration and animal development. The specific functions of collagens are generally triggered by specific interactions of collagen-binding molecules (membrane receptors, soluble factors and other ECM components) with certain structures displayed on the collagen triple helices. Thus, synthetic triple-helical peptides that mimic the structure of native collagens have been used to investigate the individual collagen–protein interactions, as well as collagen structure and stability. The first part of this article illustrates the design of various collagen-mimetic peptides and their recent applications in matrix biology. Collagen is also acknowledged as one of the most promising biomaterials in regenerative medicine and tissue engineering. However, the use of animal-derived collagens in human could put the recipients at risks of pathogen transmission or allergic reactions. Hence, the production of safe artificial collagen surrogates is currently of considerable interest. The latter part of this article reviews recent attempts to develop artificial collagens as novel biomaterials.

**Keywords:** biomaterial; collagen; extracellular matrix; peptide; supramolecule; triple helix

1. NATURAL COLLAGENS VERSUS SYNTHETIC COLLAGEN-MIMETIC PEPTIDES: AN OVERVIEW

(a) Collagens as multifaceted proteins
All multicellular animals possess collagens as major structural proteins which are responsible for maintaining the structural integrity of tissues and organs. Collagens are characterized by their unique tertiary structure, called the collagen triple helix, and by their existence in extracellular matrices (ECMs). Twenty-seven types of mammalian collagen, consisting of over 40 specific polypeptide chains, called a-chains, have been identified to date. These collagens are classified into several subfamilies according to their molecular and supramolecular shapes (Ricard-Blum et al. 2005). The triple-helical molecules corresponding to collagen types I, II, III, V and XI self-associate to form cross-striated fibrils with a typical axial periodicity of 67 nm. The network-forming collagens contain types IV, VIII and X. Among these, collagen type IV forms basement membranes with other non-collagenous components, such as laminin, entactin and perlecan. Other collagen types are classified as fibril-associated collagens with interrupted triple helices, transmembrane collagens and others. Among the collagen types, even among whole mammalian proteins, collagen type I is the most ubiquitous and abundant.

The collagen triple helix is a right-handed supercoil of three left-handed polypeptide II-like helices (Beck & Brodsky 1998). In the triple helix, adjacent polypeptide chains are staggered in a one-residue manner along the common axis of the helix. This unique tertiary structure comprises tandem repeats of Xaa-Yaa-Gly sequences. A glycine residue at every third position is essential to maintain the triple-helical structure, because the interior of the triple helix does not have sufficient space to accommodate amino acid side chains larger than a hydrogen atom. Side chains of amino acid residues at the Xaa and Yaa positions, extending from the helix, can form binding interfaces for other molecules. In mammalian collagen sequences, about one-third of Xaa and Yaa positions are occupied by Pro and 4(R)-hydroxyproline (Hyp) residues, respectively. The Hyp residues are generated post-translationally by the action of procollagen prolyl 4-hydroxylase in the endoplasmic reticulum (Myllyharju 2005). In normal collagens, most Pro residues occurring at Yaa positions are converted to Hyp residues; therefore, Pro–Hyp–Gly is the most abundant triplet in collagen (Ramshaw et al. 1998). High imino acid content in both Xaa and Yaa positions contributes to the stability of the triple helix. The hydroxylation of Pro residues to generate Hyp also contributes to further stabilization.

A well-known function of collagens is to provide physical strength to tissues as the major components of ECMs (figure 1a). Indeed, tissues subjected to various mechanical forces, such as tendons, skin, bone and cartilage, are rich in collagen fibrils. The meshwork structure of the basement membrane collagen (type IV) also functions as a partition between the epithelium...
and the mesenchyme. Such physical characteristics of collagens are largely attributed to a rigid rope-like structure of the triple helix. It should be noted that collagens are not inert scaffolding materials, but exhibit a variety of biological activities by interacting with a number of other biological molecules. The triple helix of human collagen type I with a length of 300 nm contains binding sites for over 50 other molecules (Di Lullo et al. 2002). The collagen-binding molecules include membrane-anchored receptors, secreted soluble factors and other ECM components, including themselves. Collagen triple helices can directly function as signalling ligands through specific activation of collagen receptors (figure 1a; Vogel 2001). Signalling mediated by collagen-binding integrins (α1β1, α2β1, α10β1 and α11β1; Eble et al. 1993; Golbik et al. 2000; Knight et al. 2000; Xu et al. 2000; Siljander et al. 2004) and discoidin domain receptors (Shrivastava et al. 1997; Vogel et al. 1997) are such cases. Upon binding the collagen to these receptors, intracellular kinase cascades are activated to initiate specific cell responses. 

Collagens also function as reservoirs for soluble factors such as interleukin 2 (Somasundaram et al. 2000) and pigment epithelium-derived factor (PEDF; Kozaki et al. 1998; Yasui et al. 2003; figure 1c). This property, in concert with other physical and biological properties of ECM macromolecules, contributes to creating microenvironments that controls specific cell behaviours and fates. Collagens also regulate protein functions (figure 1d); for example, the activation of von Willebrand factor is triggered by binding to the collagen triple helix (Pietu et al. 1987), and PEDF is thought to exhibit its anti-angiogenic activity in its collagen-bound form (Hosomichi et al. 2005). Additionally, collagens also show numerous indirect biological activities by interacting with other ECM macromolecules, most of which also possess multiple biological activities (figure 1e). Such complex collagen functions are further regulated by the specific degradation of collagens by collagenolytic matrix metalloproteinases (MMPs; figure 1d).

As mentioned earlier, collagen triple helices show many important and specific biological functions, and many of these are initiated by specific interaction of collagen-binding molecules with epitopes displayed on the collagen triple helices. Nonetheless, individual mechanisms that account for specific collagen functions are not well understood, because modern techniques in protein chemistry and molecular biology, which have been developed for general, soluble and globular proteins, have not been very effective in handling collagens, which are insoluble proteins with large molecular sizes.

(b) Collagen-mimetic peptides, useful tools for collagen research

Collagen-mimetic peptides are generally 15–45 residue peptides that mimic the primary and tertiary structures of collagen. Collagen-like (Xaa-Yaa-Gly)_{n} sequences have an inherent ability to form a triple-helical structure spontaneously. Since the publication of the pioneering works on chemical synthesis of simple collagen-mimetic prototypes, (Pro-Pro-Gly)_{n} and (Pro-Hyp-Gly)_{n} (Sakakibara et al. 1968, 1973), many variants have been prepared (for early works, see the reviews of Heidemann & Roth (1982) and Fields & Prockop (1996)). The peptides have been used in studies on the triple-helix structure and stability, as well as in those focusing on biological activities of collagens. Collagen-mimetic peptides are generally prepared by means of chemical synthesis, because they frequently contain Hyp residues at the Yaa positions that are difficult to incorporate using recombinant expression in bacterial cells. It may be that the initial success of work on chemical synthesis resulted in the later development of a variety of molecular designs for collagen-mimetic peptides.

Collagen-mimetic peptides are often used as a soluble surrogate that represents a certain part of native triple-helical collagens. This molecular downsizing approach allows the researcher to perform biochemical investigations with a finer resolution. The use of the model peptides in collagen research is analogous to the use of synthetic oligonucleotides in DNA research which also enables precise biochemical analysis.

2. DESIGNS AND APPLICATIONS OF COLLAGEN-MIMETIC PEPTIDES

Collagen-mimetic peptides have contributed to elucidating both the three-dimensional structure of the collagen triple helix and the origin of its conformational stability. The latter is still a hot topic for debate. We are now able to design and prepare a variety of tailored triple-helical (or even single-stranded) collagen mimetics by virtue of the knowledge accumulated in the field of molecular design and developments in the techniques of peptide synthesis. This section of the review illustrates useful designs and successful applications of collagen-mimetic peptides.

(a) Homotrimeric designs

Among vertebrate collagens, some types (e.g. II, III and VIII) are homotrimeric and the others are heterotrimeric (e.g. I and IV). As models for the homotrimeric collagens, self-associating open-chain peptides, with the sequence (Xaa-Yaa-Gly)_{n}, are most frequently used because they are most easily prepared by standard synthetic protocols. The initial model peptides of the collagen triple helix appeared as...
repeating tripeptides (figure 2a). Over three decades ago, (Pro-Pro-Gly)$_n$ and (Pro-Hyp-Gly)$_n$ were synthesized and their self-trimerizing property was revealed (Sakakibara et al. 1968, 1973). Using such repeating tripeptides, the importance of imino acids, especially Hyp residues in Yaa positions, for stabilizing the triple-helical structure was established. Although most of the collagen-mimetic peptides synthesized to date contain imino acids, the occupation of Xaa or Yaa positions by imino acid residues is not always necessary for maintaining the triple-helical structure. A peptide (Gly-Glu-Arg)$_{15}$ was also shown to form a collagen-like triple helix (Mechling & Bächinger 2000). Not only naturally occurring but also non-natural amino acid residues have been extensively incorporated into the tripeptide units. Marked progress in elucidating the origin of triple-helix stability has been made in a series of recent works that followed on from the finding that the triple helix of [Gly-Pro-4($R$)-fluoroproline (Flp)]$_n$ has outstanding thermal stability (Holmgren et al. 1998). The character of the fluorine atom in an Flp residue is very different from that of the hydroxyl group in a Hyp residue; the fluorine atom is not likely to form hydrogen bonds, but is highly electronegative when compared with the hydroxyl group. Recently, it was found that the electroinductive effect that controls proline puckering and the cis/trans ratio of the peptide bond is the major determinant of triple-helix stability (Bretschger et al. 2001; Renner et al. 2001).

The thermal stability of triple helices depends both on the amino acid sequences and the chain lengths of the peptides. When preparing a triple-helical peptide that possesses a certain collagen sequence of interest, it is important to consider the triple-helix stability of the peptide. The sequences of interest, especially imino-poor sequences, sometimes require external stabilizers to maintain their triple-helical conformation. The stabilizers are either built-in sequences of higher triple-helical propensity, artificial covalent bridges or other non-covalent interactions between the peptide strands (figure 2). One of the most commonly used peptide systems for such purposes is the host–guest peptides (Shah et al. 1996). In this design, a guest (Gly-Xaa-Yaa)$_n$ sequence of interest is flanked by several repeats of a host triplet, which enhances the thermal stability of the triple-helical molecule (figure 2b). Usually, three or four repeats of Gly-Pro-Hyp are added to both ends of the guest sequence. It is useful if the thermal stability of the triple helix can be predicted before starting the actual synthesis of the peptides. The parameters for predicting the triple-helical propensity of each amino acid were first proposed by Bächinger & Davis (1991). Recently, Brodsky and co-workers completed their comprehensive works on parametrization of the triple-helix propensity of Gly-Xaa-Yaa sequences by measuring melting temperatures (the temperature at which a half of a triple helix is denatured) of host–guest peptides with every Xaa and Yaa combination in the guest triplet (Persikov et al. 2000, 2002). They further parametrized the effect of side-chain interactions between adjacent guest triplets (Persikov et al. 2005b). Based on the parameters, computational software for predicting triple-helix stability has been developed and is available on the World Wide Web (Rainey & Goh 2004; Persikov et al. 2005a).

A prominent finding with collagens obtained using host–guest peptides is the X-ray crystallographic structure of T3-785 peptide, (Pro-Hyp-Gly)$_s$-Ile-
Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-(Pro-Hyp-Gly)\textsubscript{4} (Kramer et al. 1999, 2001). This study demonstrated that the local helical twists of the triple helix differ according to the context: the imino-rich host region takes the 7/2 helix, and the central imino-poor guest region forms the 10/3 helix. Host–guest peptides have also been used in co-crystallization with collagen-binding proteins. The crystal structure of a complex between the I domain of z2\beta1 integrin and a host–guest peptide containing a Gly-Phe-Hyp-Gly-Arg motif revealed the triple-helix recognition mechanism of the cell surface receptor at the atomic level (Emsley et al. 2000). The crystal structure of a Staphylococcus aureus collagen-binding protein, CNA, complexed with a host–guest peptide was also resolved recently, and a novel model for collagen–protein interactions was proposed (Zong et al. 2005). It should be emphasized that such fine structural analysis was made possible by the preparation of appropriately designed collagen-mimetic peptides. The structural information obtained contributes not only to basic protein research but also to structure-based drug design.

A number of host–guest peptides have also been used effectively for establishing the binding specificity of collagen-interacting molecules. To establish the binding specificity of collagen-binding integrins, host–guest peptides with various guest sequences were used in binding assays. As a result, the Gly-Glu-Arg (GER) sequence in the collagen triple helix was identified as the dominant motif for the integrins (Knight et al. 2000; Xu et al. 2000). Farndale and colleagues have recently identified other high-affinity-binding motifs for integrin z2\beta1 by the screening of 57 synthetic host–guest peptides encompassing the whole triple-helical domain of human collagen type III (Raynal et al. 2006). In addition, a system using designed host–guest peptides has been applied to the identification of a binding motif for Hsp47, a collagen-specific molecular chaperone (Koide et al. 2002a).

Covalent bridges that tether three peptide strands together are often used for enhancing the stability of self-assembled triple helices (figure 2c). Such covalently tethered peptides can be established, either on trifunctional organic templates or by cross-linking between side-chain functional groups of extra amino acids added to the ends of the peptide chains. The former approach includes the use of cis,cis-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid, called Kemp triacid (Kemp & Petrakis 1981; Goodman et al. 1998), or cyclotriverterylene templates (Rump et al. 2002). The latter also includes Fields’ Lys–Lys branching (Fields et al. 1993) and Moroder’s cystine knots (Ottl & Moroder 1999).

The triple-helical conformation of a peptide can be stabilized also by non-covalent trimerizing interactions (figure 2d). One example is the (fibrin–foldon domain)-(Gly-Xaa-Yaa), fusion protein system. The gene encoding the fibrin–foldon domain of bacteriophage T4, a homotrimer-forming, 27-residue domain, was fused in frame to a gene encoding collagenous (Gly-Xaa-Yaa), sequences, and the resulting recombinant fusion protein was expressed in Escherichia coli cells (Frank et al. 2001). The non-covalent homotrimeric association of the foldon domain was shown to increase the stability of self-assembled triple helices. Metal coordination to built-in chelating ligands has also been shown to stabilize the triple-helical conformation of collagenous peptides. Koide et al. (2002b) synthesized collagen-mimetic peptides with the N-terminal, 2,2'-bipyrindine (bpy) group. Addition of the Fe(II) ion to the peptide solution to form an Fe\textsuperscript{II}(bpy), complex increased the thermal stability of the triple helix (Koide et al. 2002b). Similarly, Goodman and co-workers used N-terminally introduced catechol groups or C-terminally introduced hydroxamate groups as the ligands for complex formation with an Fe(III) ion (Cai et al. 2004; Kinberger et al. 2006). It is worth mentioning that covalent and non-covalent stabilizers can be introduced at the N\textsubscript{2}, C- or both termini of the collagenous peptides to enhance triple-helix stability (Tanaka et al. 1998; Frank et al. 2003).

The use of covalent bridges for the stabilization of the triple helix enabled us to separately prepare triple-helical and single-stranded collagenous peptides with the same peptide sequences. In particular, by employing the cystine-knot strategy to tether three peptide strands, we were able to destabilize the triple-helical conformations simply by adding a reducing agent, such as dithiothreitol (DTT). In our case, an N-terminally tethered homotrimeric peptide, with the sequence (Gly-Pro-Hyp)\textsubscript{2}-Gly-Pro-Arg-(Gly-Pro-Hyp),-Gly-amide, was shown to be fully triple helical at 25°C, while in the presence of DTT, the peptide assumed a random-coil conformation (Koide et al. 2004). By taking advantage of a similar system involving conformationally constrained peptides, we showed that the collagen-specific molecular chaperone Hsp47 only interacts with triple-helical molecules (Koide et al. 2006a).

(b) Heterotrimeric designs

Some types of collagen are composed of different \( \alpha \)-chains to form heterotrimeric molecules. The most abundant collagen type I, for example, is composed of two \( \alpha \textsubscript{1}(I) \) chains and one \( \alpha \textsubscript{2}(I) \) chain. The \( \alpha \)-chain composition of collagen type IV is more complex: triple-helical molecules with the composition \( \{z1(IV)\}_{n} \text{z2(IV)}, \text{z3(IV)} \text{z4(IV)} \text{z5(IV)} \) or \( \{z5(IV)\}_{n} \text{z6(IV)} \) are found with a tissue-dependent distribution. In the biosynthesis of such heterotrimeric collagen molecules, selective heterotrimerization at the C-terminal propeptide domain, which proceeds to propagation of the triple helix, is responsible for the correct \( \alpha \)-chain arrangement (Koide & Nagata 2005). It is believed that repeated amino acid sequences of Xaa-Yaa-Gly do not have sufficient structural information for the selective formation of heterotrimers with distinct chain arrangements.

Heterotrimeric collagen-mimetic peptides can also be prepared by applying the cystine-knot strategy. Unlike the homotrimeric collagen-mimetic peptide system that uses intrinsic peptide self-trimerization, the chain compositions of the heterotrimeric peptides are controlled chemoselectively by stepwise disulphide bond formation (figure 3). To date, either C- or N-terminally tethered heterotrimeric peptides have been synthesized (Ottl & Moroder 1999; Ottl et al. 2000; Koide et al. 2004). In heterotrimeric triple helices, structural isomers with different chain
staggering can be formed because the peptide chains in the triple helix are staggered by one residue relative to each other. A heterotrimeric triple helix with three different peptide chains, \(a_1\), \(a_2\) and \(a_3\), can give rise to six isomers with the staggered chains \(a_1 a_2 a_3\), \(a_1 a_3 a_2\), \(a_2 a_1 a_3\), \(a_2 a_3 a_1\), \(a_3 a_1 a_2\) and \(a_3 a_2 a_1\). The peptide system with a C-terminal cystine knot has enabled us to synthesize individual isomers with unambiguously staggered chains. This distinguishing feature of the system was successfully employed for the interaction of the triple helix are staggered by one residue relative to each other.

### 3. TOWARDS TOTALLY SYNTHETIC COLLAGEN SURROGATES FOR NOVEL BIOMATERIALS

#### (a) Native collagens as biomaterials

A variety of natural, semi-synthetic and synthetic macromolecules or polymers have been developed and used as biomaterials (Rosso et al. 2005). Biomaterials are expected to function as cell scaffolds that could replace native ECMs. Among these, collagens, mostly type I, are acknowledged as one of the most useful biomaterials. Collagens of natural origin, typically from pigs or cows, are widely used for tissue engineering (Ruszczak 2003; Yarlagadda et al. 2005), cosmetic surgery (Narins & Bowman 2005) and drug delivery systems (Wallace & Rosenblatt 2003). Such animal-derived collagens are readily purified in large amounts at relatively low cost. The collagens are used either in their native fibrillar form or after denaturation (gelatinization) in variously fabricated forms, such as sponges, sheets, plugs and pellets. The physical and chemical properties of collagens are often modified by cross-linking or chemical modification depending on their purpose. Although animal-derived collagen has been recognized as a very safe material, rarely eliciting heterotrimerization of designed collagenous peptides without using the special techniques of peptide synthesis.

#### (c) Collagen-mimetic peptides with functional moieties

Chemical synthesis of peptides allows us to introduce various functional moieties to collagen-mimetic peptides. One example is the development of fluorescence resonance energy transfer (FRET) substrates for detecting the activity of collagenolytic MMPs. Using the FRET substrates, the enzymatic activity can be monitored spectroscopically. These FRET substrates use the quenching effect of the fluorescence by a quencher chromophore placed in close vicinity to the fluorophore, and the quenching is designed to be cancelled upon the cleavage of scissile peptide bonds. One design of FRET substrates for MMPs is a heterotrimeric collagen-mimetic peptide harbouring a FRET pair of a tryptophan (fluorophore) and a dansyl group (quencher) at the N-termini of two of the three peptide chains (Müller et al. 2000). The other is a homotrimeric, open-chain, collagen-mimetic peptide possessing a FRET pair of methoxy coumarin/dinitrophenol that flanks the scissile peptide bond (Laufer-Fields et al. 2000, 2001; Minond et al. 2004).

Another example of collagen-mimetic peptides with functional moieties is the recently reported photoaffinity substrates for the collagen-specific molecular chaperone Hsp47 (Koide et al. 2006a). Host–guest peptides were designed with one Pro-Arg-Gly tripeptide unit that functions as the recognition sequence for Hsp47 and a photoreactive \(\beta\)-benzoyl-\(L\)-phenylalanine residue in either the Xaa or Yaa positions. Specific photoaffinity cross-linking was observed after UV-irradiation of the chaperone–peptide complex. Such a system involving photoaffinity substrates has been shown to be useful for mapping the collagen–protein-binding interface.

---

*Phil. Trans. R. Soc. B* (2007)
significant immune responses, risks attached to the use of the collagens in humans have been emerging with increasing clinical applications. We should now consider the risk of contamination with prions that cause bovine spongiform encephalopathy (O’Grady & Bordon 2003) as well as that of allergic reactions (Sakaguchi et al. 1999; Lynn et al. 2004). The development of novel and safer biomaterials that can replace the current use of natural collagen is a great interdisciplinary challenge.

Recently, supramolecular structures that are obtained by self-assembly of designed synthetic molecules have been extensively studied, and their potential for biomedical applications has emerged (Lutolf & Hubbel 2005). In particular, the use of peptides consisting of common amino acids as units for self-assembled supramolecules is a promising approach, because they are expected to possess high biocompatibility and flexibility in molecular designs (Zhang 2003).

(b) Peptide-amphiphile nanofibres

One of the successful designs of artificial scaffolds based on molecular self-assembly is the peptide-amphiphile nanofibres developed by Stupp and co-workers (Hartgerink et al. 2001, 2002). The self-assembling unit, a peptide-amphiphile, is a hydrophilic 11-residue peptide with an N-terminal alkyl tail with 16 carbon atoms. The peptide-amphiphiles self-associate to form cylindrical micelles, or nanofibres, packing the alkyl tails in the centres of the fibres. The fibrous supramolecular structure can be stabilized by disulphide cross-links between the Cys residues in the peptide moieties. This covalent capture step is effective in enhancing the strength of the nanofibre hydrogel, similar to the naturally occurring, covalent cross-linking between collagen triple helices that enhances their fibre strength.

(c) Self-assembled peptide supramolecules with b-sheet or a-helical coiled-coil structures

The intrinsic ability of a peptide to form a certain secondary structure is also used in the creation of a supramolecular scaffold. The self-assembly process relies on the spontaneous, regular and intermolecular folding that creates a sticky end for the addition of the next peptide unit. Self-assembled nanofibres of b-sheet-forming peptides, which include the well-known disease-causing amyloid fibrils, seem to be a feasible scaffold for biomaterials (Zhang 2003). b-Sheet nanofibres have been used for in vitro primary culture and in vivo experiments of tissue repair.

The intermolecular a-helical coiled-coil formation of designed peptides has also been used for creating elongated peptide fibres (Pandya et al. 2000; Oghara et al. 2001; Potekhin et al. 2001; Ryadnov & Woolfson 2003a,b, 2004). Among these, Woolfson’s sticky-end assembly is an attractive system owing to its simplicity and flexibility in molecular design. The system is composed of two different a-helix-forming peptides that are designed for direct formation of staggered coiled-coil heterodimers. These have sticky ends to promote formation of long fibres (Pandya et al. 2000). By incorporating different pieces into such a self-assembling, molecular Lego system, one can regulate the morphology of the supramolecules by introducing branches or kinks (Ryadnov & Woolfson 2003a,b). It is also possible to decorate the peptide fibre with molecules of interest by incorporating amino acids with a molecular recognition moiety (e.g. biotin) into the Lego pieces (Ryadnov & Woolfson 2004).

(d) Supramolecular materials consisting of collagen-like triple helices

Self-assembling peptide supramolecules that rely on the triple-helix-forming ability of Xaa-Yaa-Gly repeats have not been extensively studied to date, although it seems to be reasonably straightforward to create a triple-helical collagen surrogate. One attempt at producing an artificial collagen is the development of a collagen-like peptide-amphiphile system by Fields and colleagues (Fields et al. 1998; Lauer-Fields et al. 2003). These peptide-amphiphiles are mono-alkylated, collagen-like (Xaa-Yaa-Gly)n repeat peptides. The amphiphiles self-assemble to form higher-order structures and insoluble deposits. The material containing the sequence corresponding to the cell attachment sequence (1263–1277) of the collagen z1(IV) chain was shown to exhibit specific cell attachment and focal adhesion kinase activation activities (Fields et al. 1998; Lauer-Fields et al. 2003). Kaplan and co-workers also prepared triple-helical supramolecules of a distinct higher-order structure by the self-assembly of peptides with the structure, (Glu)5-(Gly-Xaa-Hyp-Gly-Pro-Hyp)6-(Glu)5 (Valluzzi & Kaplan 2000; Martin et al. 2003).

Another direction for the development of artificial collagen-like materials is to prepare longer Xaa-Yaa-Gly repeats with a molecular size close to that of native collagen a-chains. Tandem polymers of (Xaa-Yaa-Gly)n sequences with higher-molecular masses have been prepared using bacterial expression systems or chemical condensation of synthetic peptides, leading to formation of gelatin-like hydrogels or collagen-like fibrils.

The recombinant expression of (Gly-Pro-Pro)n polypeptide was first tried by Goldberg et al. (1989) using E. coli as a host. However, they did not report the biochemical/biophysical properties of the recombinant protein. Later, Kajino et al. (2000) constructed an expression plasmid encoding six and eight repeats of (Gly-Xaa-Yaa)10 sequences, and the corresponding polypeptides, artificial gelatins, were extracellularly produced using a Bacillus brevis expression system. Although the length of these artificial gelatins (180 and 240 amino acids) was shorter than that of the native collagen I triple-helical domain (1018 residues), they exhibited a sol–gel transition, similar to gelatin. However, their threshold temperature is lower than body temperature and their gelating properties are much weaker than those of gelatin. In addition, a major drawback of such a bacterial expression system is a lack of post-translational modification enzymes, most importantly, prolyl 4-hydroxylase.

In 2005, Tanihara and colleagues reported the preparation and characterization of poly(Pro-Hyp-Gly)n. They simply polymerized commercially available (Pro-Hyp-Gly)n (n=1, 5 and 10) by the action of water-soluble carbodiimide and 1-hydroxybenzotriazole to obtain polypeptides with molecular weights greater than 10 000 dalton (figure 4a, Kishimoto et al. 2005).
By polymerizing the peptides, the triple-helix thermal stability was increased up to 80°C, and nanofibre-like suprastructure was observed by electron microscopy. However, this condensation method, in principle, requires the protection of side-chain functional groups when biologically relevant sequences are to be incorporated.

In the same year, similar poly(Xaa-Yaa-Gly)n was prepared by means of the native chemical ligation technique, in which synthetic peptide units spontaneously polymerize in aqueous solutions (figure 4b, Paramonov et al. 2005). Using this method, investigators successfully obtained collagenous peptide polymers with an apparent molecular weight less than or equal to 1 000 000 dalton. The triple-helical polymers formed a nanofibre-like structure with a length in micrometres. The polymerization by the native chemical ligation, which proceeds under mild conditions, enables incorporation into the polymer of amino acid residues with various side-chain functional groups. In addition, the Cys residues on the triple-helical polymers, which are necessary for the native chemical ligation, could be used for further cross-linking between triple helices or immobilization of functional moieties.

However, efforts to increase the size of poly(Xaa-Yaa-Gly)n for the development of artificial collagens seem to have an intrinsic problem of chain misalignment. Generally, shorter collagenous peptides readily form perfectly aligned trimers, but the thermal stability of the shorter triple helix is lower than that of the longer helices. In contrast, very long collagenous sequences can form shorter triple helix is lower than that of the longer helices. Generally, shorter collagenous peptides readily form perfectly aligned trimers, but the thermal stability of the shorter helices. In contrast, very long collagenous sequences can form perfectly aligned trimers, but the thermal stability of the shorter helices.

Recently, Raines’ group and ours have independently developed a novel peptide-based system for obtaining collagen-like triple-helical supramolecules via a spontaneous self-assembly process (figure 4c, Koide et al. 2005; Kotch & Raines 2006). The designed peptide units are self-complementary trimers based on Pro-Hyp-Gly repeats. In the trimer, peptide strands are tethered by two disulphide bridges in a staggered arrangement. The molecular design prevents intramolecular triple-helix formation, but allows the intermolecular folding to form elongated triple-helical supramolecules. The concept of self-assembly is similar to that of the sticky-end elongation of α-helical coiled-coil peptides (Pandya et al. 2000).

4. Conclusions and Future Outlooks

Collagens are peculiar proteins. As described in this article, they have unique primary, tertiary and supramolecular structures, post-translationally modified amino acids and physicochemical properties. Owing to such properties, research on collagens has evolved in a different manner from that on general, globular and soluble proteins. Instead of using recombinant proteins expressed in E. coli, researchers have chosen to study chemically synthesized model peptides that mimic a part of the collagen triple helix. From a given collagen sequence, it is now possible to design and synthesize a peptide that mimics its structure and predicts its triple-helix stability. Furthermore, it is possible to add appropriate functional groups, such as fluorophores and photoactive cross-linking agents, to the peptides. Peptides that mimic certain portions of native collagen molecules are especially useful in investigating the binding specificity of collagen-binding molecules. The binding site of a collagen-binding protein is one of the functional sites of the collagen molecule. Thus, the peptide representing the structure of a protein-binding site is expected to exhibit one of the specific functions of collagens. Such a functional collagen-mimetic peptide may be useful in giving functionality to artificial biomaterials (Reyes & Garcia 2004).

ECM is a treasure house of functional peptide sequences. The well-known Arg-Gly-Asp (RGD) sequence, first identified from fibronectin (Pierschbacher & Ruoslahti 1984), acts as a ligand for some members of the integrin family, and plays a crucial role in cell attachment/signalling. Laminin, a major component of basement membranes, is also known to interact with many proteins and proteoglycans. Many peptide sequences responsible for these interactions have been identified on laminin chains by systematic searching of synthetic peptides (Suzuki et al. 2003, 2005). Some of the functional peptides derived from laminin have been used to give functionality to biomaterials (Mochizuki et al. 2003). There are also dozens of collagen-binding proteins known to date (Di Lullo et al. 2002). Conventionally, protein-binding sites in collagen triple helices are determined by morphological assessment of the binding complex by electron microscopy. Cyanogen...
bromide-cleaved fragments of collagens are also used for \textit{in vitro} binding assays. However, the majority of the structures of the triple-helical epitopes responsible for binding have not been identified. Effective strategies for obtaining novel functional peptide sequences from native triple-helical collagens are currently required. Recently, Farndale and co-workers prepared a collection of triple-helical peptides that encompasses the whole sequence of the triple-helical domain of human collagen type III \cite{Raynal2007}.

This peptide library, named Collagen III Toolkit, should be very effective for examining the specificity of collagen type III-binding proteins. A library of triple-helical peptides can also be expressed in the yeast \textit{Saccharomyces cerevisiae}. We have demonstrated the yeast two-hybrid selection of the HS\text{p}47-binding triple-helical peptide \textit{library} \cite{Koide2000}.

An alternative, promising approach for determining protein-binding sites in collagen triple helices is the collagen-footprint method developed by \textit{Yasui \& Koide} \cite{Yasui2003}. In this system, a unique thiol tag is specifically transferred from an engineered collagen-binding protein to collagen by the reduction of UV-induced cross-linking products.

Ideally, artificial collagen-mimetic biomaterials should mimic both the physical and biochemical properties of native collagens. The excellent physical properties of native fibril-forming collagens when acting as biomaterials are largely attributable to their supramolecular structure; therefore, it is particularly important, for the development of artificial collagen-like biomaterials, to mimic the higher-order structure of collagens. Although now there is sufficient information and techniques to control the tertiary structure of collagenous peptides, it is still very difficult to mimic the higher-order structure of native collagens. The attempts towards forming self-assembled triple-helical supramolecules are still in their infancy.

I am grateful to Dr Shinichi Asada for his critical reading of this manuscript. The author’s projects on collagen research were supported by a Grant-in-Aid for Scientific Research on Priority Areas (no. 15032238) and a Grant-in-Aid for the Encouragement of Young Scientists (no. 16689004) by Ministry of Education, Culture, Sports and Technology, Japan.

REFERENCES


Phl. Trans. R. Soc. B (2007)

\textit{Ebble, J., Golbik, R., Mann, K. \& Kühn, K.} 1993 The \(\alpha1b1\) integrin recognition site of the basement membrane collagen molecule \([x1(IV)\cdot2(IV)]\).


Kishimoto, K., Morihara, Y., Osanai, M., Ogata, S., Kinberger, G. A., Taulane, J. P. & Goodman, M. 2006 FeIII-
Phil. Trans. R. Soc. B

Koide, T., Nishikawa, Y. & Takahara, Y. 2004 Synthesis of


Koide, T., Aso, A., Yorihuzi, T. & Nagata, K. 2000 Conformational requirements of collagenous peptides for
recognition by the chaperone protein HSP47. J. Biol. Chem. 275, 27 957–27 963.

Koide, T., Takahara, Y., Asada, S. & Nagata, K. 2002α Xaa-
Arg-Gly triplets in the collagen triple helix are dominant
binding sites for the molecular chaperone HSP47. J. Biol. Chem. 277, 6178–6182. (doi:10.1074/jbc.M106497200)


Koide, T., Homma, D. L., Asada, S. & Kitagawa, K. 2005 Self-complementary peptides for the formation of


Phil. Trans. R. Soc. B
Koide, T., Aso, A., Yorihuzi, T. & Nagata, K. 2000 Conformational requirements of collagenous peptides for
recognition by the chaperone protein HSP47. J. Biol. Chem. 275, 27 957–27 963.

Koide, T., Takahara, Y., Asada, S. & Nagata, K. 2002α Xaa-
Arg-Gly triplets in the collagen triple helix are dominant
binding sites for the molecular chaperone HSP47. J. Biol. Chem. 277, 6178–6182. (doi:10.1074/jbc.M106497200)


Koide, T., Homma, D. L., Asada, S. & Kitagawa, K. 2005 Self-complementary peptides for the formation of


