A yeast two-hybrid system to obtain triple-helical ligands from combinatorial random peptide libraries

Ryo Masuda[1*†], Khine Phyu Phyu Thant[2*†], Kazuki Kawahara[3], Hiroya Oki[4], Tetsuya Kadonosono[5], Yuji Kobayashi[3], Takaki Koide[1,2*]

[†]These authors contributed equally to this work.

[1]Waseda Research Institute for Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

[2]Department of Chemistry and Biochemistry, School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

[3]Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka 565-0871, Japan

[4]Department of Infection Metagenomics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan

[5]Department of Life Science and Technology, School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8501, Japan

***Corresponding author**:

Professor Takaki Koide, Ph.D.

Department of Chemistry and Biochemistry, School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

Tel.: +81-3-5286-2569, E-mail: koi@waseda.jp

Abstract

Many bioactive proteins interact with collagen, recognizing amino acid sequences displayed on the triple helix. We report here a selection strategy to obtain triple-helical peptides that interact with the proteins from a combinatorial random library constructed in yeast cells. This system enables us to select them using the standard two-hybrid protocol, detecting interactions between triple-helical peptides and target proteins fused to the GAL4-activating and binding domains, respectively. The library was constructed to contain triple-helical peptides with a "host-guest" design in which host helix-stabilizing regions flanked guest random sequences. Using this system, we selected peptides that bind to pigment epithelium-derived factor (PEDF), a collagen-binding protein that shows anti-angiogenic and neurotrophic activities, from the following libraries. Two-step selections from the total random library and subsequently from the second focused library yielded novel PEDF-binding sequences that exhibited affinity comparable to or more potent than that of the native PEDF-binding sequence in collagen. The sequences also contained a variant of the PEDF-binding motif that did not match the known motif identified from the native collagen sequences. This combinatorial library system allows the chemical space of triple-helical peptides to be screened more widely than that found in native collagen, thus increasing the expectation of obtaining more specific and high-affinity peptides.

Significance

Combinatorial random libraries of structurally constrained peptides have been constructed to obtain bioactive peptides with high affinity and/or biostability. This is the first report on the construction of a random collagen-like triple-helical peptide library by a DNA-encoded method and on an easy-to-access method of selecting triple-helical peptides that bind to proteins of interest. We identified several triple-helical peptides with unnatural sequences that bind to a collagen-binding protein. The binding affinity of these peptides was comparable to that of the peptide with the natural binding sequence in collagen. Our system demonstrates the feasibility of obtaining bioactive triple-helical peptides that could be used as drug leads or bioactive moieties of artificial extracellular matrices.

Introduction

The triple helix, an iconic tertiary structure of proteins in the collagen family, is a right-handed supercoil of three left-handed polyproline-II helices consisting of tandem Gly-Xaa-Yaa tripeptide repeats. Pro residues at Xaa positions and post-translationally modified 4-hydroxyproline (Hyp) residues at Yaa positions are frequently found in natural collagen sequences, contributing to conformational thermal stability (1,2).

Fibril-forming types of collagen, such as types I, II, III, V, and XI, have long (>1000 amino acids) uninterrupted triple helices flanked by short telopeptides at both ends of the helices. These collagen types are prominent components of the extracellular matrix and play pivotal roles in maintaining the integrity of the organs and tissues of animals by forming rigid fibers. They also exhibit various physiological functions to regulate cell attachment, differentiation, angiogenesis, and blood coagulation (3,4). Such functions are elicited through specific interactions of bioactive macromolecules with certain regions in the triple helices. In many cases, the interactions are amino acid sequence-specific: collagen-binding integrins such as α1β1, α2β1, α10β1, and α11β1 recognize Gly-Phe-Hyp-Gly-Glu-Arg, Gly-Leu-Hyp-Gly-Glu-Asn, and other similar sequences containing Glu residues at Xaa positions (5-8). Discoidin-domain receptors (DDRs) 1 and 2, collagen-specific tyrosine-kinase receptors, recognize Gly-Val-Met-Gly-Phe-Hyp-containing sequences. The same region is shared by other collagen-binding proteins such as von Willebrand factor (VWF), a serum blood coagulation factor, and secreted protein acidic and rich in cysteine (SPARC), also known as osteonectin and BM-40, a factor regulating collagen assembly (9-12). Pigment epithelium-derived factor (PEDF), a neurotrophic/anti-angiogenic serpin, binds to the Lys-Gly-His-Arg-Gly-Phe-Ser-Gly-Leu sequence [named α1(I)930–938] in the type I collagen triple helix, which overlaps with the heparan sulfate proteoglycan (HSPG)-binding sites and crosslinking sites with the allysine-containing telopeptide region between different triple helices

(13-16). Because two or more proteins share overlapping binding sites on the native collagen triple helix, the collagen triple helix performs complex regulation of physiological events with multiple intertwined macromolecular interactions.

Synthetic triple-helical peptides containing certain parts of natural collagen sequences have been effectively used to identify amino acid sequences in the collagen triple helices recognized by the collagen-binding macromolecules. Because most fragments of natural collagen generated by enzymatic or chemical cleavage lose the ability to maintain the triple-helical conformation, synthetic peptides are designed to be so-called "host-guest peptides" in which the natural collagen sequences are flanked by several repeats of Pro-Pro-Gly or Pro-Hyp-Gly that enhance the thermal stability of the triple helix (17,18). Farndale and co-workers established the collections of host-guest peptides that cover the entire amino acid sequences of fibril-forming types of collagen, and named them "Collagen Toolkits." The amino acid sequences responsible for the binding to the collagen-binding integrins, VWF, DDRs, and SPARC (described above), were identified by the screening of Collagen Toolkits II and III, which covered the amino acid sequences of α1(II) and α1(III) chains, respectively (9-12). The PEDF-binding sequence, α1(I)930–938, was also identified by Koide and co-workers using synthetic triple-helical peptides (13,14).

Generally, the searching of random peptide libraries is a strategy to identify binding sequences for proteins of interest (POIs). Combinatorial random peptide libraries are constructed chemically by split-and-mix solid-phase synthesis or biologically by a DNA-encoded method involving degenerated codons. In principle, random peptide libraries contain a greater variety of peptides, including novel peptides that are not encoded in genomes. Such libraries, combined with high-throughput screening methods, are effective at obtaining a set of amino acid sequences recognized by POIs and determining the responsible binding sequence motifs.

Previously, our group reported a yeast two-hybrid system to detect the interaction of triple-helical peptides with heat-shock protein 47 (HSP47) (19). HSP47 is a collagen-specific chaperone that completes procollagen folding by stabilizing the triple helix that is otherwise unstable at the body temperature of mammals (20,21). On the basis of this system, we here constructed combinatorial random peptide libraries designed to obtain triple-helical amino acid motifs recognized by POIs, by taking advantage of the host-guest peptide design. The peptide library expressed in yeast cells was screened by two-hybrid methods. Using genetically encoded triple-helical peptide libraries as a source, we expect to obtain more specific binders of collagen-binding proteins. In the following sections, we describe the design, construction, and validation of the libraries, followed by the results of actual selections of PEDF-binding triple-helical peptides.

Results

Design, construction, and validation of the random peptide library for Y2H selection

To detect the interaction of a POI and a triple-helical peptide, we designed a yeast two-hybrid system, as shown in Fig. 1A. Here, a POI and a peptide with collagenous Gly-Xaa-Yaa (here, we used Yaa-Gly-Xaa repeats for experimental convenience) are expressed as fusion proteins of GAL4-BD and GAL4-AD, respectively. The AD-fusion proteins were expected to trimerize, concomitant with triple-helix formation. The trimer binds to one, two, or three molecules of the BD-fusion POI. The resulting oligomeric complex can activate the expression of the reporter genes as well as the general 1:1 complex (19).

Fig. 1B shows the design of GAL4-AD-fusion random peptides with a propensity to form a triple helix. In the construct, we took advantage of the host-guest peptide for collagen-like peptides to enhance the thermal stability of the triple helix (17,18). A middle portion consisting of three Yaa-Gly-Xaa repeats (guest) was flanked at each end by ten repeats of Pro-Gly-Pro triplet

(host) with the ability to stabilize triple helices. NNK codons encoded each of three Xaa and Yaa amino acids to generate a randomized guest sequence in the peptides (N and K are equimolar mixtures of A/T/G/C and T/G, respectively). We randomized the three triplet repeats as the guest region, Yaa1-Gly2-Xaa3-Yaa4-Gly5-Xaa6-Yaa7-Gly8-Xaa9, to provide specific interactions based on the sizes of binding interfaces between known collagen-binding proteins and triple-helical peptides. Most collagen-binding proteins or domains recognize fewer than three triplets (nine amino acids), as described above (5-14), although the longest case in the catalytic domain of matrix metalloprotease-1 is 10 amino acids, Hyp-Gly-Pro-Gln-Gly-Leu-Ala-Gly-Gln-Arg (*SI Appendix*, Table S4) (22). In theory, the diversity of the library is 20^6 (= 6.4×10^7). After the transformation of library plasmids into yeast cells, the diversity of the constructed library was estimated to be 4.0×10^6, by counting the colony numbers.

The actual diversity of the random peptide library expressed in yeast cells was estimated by next-generation sequencing (NGS). The relative frequency of occurrences of Xaa and Yaa amino acid residues is shown in Fig. 2A. All possible Xaa-Yaa combinations were found in the library, and the frequencies of occurrence corresponded well to the theoretically expected ones (*SI Appendix*, Fig. S3). The frequencies of Xaa-Yaa combinations in α-chains of human fibrillar collagen, α1(I), α2(I), α1(II), α1(III), α1(V), α2(V), α3(V), α1(XI), and α2(XI), are similarly shown for comparison (Fig. 2B). In contrast to the results of the peptide library, the frequency is significantly biased to the triple-helix-stabilizing Xaa-Yaa combinations. The Pro-Hyp combination with the highest triple-helix-stabilizing effect is the most prominent (11.2%) in human fibrillar collagen. Meanwhile, the frequency of occurrence decreases as the helix-stabilizing effect decreases. Notably, the combinations categorized with the lowest stabilization score do not occur (the combinations shown in blue in Fig. 2B). These findings indicate that the random peptide library, at least at the genetic level, contains a much greater variety of Xaa-Yaa

combinations than human fibrillar collagen, including ones that had been rejected by natural selection.

Triple-helix-forming propensity of the peptides in yeast cells

The molecular chaperone HSP47 selectively binds to collagen-like peptides with a triple-helical conformation (19,23). Although the high-affinity binding site contains an Arg at the Yaa position (24), the triple-helical Pro-Pro-Gly repeating sequence is also recognized by HSP47 (25). This structural requirement of HSP47 for recognizing the repeating sequence is useful for determining the conformational states of the peptides expressed in yeast cells.

We expressed (Pro-Gly-Pro)-repeating peptides with a repeat number of 10, 13, or 23, as GAL4-AD-fusion proteins in yeast cells. The lengths of $(\text{Pro-Gly-Pro})_{10}$, $(\text{Pro-Gly-Pro})_{13}$, and $(\text{Pro-Gly-Pro})_{23}$ correspond to those of the *N*-terminal host, the guest plus the *N*-terminal host, and the guest plus the *N*- and *C*-hosts in the above peptide design (Fig. 1B), respectively. The shorter peptides have lower thermal stability of the helix. We set three different culture temperatures, 16, 30, and 37°C, with the expectation of identifying binding of BD-fusion HSP47 to these three peptides with GAL4-AD by the yeast two-hybrid system (Fig. 3). The mated yeast cells were selected by His and Ade auxotrophy in the presence of aureobasidin and X-α-Gal. The interaction between HSP47 (bait) and peptides (prey) was detected by the growth of blue colonies on the selection plates. Overall, culturing at lower temperatures resulted in the growth of blue colonies expressing the peptides with shorter Pro-Gly-Pro repeats. HSP47 bound only to $(\text{Pro-Gly-Pro})_{23}$ in yeast cells at 37°C. At 30°C, HSP47 bound to $(\text{Pro-Gly-Pro})_{13}$ as well as $(\text{Pro-Gly-Pro})_{23}$, but not to $(\text{Pro-Gly-Pro})_{10}$. At 16°C, HSP47 bound to all three of the peptides. This temperature-dependent result demonstrates that HSP47 distinguished the conformational state of the host Pro-Gly-Pro repeats in yeast cells. This indicates that BD-fusion HSP47 can be used as

an indicator for determining the triple-helical conformation of peptides expressed in yeast cells because it binds to (Pro-Gly-Pro) repeats in the host region of the AD-fusion triple-helical peptides.

Y2H selection of PEDF-binding peptides from the random peptide library

Two-hybrid screening of the random peptide library was conducted using PEDF as a target protein. PEDF is a collagen-binding serpin that exhibits anti-angiogenic and neurotrophic activities (26-28). It binds to the Lys-Gly-His-Arg-Gly-Phe-Ser-Gly-Leu sequence at colα1(I) 930–938 (14,29). The motif on the triple helix for binding to PEDF was identified as Lys-Gly-Xaa-Arg-Gly-Phe-Yaa-Gly-Leu by Ala-scanning of α1(I)930–938 (14).

The full-length mouse PEDF was expressed as a GAL4-BD-fusion protein in yeast cells and used as bait for the screening. The selection of 1.2×10^6 clones at 25°C yielded three hit clones. Their guest sequences were determined as follows: Lys-Gly-Cys-Arg-Gly-Leu-His-Gly-Leu (named 1-1), Arg-Gly-Val-Arg-Gly-Phe-Glu-Gly-Cys (named 1-2), and Arg-Gly-His-Arg-Gly-Phe-Leu-Gly-Leu (named 1-3) (Fig. 4A). Although none of these sequences matches those reported in NCBI and UniProt databases, the above three hit clones satisfied the reported structural requirements for binding to PEDF (14): basic amino acids at Yaa1, Arg at Yaa4, and hydrophobic amino acids at Xaa6. Leu at Xaa9, whose replacement by Ala in α1(I)930–938 had led to partial reduction of the binding affinity to PEDF (14), was also found in the 1-1 and 1-3 sequences.

To determine whether the peptides with the obtained sequences bind to PEDF in the same manner as the native PEDF-binding sequence, α1(I)930–938, we examined independent two-hybrid interactions of the novel three peptides and their mutants in which Arg at Yaa4 was substituted with Ala. The peptide with an α1(I)930–938 sequence in which Arg at Yaa4 is

substituted with Ala does not bind to PEDF (14). We also examined the binding of the peptides to collagen-binding-null or heparin-binding-null PEDF mutants that we already reported (30). As shown in Fig. 4B, interactions with the AD-fusion peptides with 1-1 and 1-3 sequences, as well as α1(I)930–938, were observed. Meanwhile, interaction was not observed for the peptide with the 1-2 sequence, probably because Cys at Xaa9 led to weaker interaction with PEDF than Leu (14). The peptide mutants substituted with Ala at Yaa4 (1-1_R4A and 1-3_R4A) did not show interactions with PEDF. In addition, in the assay using the collagen-binding-null PEDF mutant (PEDF D255N), no interactions were observed. By contrast, the interactions were maintained using heparin-binding-null PEDF mutant (PEDF R148A) as a bait protein. Besides, the two-hybrid interaction with HSP47 was observed for all peptides. Because we already demonstrated that HSP47 bound to host (Pro-Gly-Pro)-repeating sequences on the triple helix (Fig. 3), all peptides formed the triple helix in yeast cells at the assay temperature of 25°C. These results indicated that the two-hybrid selection from the peptide library yielded novel PEDF-binding sequences that are recognized in the same manner as those in the native collagen.

The second-round selection of PEDF-binding peptides from a focused library

Narrowing down the searchable chemical space after acquiring essential structural elements from the screening of a totally random (and hence broad) library is a useful strategy. All three positive clones obtained by the first selection for PEDF-binding peptides contained Arg at the Yaa4 position. Furthermore, the importance of this amino acid residue had been evidenced by the crystal structure of the PEDF-peptide complex (29). We thus further constructed the second library in which the Arg residue was fixed, and the other five residues at the Xaa and Yaa positions were randomized. The guest region of the library encodes the peptide with Yaa1-Gly-Xaa3-Arg4-Gly-Xaa6-Yaa7-Gly-Xaa9. The selection by two-hybrid interaction with PEDF from among 2.1

$\times 10^6$ clones yielded 52 additional ones, and 12 individual amino acid sequences were obtained as new candidates for PEDF-binding sequences (Table 1). The interaction between the peptides with these amino acid sequences and PEDF was confirmed by two-hybrid assay of the individual clones (Fig. 5). All clones generated blue colonies on the plates. This indicated that the AD-fusion peptides with the sequences obtained from the focused library bind to BD-fusion PEDF.

They are all novel amino acid sequences—none of them is found in the NCBI and UniProt databases. Overall, the conventional binding motif, Lys, Arg, or His at Yaa1 and Phe or Leu at Xaa6, was conserved in seven sequences, 2-4, 2-5, 2-7, 2-8, 2-9, 2-10, and 2-15 (14). Meanwhile, the other five sequences, 2-6, 2-11, 2-12, 2-13, and 2-14, did not have a basic amino acid at Yaa1, which suggested that these sequences would interact with PEDF in a different manner from α1(I)930–938.

Relative PEDF-binding affinity of the chemically synthesized peptides with the sequences obtained from the selections

The amino acid sequences obtained by the two-step yeast two-hybrid selections for PEDF binders were reconstructed as chemically synthesized triple-helical peptides, and their interactions with recombinant PEDF were evaluated. The peptides were designed as host-guest peptides with the sequence of H-Tyr-(Pro-Hyp-Gly)$_4$-Pro-[guest sequence]-Hyp-Gly-(Pro-Hyp-Gly)$_4$-NH_2. An N-terminal Tyr residue was introduced to facilitate the measurement of concentration as represented by the absorbance of UV. As a guest sequence, we used the native PEDF-binding sequence and the sequences obtained by the yeast two-hybrid selection, omitting Cys-containing ones to avoid possible oxidative oligomerization. We also prepared **pepfreq**, a peptide whose guest sequence at each position was composed of the amino acid residue most frequently occurring in the two-hybrid selection (Table 2).

The peptides were synthesized by Fmoc-based solid-phase synthesis and purified by RP-HPLC (*SI Appendix*, Fig. S4). The conformational state of the synthetic peptides was analyzed by circular dichroism (CD)-spectrometry. All of the peptides showed positive signals at approximately 225 nm, indicating that they formed triple helices (31). In addition, we monitored the $[\theta]_{225}$ signal with increasing temperature (Table 2 and *SI Appendix*, Fig. S5). This confirmed that the peptides formed triple helices at 25°C, set in the following PEDF-binding assay, because $[\theta]_{225}$ signals at 25°C retained the top positions in sigmoidal curves.

We performed a competitive enzyme-linked immunosorbent assay (ELISA) to evaluate the relative binding affinity of the synthetic peptides to PEDF. In this assay, a competitive decrease in the binding of glutathione *S*-transferase (GST)-PEDF-fusion protein to the type I collagen-coated wells by the addition of the peptides was quantified by immuno-enzymatic colorimetric detection. All of the peptides showed concentration-dependent inhibition for PEDF-collagen interaction (Fig. 6). The relative PEDF-binding affinity of the peptides was evaluated using the IC_{50} values derived from the binding inhibitory curves (Table 2 and Fig. 6). **Pep7** exhibited more potent binding affinity than **pepα1(I)930–938**, the peptide with the native binding sequence (p = 0.0287). **Pep10** and **pepfreq** exhibited binding affinities comparable to that of **pepα1(I)930–938**.

Prediction of a new salt bridge enhances the binding affinity: a docking model simulation

Some of the sequences obtained from the selections, such as 2-6, did not have basic amino acids at Yaa1 but at Xaa3, which does not correspond to the conventional PEDF-binding motif, Lys-Gly-Xaa-Arg-Gly-Phe-Yaa-Gly-Leu. In addition, **pep7** and **pep10** showed higher binding affinity to PEDF than **pepα1(I)930–938**. To investigate the propensity of these peptides to bind to PEDF, we calculated the optimized complex structure of PEDF with **pep6** (Asn1-Gly2-Arg3), **pep10** (Arg1-Gly2-Pro3), and **pep7** (Lys1-Gly-Arg3) from the crystallographic data by

molecular modeling. The calculation suggested that Arg3 in **pep6** and Arg1 in **pep10** interacted with the PEDF acidic patch (red color in Fig. 7) that interacts with Lys1 in **pepα1(I)930–938** (Fig. 7A–C). In the calculation of **pep7**, Lys1 pointed toward the acidic patch, like Arg1 in **pep10**. Meanwhile, unlike Arg3 in **pep6**, Arg3 in **pep7** was found to form a new salt bridge with the other acidic amino acids in PEDF (Fig. 7D). This suggests that **pep7** binds to PEDF in a different manner from **pepα1(I)930–938**.

Discussion

This work demonstrated the yeast two-hybrid selection of triple-helical peptides that bind to a certain target protein, PEDF, from random combinatorial libraries containing collagen-like Gly-Xaa-Yaa-repeating peptides. The first blind screening without prior knowledge of the binding peptides yielded a few peptides that bind to PEDF within 2 weeks. The structures of the selected peptides matched the previously identified PEDF-binding amino acid sequence motif as well as the structural requirements revealed by analysis of the co-crystal of the PEDF-triple-helical peptide complex (29). The second round of screening of the focused library, in which essential Arg residue was fixed at the fourth amino acid residue of the randomized portion, resulted in 12 additional peptides being obtained. As demonstrated here, the combinatorial use of the completely random first screening and the subsequent second selection from a focused library is a powerful strategy for identifying binding motifs that can be recognized by a POI. By fixing one essential amino acid residue, the coverage of screening theoretically rises 20-fold (in this study, the library coverage increased from 6.3% to 66%). All of the peptides obtained by these two screenings were novel ones that are not present in the context of natural collagen. **Pep7** and **pep10** exhibited higher affinity to the target protein than **pepα1(I)930–938**, the peptide with the natural PEDF-binding sequence (Table 2).

Screening from the random triple-helical library yielded several sequences that do not correspond to the conventional motif for binding to PEDF. The docking simulation suggests that Arg3 of **pep6** points toward the acidic patch, like Arg1 of **pepα1(I)930–938**. Meanwhile, although Arg1 of **pep7** retains the interaction with the acidic patch, Arg3 of **pep7** additionally interacts with a different PEDF acidic patch with which His3 of **pepα1(I)930–938** does not interact (Fig. 7). **Pep7** exhibited more potent binding to PEDF than **pepα1(I)930–938**, probably because of this additional interaction. This analysis suggests that the yeast two-hybrid system is

useful for finding non-natural motifs for binding to collagen-binding proteins.

The procedure of the screening based on the standard yeast two-hybrid technique is simple and easy to access without any purification or immobilization of proteins or peptides. Screening of 10^6 clones out of the theoretical diversity of 20^6 can be conducted in a single round. The system relies on the self-trimerizing propensity of the peptides consisting of collagenous Gly-Xaa-Yaa repeats, and on following oligomeric interaction with the target protein in the yeast nuclei. By taking advantage of the host-guest design that stabilizes the triple-helical conformation, we achieved an increase in the population of triple-helical molecules in the library. The melting temperature of **pep7** (T_m = 35.0°C) was lower than that of **pepα1(I)930–938** (T_m = 37.0°C), whose guest sequence is known as the most unstable region of type I collagen (18). This indicates that the peptide design enabled the library to contain even amino acid sequences that are too unstable to form a triple helix in the context of natural collagen. As the yeast two-hybrid system allows us to conduct screening at temperatures ranging from 16°C to 37°C (Fig. 3), the population of the triple-helical molecules in the library can be expected to increase by lowering the screening temperature. In addition, the two-hybrid interaction with HSP47 is applicable to confirm the triple-helical conformation of the peptides expressed in the yeast clones (Fig. 3).

Collagen Toolkits have contributed to identifying the amino acid sequences on collagen triple helices responsible for the protein interactions (4-11). From these previous studies, we realized that there is a large overlap in recognition among different collagen-binding proteins. As the most typical case, VWF, DDR, SPARC, and aegyptin, a mosquito salivary gland protein, were found to share the same binding site in native collagen (9-12,32). One reason for the peculiar interactions between collagen and proteins is the low diversity of amino acid sequences displayed on the triple helix of native collagen. As shown in Fig. 2, 226 Xaa-Yaa combinations appeared in the fibril-forming collagen out of 400 (= 20 × 20) possible combinations, and they tended not to

include triple-helix-destabilizing Xaa-Yaa combinations. The use of the more diverse combinatorial peptide libraries unrelated to the consequences of the molecular evolution of collagen raised expectations of obtaining selective binders for individual collagen-binding proteins. Such selective binders would be useful tools for resolving the complex intertwined interactions between protein and collagen and hence in elucidating collagen's physiological regulatory functions.

The triple-helical peptides that specifically bind to bioactive proteins have the potential to be used as lead compounds of drug development. Because chemically synthesized triple-helical peptides show remarkable stability in living animals by resisting attacks by proteases (33,34), the triple helix is a promising scaffold for peptide drugs. Alternatively, the triple-helical peptides with protein-binding sequences obtained by the screening can be used in artificial collagen-like polymer systems to make novel functional biomaterials (35-38). In addition, targets are not limited to the collagen-binding proteins; any proteins could serve as bait for the selection.

Acknowledgments

We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript. The work was supported by JSPS KAKENHI Grant Number 20K05756, and Waseda University Grant for Encouragement of Scientists.

References

1. R. A Berg, D. J. Prockop, The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem. Biophys. Res. Commun.* **52**, 115–120 (1973).
2. J. A. Ramshaw, N. K. Shah, B. Brodsky, Gly-X-Y tripeptide frequencies in collagen: a context for host–guest triple-helical peptides. *J. Struct. Biol.* **122**, 86–91 (1998).
3. B. Leitinger, E. Hohenester, Mammalian collagen receptors. *Matrix Biol.* **26**, 146–155 (2007).
4. S. M. Sweeney *et al.*, Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* **283**, 21187–21197 (2008).
5. N. Raynal *et al.*, Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J. Biol. Chem.* **281**, 3821–3831 (2006).
6. S. W. Hamaia *et al.*, Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin α1β1 using collagen toolkits II and III. *J. Biol. Chem.* **287**, 26019-26028 (2012).
7. S. W. Hamaia *et al.*, Unique charge-dependent constraint on collagen recognition by integrin α10β1. *Matrix Biol.* **59**, 80–94 (2017).
8. W. M. Zhang *et al.*, α11β1 integrin recognizes the GFOGER sequence in interstitial collagens. *J. Biol. Chem.* **28**, 7270–7277 (2003).
9. H. Xu *et al.*, Collagen binding specificity of the discoidin domain receptors: Binding sites on collagens II and III and molecular determinants for collagen IV recognition by DDR1. *J. Biol. Chem.* **286**, 26364–26374 (2011).
10. A. D. Konitsiotis *et al.*, Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **283**, 6861–6868 (2008).
11. T. Lisman *et al.*, A single high-affinity binding site for von Willebrand factor in collagen III,

identified using synthetic triple-helical peptides. *Blood.* **108**, 3753–3756 (2006).

12. C. Giudici *et al.,* Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens *J. Biol. Chem.* **283**, 19551–19560 (2008).
13. N. Yasui, T. Koide, Collagen−protein interactions mapped by phototriggered thiol introduction. *J. Am. Chem. Soc.* **125**, 15728–15729 (2003).
14. A. Sekiya, H. Okano-Kosugi, C. M. Yamazaki, T. Koide, Pigment epithelium-derived factor (PEDF) shares binding sites in collagen with heparin/heparan sulfate proteoglycans. *J. Biol. Chem.* **286**, 26364–26374 (2011).
15. S. M. Sweeney, C. A. Guy, G. B. Fields, J. D. S. Antonio, Defining the domains of type I collagen involved in heparin-binding and endothelial tube formation. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7275–7280 (1998).
16. J. E. Koda, A. Rapraeger, M. Bernfield, Heparan sulfate proteoglycans from mouse mammary epithelial cells. Cell surface proteoglycan as a receptor for interstitial collagens. *J. Biol. Chem.* **260**, 8157–8162 (1985).
17. N. K. Shah *et al.,* A host-guest set of triple-helical peptides: stability of Gly-X-Y triplets containing common non-polar residues. *Biochemistry.* **35**, 10262-10268 (1996).
18. A. V. Persikov, J. A. Ramshaw, B. Brodsky, Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343–19349 (2005).
19. T. Koide, A. Aso, T. Yorihuzi, K. Nagata, Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47. *J. Biol. Chem.* **275**, 27957–27963 (2000).
20. K. K. Fuji *et al.*, Lowering the culture temperature corrects collagen abnormalities caused by HSP47 gene knockout. *Sci. Rep.* **9**, 17433 (2019).
21. K. K. Fujii *et al.*, The thermal stability of the collagen triple helix is tuned according to the

environmental temperature. *Int. J. Mol. Sci.* **23**, 2040 (2022).

22. S. W. Manka *et al.*, Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12461–12466 (2012).
23. M. Tasab, M. R. Batten, N. J. Bulleid, Hsp47: a molecular chaperone that interacts with and stabilizes correctly-folded procollagen. *EMBO J.* **19**, 2204–2211 (2000).
24. T. Koide, Y. Takahara, S. Asada, K. Nagata, Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J. Biol. Chem.* **277**, 6178–6182 (2002).
25. T. Koide, S. Asada, K. Nagata, Substrate recognition of collagen-specific molecular chaperone HSP47. Structural requirements and binding regulation. *J. Biol. Chem.* **274**, 34523–34526 (1999).
26. F. R. Steele, G. J. Charder, L. V. Johnson, J. Tombran-Tink, Pigment epithelium-derived factor: neurotrophic activity and identification as a member of the serine protease inhibitor gene family. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1526–1530 (1993).
27. S. P. Becerra, A. Sagasti, P. Spinella, V. Notario, Pigment epithelium-derived factor behaves like a noninhibitory serpin. Neurotrophic activity does not require the serpin reactive loop. *J. Biol. Chem.* **270,** 25992–25999 (1995).
28. S. Filleur, T. Nelius, W. de Riese, R. C. Kennedy, Characterization of PEDF: a multi-functional serpin family protein. *J. Cell. Biochem.* **106**, 769–775 (2009).
29. K. Kawahara *et al.*, Spatiotemporal regulation of PEDF signaling by type I collagen remodeling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11450–11458 (2020).
30. N. Yasui *et al.*, Dual-site recognition of different extracellular matrix components by anti-angiogenic/neurotrophic serpin, PEDF. *Biochemistry.* **42**, 3160–3167 (2003).
31. A. V. Persikov, J. A. Ramshaw, B. Brodsky, Collagen model peptides: Sequence dependence

of triple-helix stability. *Biopolymers.* **55**, 436–450 (2000).

32. E. Calvo *et al.*, Aegyptin displays high-affinity for the von Willebrand factor binding site (RGQOGVMGF) in collagen and inhibits carotid thrombus formation in vivo. *FEBS J.* **277**, 413–427 (2010).
33. H. Yasui *et al.*, Potential of collagen-like triple helical peptides as drug carriers: Their in vivo distribution, metabolism and excretion profiles in rodents. *Biopolymers.* **100**, 705–713 (2013).
34. T. Koide, N. Yamamoto, K. B. Taira, H. Yasui, Fecal excretion of orally administered collagen-like peptides in rats: contribution of the triple-helical conformation to their stabilities. *Biol. Pharm. Bull.* **39**, 135–137 (2016).
35. T. Koide, D. L. Homma, S. Asada, K. Kitagawa, Self-complementary peptides for the formation of collagen-like triple helical supramolecules. *Bioorg. Med. Chem. Lett.* **15**, 5230–5233 (2005).
36. F. W. Kotch, R. T. Raines, Self-assembly of synthetic collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3028–3033 (2006).
37. C. M. Yamazaki *et al.*, A collagen-mimetic triple helical supramolecule that evokes integrin-dependent cell responses. *Biomaterials.* **31**, 1925–1934 (2010).
38. S. F. Ichise *et al.*, Development of a collagen-like peptide polymer via end-to-end disulfide crosslinking and its application as a biomaterial. *Acta Biomaterialia.* **94**, 361–371 (2019).

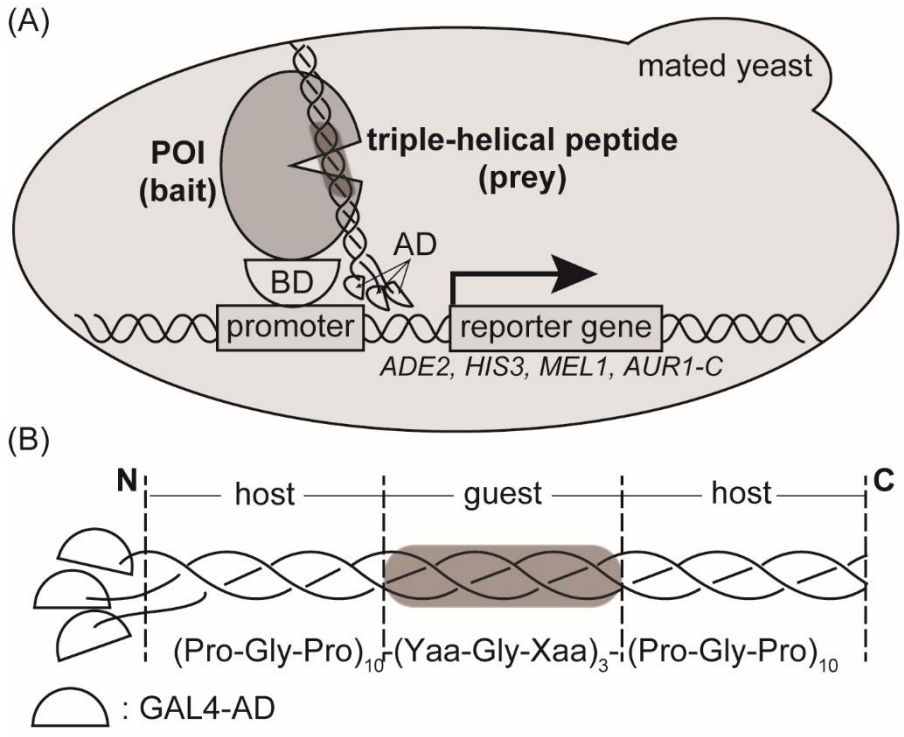
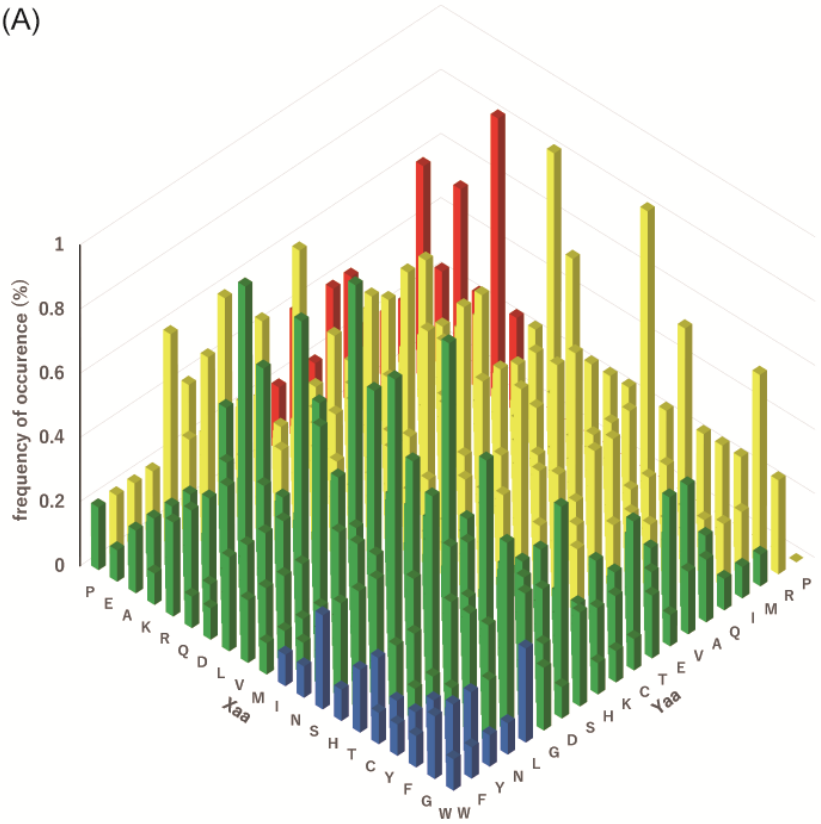


Fig. 1. Design of the two-hybrid system for selecting protein-binding triple-helical peptides in yeast cells. (A) Mechanism of the two-hybrid assay detecting the binding between a protein of interest (POI) and a triple-helical peptide in a yeast cell. The interaction brings GAL4-AD and GAL4-BD into proximity, leading to expression of the reporter genes. (B) A construct of a GAL4-AD-fusion triple-helical peptide with a random region. Xaa and Yaa are randomized by setting the corresponding codons to NNK codons.

(A)


frequency of occurence (%)
1
0.8
0.6
0.4
0.2
0
P E A K R Q D L V M I N S H T C Y F G W W F Y N L G D S H K C T E V A Q I M R P
Xaa
Yaa

(B)

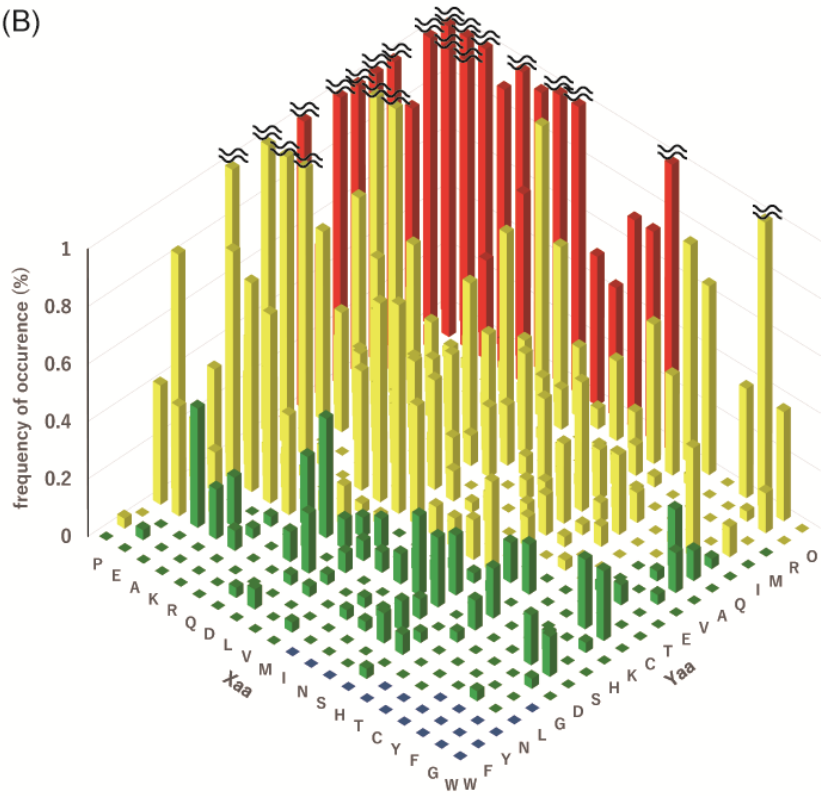

frequency of occurence (%)
1
0.8
0.6
0.4
0.2
0
P E A K R Q D L V M I N S H T C Y F G W W F Y N L G D S H K C T E V A Q I M R O
Xaa
Yaa

Fig. 2. Distribution of Gly-Xaa-Yaa triplets in (A) the guest region of the yeast library and (B) the triple-helical region of human fibrillar collagen. Each triplet is categorized into four groups according to its predicted or experimentally observed effect on the thermostability of the triple helix, as previously reported (18). Here, the effect on the thermal stability, ΔT_m^{GPO}, is defined as the difference of melting temperature of the triple helix (T_m) between $(Gly\text{-}Pro\text{-}Hyp)_8$ and peptides in which a Gly-Pro-Hyp triplet is replaced with a corresponding Gly-Xaa-Yaa: Red: $-10°C<\Delta T_m^{GPO}$, yellow: $-20°C<\Delta T_m^{GPO}<-10°C$, green: $-30°C<\Delta T_m^{GPO}<-20°C$, and blue $\Delta T_m^{GPO}<-30°C$. Maximum values on the vertical axis are (A) 1% and (B) 12%. A reduced view of Fig. 2B is shown in Fig. S3B (*SI Appendix*).

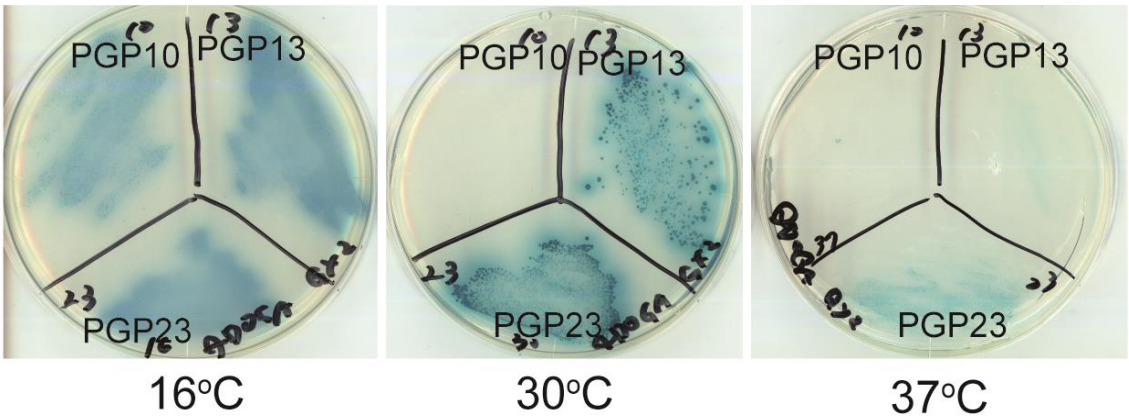


Fig. 3. Two-hybrid interaction of HSP47 with (Pro-Gly-Pro)-repeating peptides. PGP10, PGP13, and PGP23 are GAL4-AD-fusion $(Pro\text{-}Gly\text{-}Pro)_{10}$, $(Pro\text{-}Gly\text{-}Pro)_{13}$, and $(Pro\text{-}Gly\text{-}Pro)_{23}$, respectively. They were cultured at 16, 30, or 37°C for 3 days.

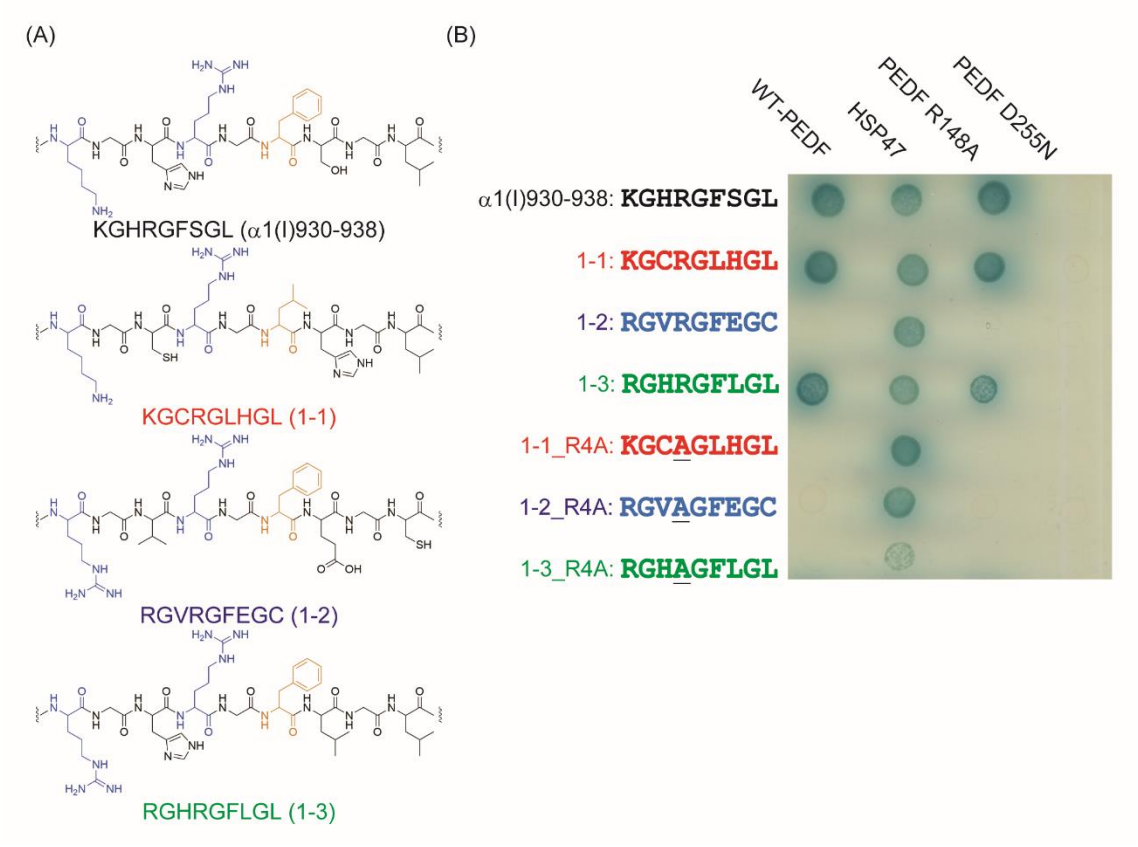


Fig. 4. Selection of PEDF-binding triple-helical peptides. (A) The guest sequences of the clones obtained by the selection and native PEDF-binding sequence in type I collagen. (B) The two-hybrid interactions of the AD-fusion peptides with the sequences obtained and their mutants with BD-fusion WT-PEDF, its mutants (R148A and D255N), and HSP47. The cells were cultured at 25°C for 5 days. The sequences of peptides are shown as one-letter codes of amino acids. In the sequences, mutated amino acids are underlined.

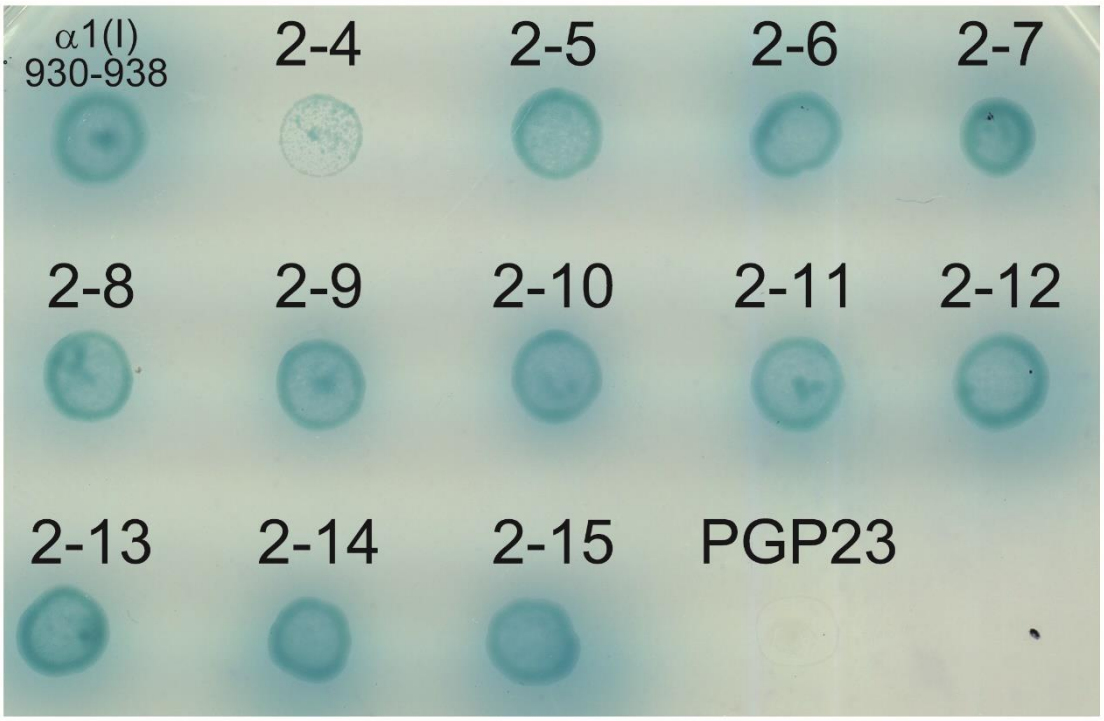


Fig. 5. Interaction of the AD-fusion peptides with the sequences obtained from the second selection with BD-fusion PEDF in yeast cells. PGP23 is GAL4-AD-fusion $(PGP)_{23}$. The cells were cultured at 25°C for 5 days.

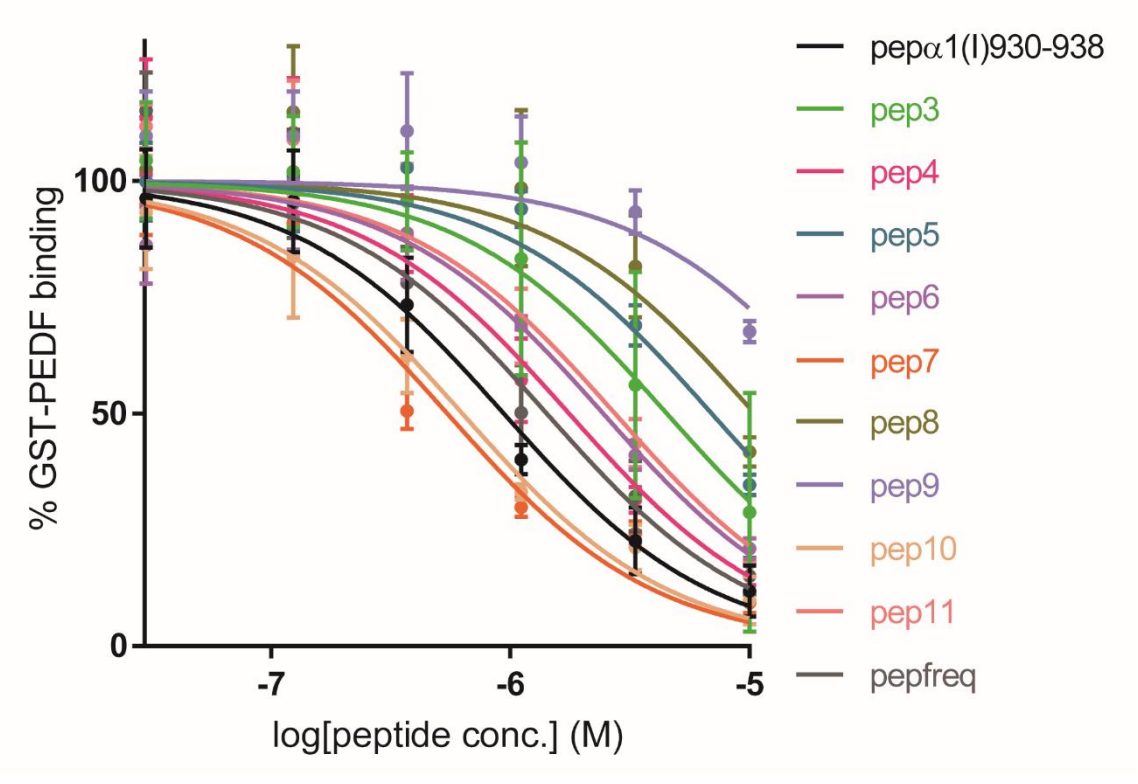


Fig. 6. Competitive binding of GST-PEDF to surface-immobilized collagen in the presence of the synthetic peptides. n = 3, mean ± S.D.

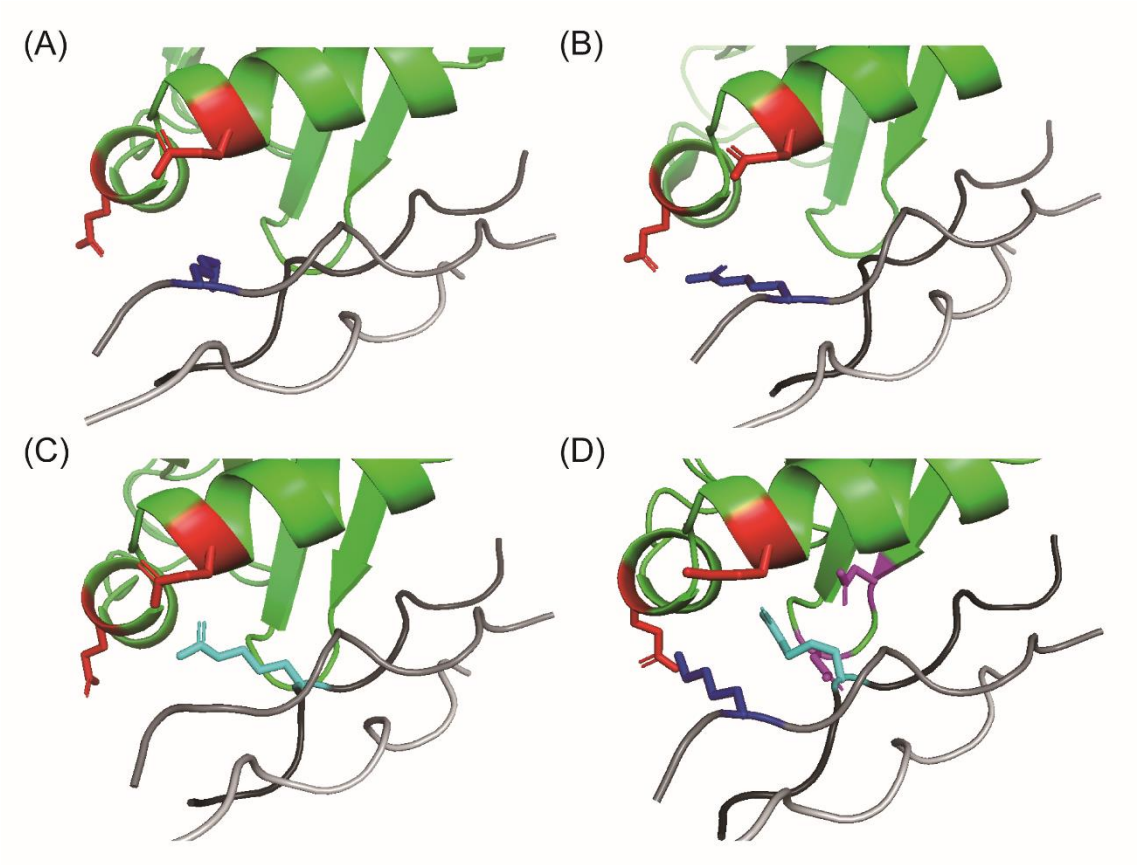


Fig. 7. Calculated models of the interactions between PEDF and the peptides obtained [(A) **pepα1(I)930–938**, (B) **pep10**, (C) **pep6**, and (D) **pep7**]. PEDF and the peptides are colored green and gray, respectively. Glu290 and Glu295 in PEDF form the acidic patch with which Lys1 in **pepα1(I)930–938** interacts and are colored red. Asp255 and Asp257 in PEDF are colored magenta. Lys1/Arg1 and Arg3 are colored blue and cyan, respectively. The model was calculated from the reported crystal structure (29, PDB ID: 6LOS).

Table 1. Guest sequences of the clones obtained from the focused second library

Name	Guest sequence	Frequency
2-4	Arg-Gly-Ala-Arg-Gly-Leu-His-Gly-Leu	29/52
2-5	Arg-Gly-Pro-Arg-Gly-Leu-Thr-Gly-Phe	8/52
2-6	Asn-Gly-Arg-Arg-Gly-Phe-Met-Gly-Met	3/52
2-7	Lys-Gly-Arg-Arg-Gly-Phe-His-Gly-Leu	2/52
2-8	Arg-Gly-Pro-Arg-Gly-Leu-Leu-Gly-Leu	2/52
2-9	Arg-Gly-Phe-Arg-Gly-Leu-Met-Gly-Leu	1/52
2-10	Arg-Gly-Pro-Arg-Gly-Phe-Thr-Gly-Phe	1/52
2-11	Val-Gly-Arg-Arg-Gly-Leu-Ser-Gly-Leu	1/52
2-12	Thr-Gly-Arg-Arg-Gly-Leu-Cys-Gly-Leu	1/52
2-13	Cys-Gly-Val-Arg-Gly-Leu-Thr-Gly-Arg	1/52
2-14	Cys-Gly-Val-Arg-Gly-Leu-Val-Gly-Met	1/52
2-15	His-Gly-Arg-Arg-Gly-Leu-Cys-Gly-Leu	1/52

Table 2. Melting temperature of the triple helix (T_m) and IC_{50} values of the synthetic peptides H-Tyr-(Pro-Hyp-Gly)$_4$-Pro-[guest sequence]-Hyp-Gly-(Pro-Hyp-Gly)$_4$-NH_2

Name	Guest sequence	T_m (°C)	IC_{50}[a] (μM)
pepα1(I)930-938	Lys-Gly-His-Arg-Gly-Phe-Ser-Gly-Leu	37.0	0.94 ± 0.12
pep3	Arg-Gly-His-Arg-Gly-Phe-Leu-Gly-Leu	43.6	4.7 ± 1.2
pep4	Arg-Gly-Ala-Arg-Gly-Leu-His-Gly-Leu	47.6	1.8 ± 0.3
pep5	Arg-Gly-Pro-Arg-Gly-Leu-Thr-Gly-Phe	53.8	7.0 ± 0.8
pep6	Asn-Gly-Arg-Arg-Gly-Phe-Met-Gly-Met	37.0	2.5 ± 0.3
pep7	Lys-Gly–Arg-Arg-Gly-Phe-His-Gly-Leu	35.0	0.55 ± 0.66
pep8	Arg-Gly-Pro-Arg-Gly-Leu-Leu-Gly-Leu	51.0	>10
pep9	Arg-Gly-Phe-Arg-Gly-Leu-Met-Gly-Leu	50.2	>10
pep10	Arg-Gly-Pro-Arg-Gly-Phe-Thr-Gly-Phe	57.6	0.63 ± 0.74
pep11	Val-Gly-Arg-Arg-Gly-Leu-Ser-Gly-Leu	41.8	2.8 ± 0.5
pepfreq	Arg-Gly-Arg-Arg-Gly-Leu-His-Gly-Leu	44.4	1.4 ± 0.2

[a] IC_{50} values are mean ± S.E.M.