biomolecules

MDPI

Article

Identification of HSP47 Binding Site on Native Collagen and Its Implications on The Development of HSP47 Inhibitors

Haiyan Cai[1,#], Parvathy Sasikumar[2,#], Gemma Little[2], Dominique Bihan[3], Samir W. Hamaia[3], Aiwu Zhou[1], Jonathan M. Gibbins[2,*] and Richard W. Farndale[3,*§]

[1] Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai JiaoTong University, School of Medicine, Shanghai, China; hycai@shsmu.edu.cn (H. C.), awz20@shsmu.edu.cn (A.Z.)
[2] Institute for Cardiovascular & Metabolic Research, School of Biological Sciences, University of Reading, Health and Life Sciences Building, Whiteknights, Reading, RG6 6EX, UK; j.m.gibbins@reading.ac.uk (J.M.G), gemma.little@reading.ac.uk (G.L.), p.sasi-kumar@imperial.ac.uk (P.S.)
[3] Department of Biochemistry, University of Cambridge, Downing Site, Cambridge CB2 1QW, United Kingdom; dominique.bihan@ucalgary.ca (D.B.), swh23@cam.ac.uk (S.W.H.)
§ Present address: CambCol Laboratories, Ely, UK
* Correspondence: rwf10@cam.ac.uk (R.W.F)
# These authors contributed equally to this work

Citation: Cai, H.; Sasikumar, P.; Little, G.; Bihan, D.; Hamaia, S.W.; Zhou, A.; Gibbins, J.M.; Farndale, R.W. Identification of HSP47 Binding Site on Native Collagen and Its Implications on The Development of HSP47 Inhibitors. *Biomolecules* **2021**, *11*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

Abstract: HSP47 (Heat Shock Protein 47) is a collagen-specific molecular chaperone that is essential for procollagen folding and function. Previous studies have shown that HSP47 binding requires a critical Arg residue at the Y position of the (Gly-Xaa-Yaa) repeats of collagen, however, the exact binding sites of HSP47 on native collagens are not fully defined. To address this, we mapped the HSP47 binding sites on collagens through an ELISA binding assay using collagen Toolkits, synthetic collagen peptides covering the entire amino acid sequences of collagen types II and III assembled in triple-helical conformation. Our results showed that HSP47 binds to only a few of the GXR motifs in collagen and most of the HSP47 binding sites identified are located near the N-terminal part of collagen triple-helical region. Molecular modelling and binding energy calculation indicated that residues flanking the key Arg in the collagen sequence also play an important role in defining the high affinity HSP47 binding site of collagen. Based on this binding mode of HSP47 to collagen, virtual screening targeting both the Arg binding site and its neighboring area on the HSP47 surface, and subsequently bioassay, identified two novel compounds with blocking activity towards HSP47 binding of collagen. Overall, our study revealed the native HSP47 binding sites on collagen and provided novel information for the design of small molecule inhibitors of HSP47.

Keywords: HSP47 inhibitor; collagen; fibrosis; molecular docking; structural analysis

1. Introduction

Collagen is the most abundant protein in mammals and is critical for forming specialized extracellular networks that bind cells together. The folding, processing, and assembly of collagen is tightly regulated in eukaryotic cells. Procollagen, the precursor molecule of collagen, undergoes extensive posttranslational processing to assemble into a triple-helical collagen, following which mature collagen is then secreted [1,2]. During these processes, several molecular chaperones and enzymes such as prolyl hydroxylase and heat shock protein 47 (HSP47) are involved.

HSP47 is a member of the serpin family, but it lacks serine protease inhibitory activity [3]. It normally resides in the endoplasmic reticulum (ER) and functions as a collagen-specific molecular chaperone [4–6]. HSP47 associates transiently with procollagen in the ER, thereby preventing premature interactions between nascent procollagen molecules,

and dissociates when procollagen transfers to cis-Golgi [7]. HSP47 knockout mice are embryonic lethal 11.5 days post-coitus [8], and aberrant formation of triple-helical collagen I molecules as well as defects in collagen production and basement membrane formation have been detected in these embryos. This indicates that HSP47 is essential for the correct folding of procollagen.

However, overexpression of HSP47 is associated with abnormal deposition of collagens in the extracellular matrix (ECM) with the onset of various fibroses including liver cirrhosis, lung, bone marrow, and idiopathic pulmonary fibroses [9,10]. No specific treatment is currently available for these fibrotic diseases which often lead to organ or tissue failure [11]. Nevertheless, the progression of fibrosis could be markedly reduced when HSP47 expression is suppressed using siRNA-mediated knockdown [12,13]. Altered HSP47 expression level is also found to correlate with the development of several types of cancer, such as cervical, breast, pancreatic and gastric cancers [13,14]. Silencing HSP47 significantly inhibits cell invasion in breast cancer cells and inhibited tumor growth in the xenograft mammary tumor model [4,15]. Our recent studies showed that HSP47 is present on the platelet surface and is involved in the recognition and response to collagen [16]. It is plausible that aberrant expression of HSP47 may contribute to the onset of various cardiovascular diseases. As HSP47 is the only collagen-specific molecular chaperone, targeting its binding interactions with collagen could be an effective therapeutic approach in managing these HSP47 overexpression-related diseases. Several types of small molecules have been developed as HSP47 inhibitors [17–20]. Recently, one of these compounds (COM IV) has been reported to effectively prevent collagen synthesis, cell viability, and migration of lung fibroblasts [21].

There have been extensive biochemical and structural studies on the binding interactions between HSP47 and collagen using short synthetic collagen-mimetic peptides [22–24]. These showed that HSP47 binds to triple-helical peptides containing a key Arg residue at the Y position of collagenous repeats (Gly-Xaa-Yaa). These results were confirmed by X-ray crystal structures of collagen model peptides (designated CMPs) complexed with canine HSP47 [25]. However, the exact binding sites of HSP47 on native collagens are not fully defined. For example, there are more than 40 arginine residues at similar Y positions in collagen II or III, and it is unclear if HSP47 could bind them all. To clarify this, here we have mapped the HSP47 binding sites using collagen Toolkits and identified HSP47 binding sites on collagen II and III. This provides novel information on the binding mode of HSP47 on collagen and sheds light on the development of novel HSP47 inhibitors.

2. Materials and Methods

2.1. Expression and purification of HSP47

Genes encoding human HSP47 (amino acids 19-418) were cloned in pET-28a. Proteins were expressed with C-terminal His-tag in *E. coli* BL21 (DE3). Cells were cultured at 37°C in LB and induced with 0.5 mM IPTG (isopropyl-β-D-thiogalactoside) until A_{600} reached 1. After a further 16 h incubation at 25°C, cells were collected by centrifugation, then resuspended in lysis buffer (20 mM Tris-HCl pH 7.4, 0.5 M NaCl) and lysed by high pressure. His-tagged HSP47 in the supernatant was isolated by nickel-affinity chromatography (GE Healthcare). The column was washed with wash buffer (20 mM Tris-HCl pH 7.4, 0.5 M NaCl, 20 mM imidazole) and step-eluted with elution buffer (20 mM Tris-HCl pH 7.4, 0.5 M NaCl, 300 mM imidazole). The peak fractions were pooled and dialyzed against final buffer (20 mM Tris-HCl pH 7.4, 0.5 M NaCl), and purity determined by SDS–PAGE.

2.2. Binding of HSP47 to Toolkit peptides

The peptides of the collagen II and III Toolkits were synthesized on TentaGel R Ram resin using a CEM Liberty microwave-assisted peptide synthesizer as described previously [23,26]. Briefly, the host-guest strategy was applied, in which the guest sequence,

native collagen sequence of interest, is placed between $(GPP)_5$ hosts, the flanking sequences that impart triple-helical conformation on the whole peptide. Thus, Toolkits II and III encompassed the entire collagen domains of human collagens II and III, respectively. Binding of HSP47 to the Toolkit peptides was determined colorimetrically by solid-phase binding assay (SPBA). Peptides were coated at 10 μg/mL overnight at 4 °C on Immulon 2HB 96-well plates (Thermo Scientific) and blocked for 1 h with 175 μl of binding buffer (TBS containing 5% BSA) before the addition of 100 μl of binding buffer containing 1 μg/ml recombinant HSP47 for 1 h at room temperature. Wells were washed five times with 200 μl of binding buffer (TBS containing 0.1% BSA w/v). Then binding buffer added containing HRP (Horseradish Peroxidase)-conjugated anti-His antibody at 1:1000 dilution and incubated for 1 h at room temperature. After washing, color was developed using a TMB (3,3′,5,5′- Tetramethylbenzidine) Substrate Kit (Pierce) according to the manufacturer's instructions.

2.3. Virtual screening of small inhibitors targeting human HSP47/collagen binding interface

As the sequence similarity between canine and human HSP47 is 99%, and residues involved in collagen binding are all conserved, human HSP47 structure was modelled by swiss-model [27] using canine HSP47 crystal structure (PDB ID: 3ZHA) [25] as template. The triple-helical structure of HSP47-integrin peptide I was built by software FoldX [28] based on the structure of HSP47-CMP complex (PDB ID: 3ZHA) and our previously reported integrin peptide I structure (PDB ID: 1Q7D).

To select potential HSP47 inhibitors, the SPECS database (http://www.specs.net) containing structural information of approximately 100,000 chemicals ($logP \leq 5.5$ & $logS \geq -5.5$) was adopted for virtual screening using software Glide6.9 (www.schrödinger.com). The docking model of HSP47 was prepared using the Protein Preparation and Grid Preparation tools in the Schrödinger Maestro interface. In the grid preparation process, the default settings were adopted for the cutoff, neutralization, scaling and dimensions of the binding pocket. The standard precision (SP) mode of Glide was used to explore favorable binding poses. Ligand conformation was allowed to be flexible while the protein was held as a rigid structure during the docking process. After the first round of screening, 10,000 compounds with the highest scores were selected for the second round of screening using the extra precision (XP) mode of Glide, leading to the selection of ~2,000 molecules. These molecules were filtered by AutoDock4.2 for binding conformation analysis [29]. The binding site on HSP47 was defined around the center of the collagen model peptide in the crystal structure (PDB ID: 3ZHA) [25], and was covered by preparing a 75×75×75 grid box with 0.375Å spacing between grid points. The Lamarckian genetic algorithm was applied to obtain the protein-ligand binding free energies and the spatial conformations of the bound compounds.

2.4. Turbidity assay

To analyze the inhibitory activity of the compounds against HSP47, turbidity assay was carried out as described by Thomson [17]. Collagen from calf skin (Sigma, C9791) was solubilized as follows: Collagen was dissolved in 0.01 M acetic acid and stirred at 4 °C for 48 hours to obtain 6 mg/mL collagen solution. On the day of experiment, the collagen solution was diluted by the reaction buffer (PBS: 20 mM phosphate buffer/50 mM NaCl, pH 7.4) on ice and then added to each well of a 96-well plate on ice. HSP47 protein together with compounds was then added to give final HSP47 concentration of 1-5 μM and final collagen concentration of 0.2-1.2 mg/mL. Turbidity was measured at 313 nm for a period about 90 min at 34 °C. The IC_{50} value for each compound to block the HSP47/collagen interaction was determined by nonlinear regression using GraphPad Prism software (San Diego, CA, USA).

2.5. Binding affinity measurement of HSP47 with small molecular inhibitor

The binding affinity between HSP47 and small molecular inhibitors was assessed using the LabelFree Microscale thermophoresis (MST) assay. The concentration of HSP47 was kept at 1 μM, while the concentration of the compound Hs1 was varied between 0.15 μM- 78 μM. The assay was performed in PBS buffer containing 0.05% Tween-20. The final concentration of DMSO in the assay was 5%. After a short incubation, the samples were loaded into MST NT.115 LabelFree standard capillaries and the MST analysis was performed using Monolith NT.115 LabelFree. Concentrations on the x-axis are plotted in μM.

2.6. Molecular docking of small inhibitors on HSP47

The binding modes between HSP47 and the small molecule inhibitors were analyzed using software AutoDock4.2, the same parameters were used as in the virtual screening. The results showed that Col003 and its analogs and the compounds reported by others preferred binding to the large subpocket1 near Asp385 on HSP47. However, for compound Hs1, docking showed part of the molecule binding to the small subpocket2 that is close to Arg222 on HSP47. Therefore, we used software AutoDock Vina [5] to confirm the docking results. The binding site on HSP47 was also centered on the collagen model peptide in the crystal structure (PDB: 3ZHA and the search spaces was set as 30 x 30 x 30 Å, other parameters were set at default values. The binding conformations of the ligands for HSP47 were selected for further analysis by taking account of both the predicted binding free energy and the number of conformations identified.

3. Results and discussion

3.1. Binding of HSP47 to the collagen Toolkit peptides

To identify the HSP47 binding site on native collagen II and III, we screened HSP47 binding sites on collagen using the collagen Toolkits II and III. These kits contain sets of overlapping triple-helical peptides (THPs) as briefly described above [23]. Each peptide contains a guest sequence of 27 amino acids, the C-terminal 9 amino acids of which form the first 9 guest amino acids of the next peptide. Thus, the Toolkit advances 18 amino acids along the triple-helical sequence of human collagen II or III with each successive peptide, and a 9-amino acid overlap is included between adjacent peptides. These kits have been successfully used in identifying many ligand binding motifs on collagen [23,30-31]. Recombinant HSP47 with a C-terminal His-tag was incubated in wells of Immulon 2HB 96-well plates coated with Toolkit peptides. Some integrin-binding collagen peptides ($GX^4O^3GE^1R^0$) with an Arg at position Y^0 and different residues at position X^4, were selected as control peptides [32-34].

Significant binding of HSP47 could be detected from 12 peptides，9 derived from collagen II and 3 from collagen III (Table 1, Figure 1). Amongst them, four collagen II peptides (II-13, II-14, II-20, II-26) and one collagen III peptide (III-5) showed a stronger binding signal. Interestingly, most of these peptides are located near the N-terminus of collagen. As procollagen triple helix assembly occurs from C- to N-terminus, HSP47 is likely to play a stabilization role only at the late stage of procollagen triple helix assembly. All these identified HSP47-binding peptides contain the 'Gly-Xaa-Arg' or 'GXR' region, which is consistent with previous findings that the arginine residue in these THPs is critical for HSP47 binding. However, there are more than 40 'GXR' regions in either collagen II or III and only 12 of them in total showed appreciable HSP47 binding (Figure S1, Figure S2). Notably, four of the five integrin binding peptides containing 'GXOGER' motifs showed strong HSP47 binding while HSP47 binding on the fifth peptide with R at the X position was relatively weak (Table 1). Together, these observations indicate that the flanking sequences around the 'GXR' region also play an important role in defining an efficient HSP47 binding site. This is consistent with previous observations that residues at position^{-3} of collagen also affected the binding of HSP47 [31]. It should also be noted that the parent Toolkit peptides that contain the same HSP47-positive integrin-binding 'GXOGER' motifs do not generally support HSP47 binding. Peptides II-7/8, II-18/19, II-28

and II-31 contain GLOGER, GAOGER, GFOGER and GMOGER, respectively, but show negligible HSP47 binding. The same applies to III-4, III-8, III-29 and III-31. This indicates that nearby residues (-6 and -7) may restrict HSP47 binding, discussed further below.

Table 1. Identification of HSP47 binding THPs from the Toolkit II and III. *GPP* represents the start or finish of the [*GPP*]$_5$ host peptide, O represents hydroxyproline (Hyp). The GXR region that is predicted to play a key role in HSP47 binding is colored red. For the sequences that contain more than one 'GXR' motif, we highlighted the most plausible one according to our energy prediction results (Figure S4c). The integrin-binding THPs with sequence of 'GXOGER' flanked by [GPP]$_5$ at both ends where X (colored blue in the table) represents F, A, M, L or R were used as control peptides.

Peptide Name	Sequence	Mean A_{450}
Toolkit II		
14	*GPP***GPR**GPOGPQGATGPLGPKGQTGEOGIA*GPP*	>2
20	*GPP*GANGDOGROGEOGLO**GAR**GLTGROGDA*GPP*	>2
13	*GPP*GAKGSAGAOGIAGAOGFO**GPR**GPOGPQ*GPP*	1.5-2
26	*GPP***GER**GEQGAOGPSGFQGLOGPOGPOGEG*GPP*	1.5-2
11	*GPP*GARGPEGAQ**GPR**GEOGTOGSOGPAGAS*GPP*	1-1.5
10	*GPP*GGOGFOGAOGAKGEAGPT**GAR**GPEGAQ*GPP*	0.5-1
17	*GPP*GKRGARGEOGGVGPIGPO**GER**GAOGNR*GPP*	0.5-1
24	*GPP*GKAGEKGLOGAO**GLR**GLOGKDGETGAA*GPP*	0.5-1
39	*GPP***GAR**GAQGPOGATGFOGAAGRVGPOGSN*GPP*	0.5-1
Toolkit III		
5	*GPP***GER**GLOGPOGIKGPAGIOGFOGMKGHR*GPP*	1.5-2
30	*GPP*GAO**GLR**GGAGPOGPEGGKGAAGPOGPO*GPP*	0.5-1
14	*GPP*GIOGAOGLM**GAR**GPOGPAGANGAOGLR*GPP*	0.5-1
Integrin Peptide		
1	*GPP*G**F**O**GER***GPP*	>2
2	*GPP*G**M**O**GER***GPP*	1.5-2
3	*GPP*G**A**O**GER***GPP*	1.5-2
4	*GPP*G**L**O**GER***GPP*	1.5-2
5	*GPP*G**R**O**GER***GPP*	0-0.5

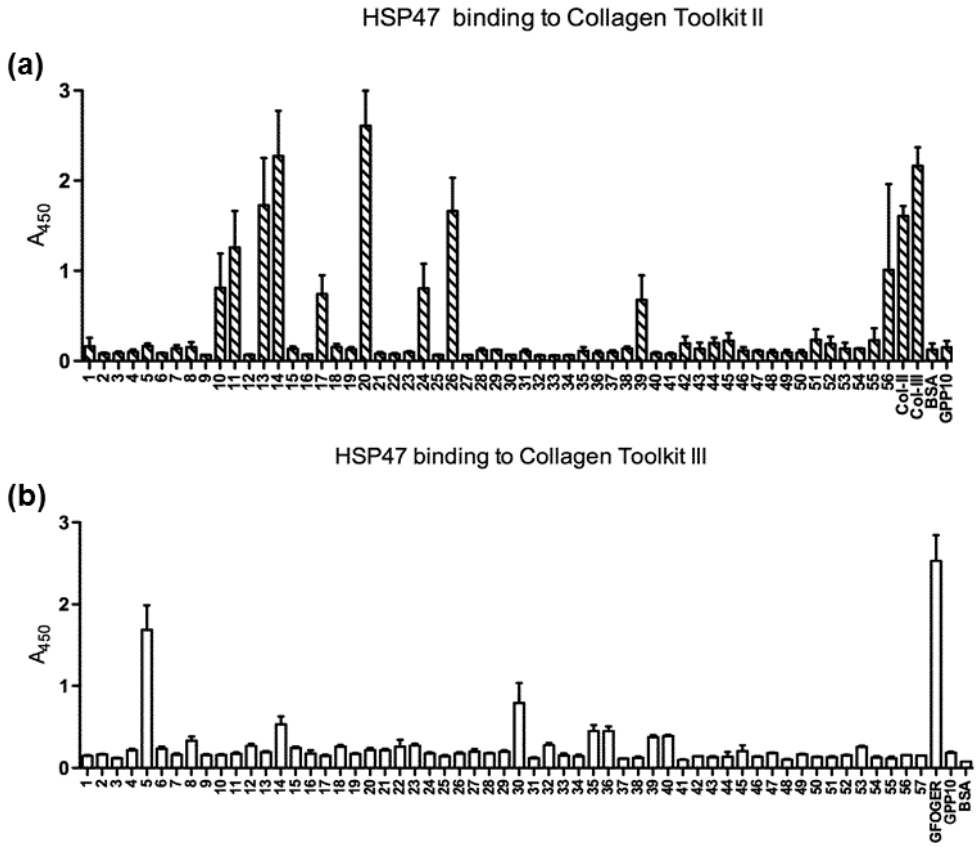


Figure 1. Binding of HSP47 to collagen Toolkits II and III, which encompassed the entire collagen domains of human collagens II and III, respectively. The recombinant HSP47 was incubated in wells of Immulon 2HB 96-well plates coated with peptides, and adhesion was measured as described under "Materials and Methods". "O" represents hydroxyproline (Hyp). Col, collagen. BSA, GPP10 are served as control surface coatings. Data are the mean + S.E. of three independent experiments.

3.2. Binding mechanism analysis between HSP47 and native collagen peptides

To elucidate the structural basis underlying the specific binding of these collagen peptides with HSP47, we analyzed their binding characteristics through structural modelling and binding energy calculation. A model of human HSP47 was built based on the reported crystal structure of canine HSP47 complexed with the triple-helical CMP (PPGP^{-7}P^{-6}GP^{-4}T^{-3}GP^{-1}R^{0}GPPGPPG, the key arginine residue is numbered as position 0). According to the crystal structure of the HSP47 and CMP complex, there are two copies of HSP47 molecules binding to a triple-helical collagen peptide in the asymmetric unit cell (PDB ID: 3ZHA). Each HSP47 molecule interacts mainly with one strand of the collagen and these two HSP47 molecules have similar binding interactions with the corresponding collagen strands. [25]. Therefore, the binding interactions between HSP47 and one strand of collagen were selected for subsequent analysis. The binding sites of CMP on human and canine HSP47 are almost identical as human and canine HSP47 are highly similar with only a few residues different. The CMP is 18 residues long and only 8 of them (P^{-7}P^{-6}GP^{-4}T^{-3}GP^{-1}R^{0}) are involved in direct interaction with the β-sheet C region (yellow colored) of HSP47 (Figure 2). The Arg^{0} in collagen formed a salt bridge to the conserved residue Asp385 in HSP47 (Figure 2b), which was essential for collagen binding as confirmed by mutagenic study [25].

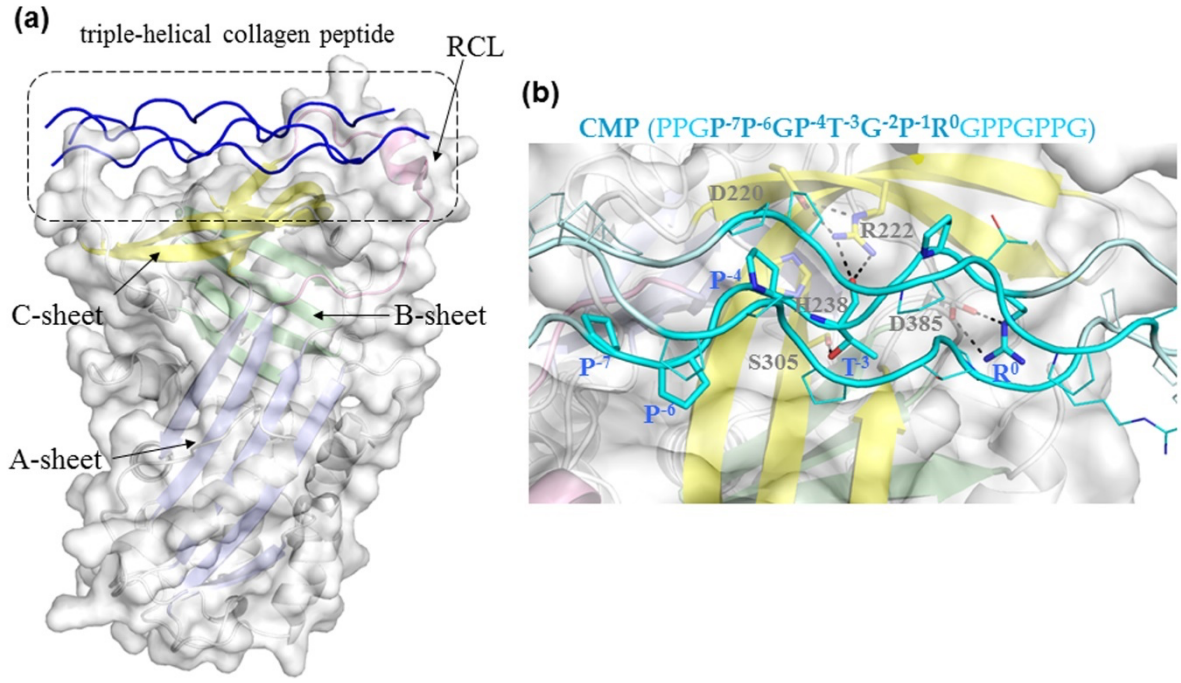


Figure 2. The model of human HSP47 complexed with triple helical collagen peptide. (**a**) The model was derived from the crystal structure of canine HSP47 complexed with collagen model peptide: CMP (PDB: 3ZHA[25]). HSP47 is shown in cartoon with three beta-sheets (A, B, C) colored in pale blue, green and yellow respectively and its reactive center loop is in pink. The surface of HSP47 is in pale grey. Triple-helical CMP is shown in blue cartoon. (**b**) Close-up of the interaction between human HSP47 and the collagen peptide (PPGP^{-7}P^{-6}GP^{-4}T^{-3}GP^{-1}**R**^{0}GPPGPPG, cyan cartoon and lines). Residues of both HSP47/CMP that are involved in interactions are shown in sticks.

Based on this model, we then analyzed the structure-activity relationship of the binding between collagen peptides and HSP47 using software FoldX, which is a commonly-used protein stability prediction algorithm and can estimate changes in protein folding free energy caused by residue variations [28]. We applied FoldX to analyze single residue variations at different positions in CMP to estimate changes of the folding free energy compared to WT complex. The lower folding free energy change (ΔΔG) indicates favorable binding of HSP47 and CMP. Here we first validated this method by calculating the binding energies of a set of collagen model peptides with variations at position -3 presented in a previous report where the bioactivities of these peptides were measured [31]. This showed the calculated energy scores of these peptides are largely consistent with the HSP47 binding activities (Figure S3), where the most optimal residue at position -3 is Pro or Thr. Even though the hydroxyl group and the main chain of Thr^{-3} formed polar interactions with residue Arg222 and Ser305 seen in the crystal structure of HSP47 (Figure 2b), Pro at this position could bind into the same pocket and has similar binding activities [31]. Other residues such as Hyp and Met at position -3 seemed compatible with HSP47 binding as well (Figure S3), while bulky or charged residue such as Phe, His, Arg, Trp and Tyr were not. This is consistent with the HSP47 binding peptides identified here from the Toolkits with Pro, Hyp, Met (III-14) or Thr (II-10) at position -3 (Table 1).

Subsequently FoldX was applied to analyze the impact of substitutions at position -4 on the HSP47-collagen binding energy. We found that Pro or Phe at this position binds better than Hyp, Leu, Met and Ala and molecular modelling of the integrin peptide 1 (GF^{-4}OGER^{0}) showed that Phe at position -4 could fit into a shallow surface cleft of HSP47 providing potential stabilizing interactions (Figure 3). Smaller hydrophobic residues such as Met and Leu seen in other integrin peptides could also be accommodated in this surface cleft while the charged bulky side chain of an arginine residue could not. This readily explains why HSP47 binds tightly on the integrin peptide 1 but very poorly on the integrin

peptide 5 (Table 1). However, there are differences between the experimental and predicted result for Ala in position -4, which is likely an outlier from the calculation by FoldX (Table 1, Figure3)

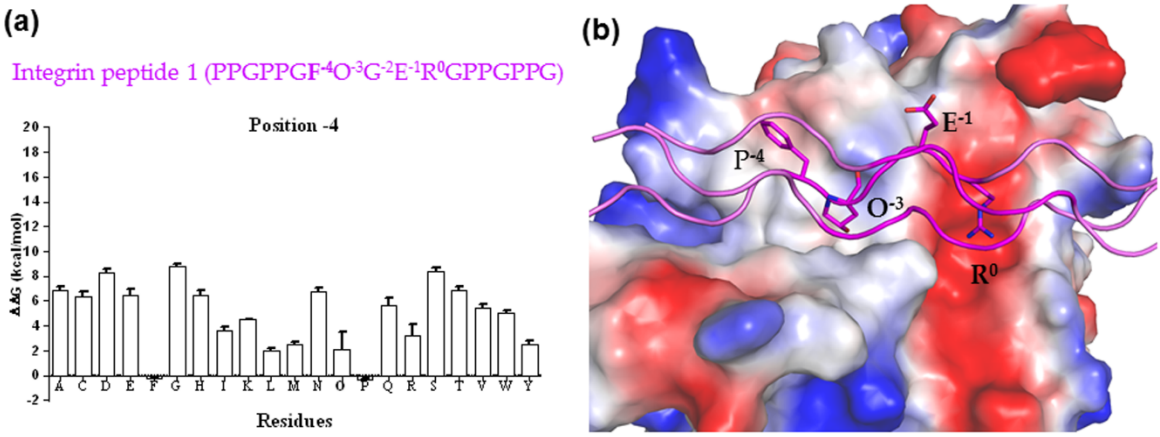


Figure 3. (**a**) Predicted total folding energies for the substitution of residues at position -4 in integrin peptide 1 (PPGP^{-7}P^{-6}GF^{-4}O^{-3}GE^{-1}R^{0}GPPGPPG) and HSP47 complex compared to that of WT (ΔΔG± SD, n=5, kcal/mol). The lower energy indicates favorable binding. (**b**) The predicted binding mode between triple helical integrin peptide 1 (shown in magenta cartoon and sticks) and HSP47 (shown in electrostatic potential surface).

We also assessed the effect of substitutions in positions -7 and -6 in CMP (PPGP^{-7}P^{-6}GP^{-4}T^{-3}GP^{-1}R^{0}GPPGPPG), which are located at the other end of the HSP47 binding site on collagen. It appeared that all the non-Pro substitutions showed worse binding to HSP47 (Figure S4). This may be due to these positions being far away from the key anchoring Arg^{0} and replacement of proline, which has a rigid sidechain, will increase the flexibility in this part of the triple-helical conformation, leading to lower binding activities of these peptides. Although not all the unbound peptides with GXR motif in Toolkits II and III could be readily explained by simple free folding energy calculation (Figure S2), the binding activities of collagen peptides towards HSP47 most likely reflect the overall cumulative effects of all the residues involved in the HSP47-peptide binding interface. Arg^{0} is the most critical residue in collagen, but residues at position -3 and -4 of collagen peptides play a critical supporting role for efficient HSP47 binding. This indicates that targeting the binding surface of HSP47 corresponding to positions -4, -3 and 0 of collagen simultaneously would provide a more efficient approach in blocking HSP47 activity than targeting the binding surface of HSP47 at position 0 only.

3.3. Screening and identification of HSP47 small molecular inhibitors

The binding mode between HSP47 and the collagen peptide shown above was used to design a virtual screening for potential HSP47 inhibitors, focusing on the footprint of collagen residues 0, -3 and -4 upon the surface of HSP47. After initial screening, we manually selected 58 compounds for subsequent inhibitory activity assay according to the diversity and potential interactive mechanisms between the compounds and HSP47. Collagen molecules in a buffer of physiological pH spontaneously associate to form fibrils [17,36], and the turbidity gradually increased in a concentration-dependent manner and then reached a plateau phase (Figure 4a). HSP47 could retard this fibril formation process within the range of 2.0-5.2 μM (Figure 4b). An HSP47 concentration of 2.6 μM was used for the compound activity assay (Figure 4c, d, e), and a previously reported HSP47 inhibitor, Com II, was selected as a positive control.

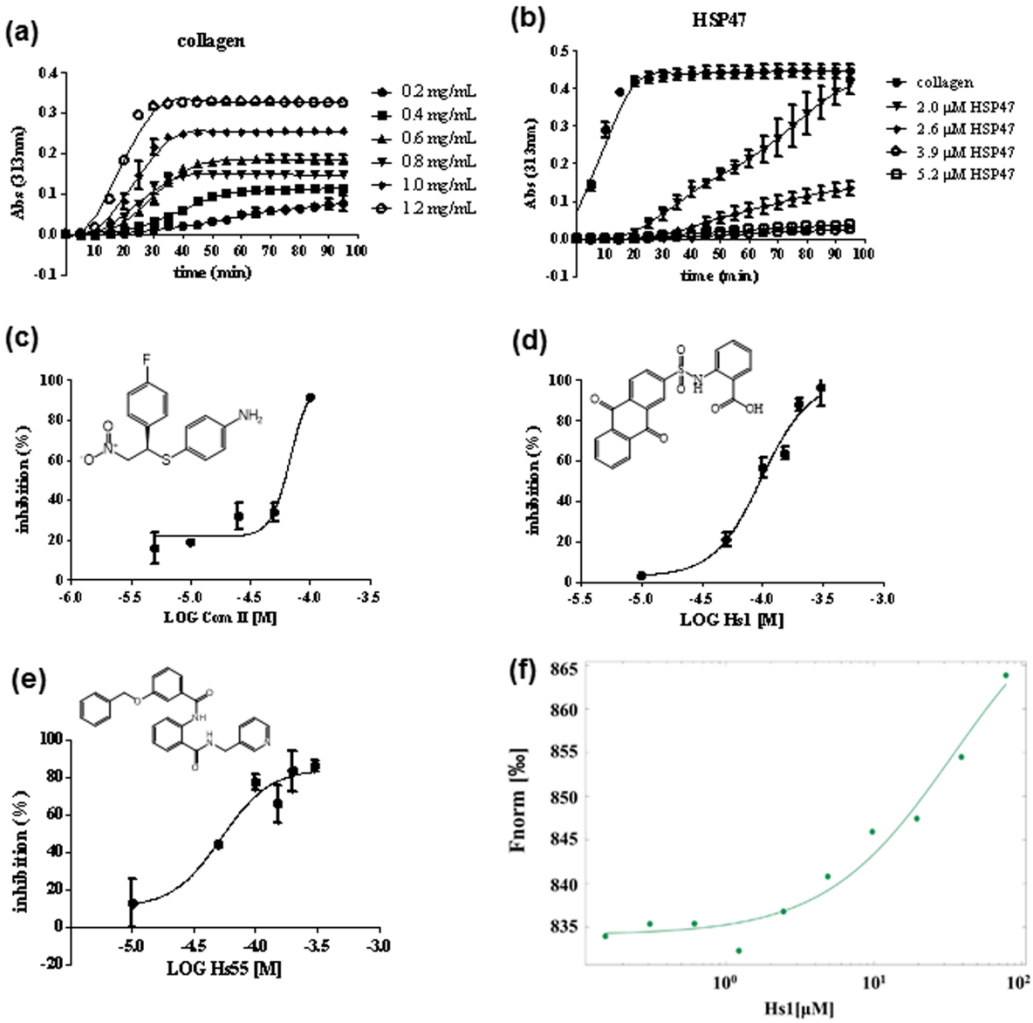


Figure 4. Screening and discovery of HSP47 small molecular inhibitors. Collagen fibril formation was measured in the absence (**a**) and presence (**b**) of HSP47. Collagen fibril formation was monitored at 313 nm in phosphate buffer. Inhibition effect of Com II (**c**), Hs1 (**d**) and Hs55 (**e**) against HSP47 in turbidity assay. (**f**) The binding activity of Hs1 against HSP47 in the LabelFree MST experiment.

Our measurement showed IC_{50} value for Com II is in the 50-100 μM range (Figure 4c), which is comparable to the reported value of 26.6 μM [17]. Amongst our selected compounds, compound Hs1 and Hs55 have most promising activities with IC_{50} values of 97 μM and 55 μM respectively (Figure 4d, 4e). As Hs1 had better water solubility, we determined its binding affinity to HSP47 using MicroScale Thermophoresis (MST). This revealed that Hs1 binds human HSP47 with a dissociation constant (Kd) of ~33 μM (Figure 4f), consistent with its blocking activity in inhibiting collagen fibril formation (Figure 4d).

3.4. Binding modes of small molecular inhibitors on HSP47

To predict how these inhibitors would bind on HSP47, we docked these molecules on HSP47 using softwares AutoDock4 and AutoDock Vina. The docking grid box was defined to include the whole collagen binding region on the HSP47 surface. Apart from Com II, Hs1 and Hs55, a recently identified HSP47 inhibitor Col003 and its analogues were also selected for docking [20,37]. The docking results indicated that Com II and Col003 bind to the surface area of HSP47 involved in binding 'GXR' of collagen (colored yellow in Figure 5, subpocket1). For Com II, the nitrogen atom of the aniline group forms a hydrogen bond with Asp385 and NO_2 group forms a hydrogen bond with Gln254 and the main chain of Tyr383, which might play essential roles for the compound binding. And

the fluorophenyl group of Com II stacks on the relatively hydrophobic area of HSP47 formed by Tyr245, Met271, Leu381, and Phe382. Col003 binds to a similar position, forming hydrogen bonds with His274 and main chain of His273 of HSP47 through its hydroxy group, and a hydrogen bond with the main chain of Val275 through the carbonyl group (Figure 5a, Figure S5). Its benzyl moiety binds the same hydrophobic patch of HSP47 as Com II (Figure 5b, Figure S5). In contrast, the binding of both Hs1 and Hs55 seems to involve both subpocket1 (Figure 5a) and the neighboring surface area for the docking of collagen residues at position -3 and -4 (pale blue, subpocket2, Figure 5a) mimicking a collagen peptide chain binding both these surface pockets simultaneously. The carboxyl group of Hs1 forms hydrogen bonds with Arg222 of HSP47 and the hydroxyl group on 9,10-dihydroanthracene forms a hydrogen bond with Asp385. At the same time, the 9,10-dihydroanthracene ring forms hydrophobic interactions with Met271, Leu381, Phe382 and Tyr383. Hs55 with a longer extended conformation docks in a similarly orientation as Hs1 and bridges both collagen binding pockets of HSP47 simultaneously. Although the binding mode of these compounds on HSP47 needs to be verified by structural studies, these docking results indicate that it is feasible to target two sub-pockets on the HSP47 surface at the same time. Subsequent chemical structure modification may improve the activities and specificity of these compounds for efficient HSP47 activity inhibition.

4. Conclusions

As HSP47 is an attractive therapeutic target for the treatment of fibrotic diseases, understanding of its physiological binding interactions would provide valuable information for the selection and design of small HSP47 inhibitors. Here we identified the HSP47 binding sites on native collagen II and III using the collagen Toolkits and showed collagen binding involves two neighboring subpockets on HSP47 surface. Subsequent structure-based drug screening through targeting these surface areas identified two novel small molecule HSP47 inhibitors, which might serve as lead compounds for the development of HSP47 inhibitors through chemical structure modification.

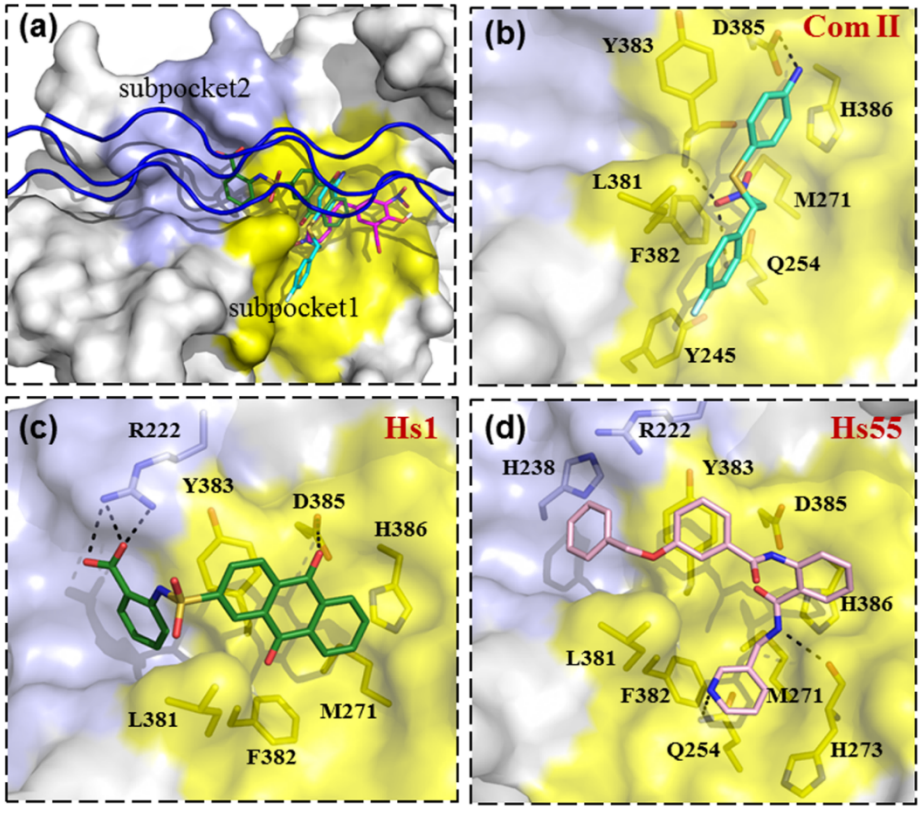


Figure 5. Predicted binding mode of HSP47 inhibitors. (**a**) HSP47 is shown as surface presentations and the collagen mode peptide (blue) is shown in cartoon. The Arg0 binding area on HSP47 is colored yellow (subpocket 1) while the binding area for collagen residues at position -4 and -3 is colored slate (subpocket 2). The potential binding interactions between HSP47 and the compound Col003 (a) COM II (**b)**, Hs1 (**c**) or Hs55 (**d**) were showed with the compounds in sticks and labeled. Compound Col003, Com II, Hs1 and Hs55 is shown in magenta, cyan, green and pink sticks, respectively. The black dashed lines present the hydrogen bonds.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1

Author Contributions: Conceptualization, Aiwu Zhou, Jonathan Gibbins and Richard Farndale; Data curation, Haiyan Cai; Funding acquisition, Richard Farndale and Aiwu Zhou; Investigation, Haiyan Cai and Parvathy Sasikumar; Methodology, Haiyan Cai, Parvathy Sasikumar, Gemma Little, Dominique Bihan and Samir Hamaia; Resources, Dominique Bihan; Supervision, Samir Hamaia, Aiwu Zhou, Jonathan Gibbins and Richard Farndale; Writing – original draft, Haiyan Cai; Writing – review & editing, Parvathy Sasikumar, Aiwu Zhou, Jonathan Gibbins and Richard Farndale.

Funding: This work was funded by National Natural Science Foundation of China,(81502981); The National Key Research and Development Program of China (2016YFA0500601); British Heart Foundation, grant numbers FS/17/31/32848, FS/11/86/29137 and RG/15/4/31268, and Wellcome Trust, Biomedical Resource Grant, 094470/Z/10/Z.

Conflicts of Interest: RF is Chief Scientific Officer of CambCol Laboratories, Ltd, Ely, UK. The authors declare no other conflict of interest.

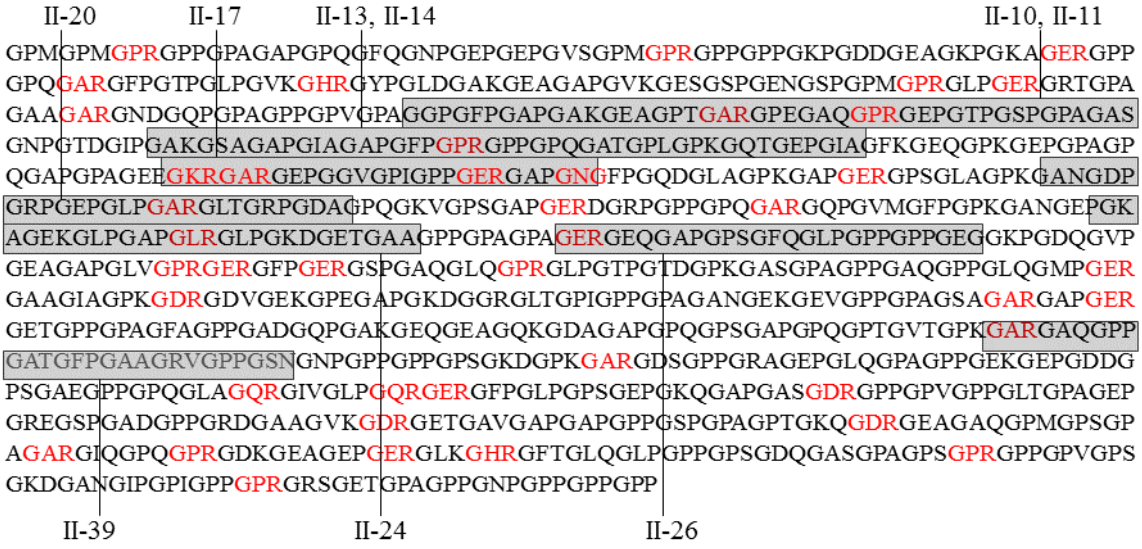
References

1. Prockop, D.J,; Kivirikko, K.I. Collagens: Molecular Biology, Diseases, and Potentials for Therapy. *Annu. Rev. Biochem.* **1995**, 64, 403-34.
2. Lamande, S.R.; Bateman, J.F. Procollagen folding and assembly: The role of endoplasmic reticulum enzymes and molecular chaperones. *Semin. Cell Dev. Biol.* **1999**, 10, 455-464.
3. Hirayoshi, K.; Kudo, H.; Takechi, H.; Nakai, A.; Iwamatsu, A.; Yamada, K.M.; Nagata, K. HSP47: a tissue-specific, transformation-sensitive, collagen-binding heat shock protein of chicken embryo fibroblasts. *Mol. Cell. Biol*. **1991**, 11, 4036-4044.
4. Ito, S.; Nagata, K. Biology of Hsp47 (Serpin H1), a collagen-specific molecular chaperone. *Semin. Cell Dev. Biol*. **2017**, 62, 142-151.
5. Nagata, K.; Saga, S.; Yamada, K.M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J. Cell Biol*. **1986**, 103, 223-229.
6. Thomson, C. a; Ananthanarayanan, V.S. Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro. *Biochem. J*. **2000**, 349, 877-883.
7. Koide, T.; Takahara, Y.; Asada, S.; Nagata, K. Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J. Biol. Chem*. **2002**, 277, 6178-6182.
8. 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**, 150, 1499-1506.
9. Masuda, H.; Fukumoto, M.; Hirayoshi, K.; Nagata, K. Coexpression of the collagen-binding stress protein HSP47 gene and the α1 (I) and α1 (III) collagen genes in carbon tetrachloride-induced rat liver fibrosis. *J. Clin. Invest*. **1994**, 94, 2481-2488.
10. Honzawa, Y.; Nakase, H.; Shiokawa, M.; Yoshino, T.; Imaeda, H.; Matsuura, M.; Kodama, Y.; Ikeuchi, H.; Andoh, A.; Sakai, Y.; et al. Involvement of interleukin-17A-induced expression of heat shock protein 47 in intestinal fibrosis in Crohn's disease. *Gut*. **2014**, 63, 1902-1912.
11. Friedman, S.L.; Sheppard, D.; Duffield, J.S.; Violette, S. Therapy for fibrotic diseases: Nearing the starting line. *Sci. Transl. Med*. **2013**, 5, 167sr1.
12. Sato, Y.; Murase, K.; Kato, J.; Kobune, M.; Sato, T.; Kawano, Y.; Takimoto, R.; Takada, K.; Miyanishi, K.; Matsunaga, T.; et al. Resolution of liver cirrhosis using vitamin A-coupled liposomes to deliver siRNA against a collagen-specific chaperone. *Nat. Biotechnol*. **2008**, 26, 431-442.
13. Ishiwatari, H.; Sato, Y.; Murase, K.; Yoneda, A.; Fujita, R.; Nishita, H.; Birukawa, N.K.; Hayashi, T.; Sato, T.; Miyanishi, K.; et al. Treatment of pancreatic fibrosis with siRNA against a collagen-specific chaperone in vitamin A-coupled liposomes. *Gut*. **2013**, 62, 1328-1339.
14. Hirai, K.; Kikuchi, S.; Kurita, A.; Ohashi, S.; Adachi, E.; Matsuoka, Y.; Nagata, K.; Watanabe, M. Immunohistochemical distribution of heat shock protein 47 (HSP47) in scirrhous carcinoma of the stomach. *Anticancer. Res*. **2006**, 26, 71-78.
15. Zhu, J.; Xiong, G.; Fu, H.; Evers, B.M.; Zhou, B.P.; Xu, R. Chaperone Hsp47 drives malignant growth and invasion by modulating an ECM gene network. *Cancer Res*. **2015**, 75, 1580-1591.
16. Sasikumar P, AlOuda KS, Kaiser WJ, et al. The chaperone protein HSP47: a platelet collagen binding protein that contributes to thrombosis and hemostasis. *J. Thromb. Haemost*. **2018**, 16, 946-959.

17. Thomson, C.A.; Atkinson, H.M.; Ananthanarayanan, V.S. Identification of small molecule chemical inhibitors of the collagen-specific chaperone Hsp47. *J. Med. Chem*. **2005**, 48, 1680-1684.
18. Okano-Kosugi, H.; Matsushita, O.; Asada, S.; Herr, A.B.; Kitagawa, K.; Koide, T. Development of a high-throughput screening system for the compounds that inhibit collagen-protein interactions. *Anal. Biochem*. **2009**, 394, 125–131.
19. Katarkar, A.; Haldar, P.K.; Chaudhuri, K. De novo design based pharmacophore query generation and virtual screening for the discovery of Hsp-47 inhibitors. *Biochem. Biophys. Res. Commun*. **2015**, 456, 707-713.
20. Ito, S.; Ogawa, K.; Takeuchi, K.; Takagi, M.; Yoshida, M.; Hirokawa, T.; Hirayama, S.; Shin-ya, K.; Shimada, I.; Doi, T.; et al. A small-molecule compound inhibits a collagen-specific molecular chaperone and could represent a potential remedy for fibrosis. *J. Biol. Chem*. **2017**, 292, 20076–20085.
21. Miyamura, T.; Sakamoto, N.; Kakugawa, T.; Taniguchi, H.; Akiyama, Y.; Okuno, D.; Moriyama, S.; Hara, A.; Kido, T.; Ishimoto, H.; et al. Small molecule inhibitor of HSP47 prevents pro-fibrotic mechanisms of fibroblasts in vitro. *Biochem. Biophys. Res. Commun*.**2020**, 530, 561-565.
22. Koide, T.; Takahara, Y.; Asada, S.; Nagata, K. Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J. Biol. Chem*. **2002**, 277, 6178-6182.
23. Raynal, N.; Hamaia, S.W.; Siljander, P.R.M.; Maddox, B.; Peachey, A.R.; Fernandez, R.; Foley, L.J.; Slatter, D.A.; Jarvis, G.E.; Farndale, R.W. Use of synthetic peptides to locate novel integrin α2β1-binding motifs in human collagen III. *J. Biol. Chem*. **2006**,281,3821-3831.
24. Koide, T.; Nishikawa, Y.; Asada, S.; Yamazaki, C.M.; Takahara, Y.; Homma, D.L.; Otaka, A.; Ohtani, K.; Wakamiya, N.; Nagata, K.; et al. Specific recognition of the collagen triple helix by chaperone HSP47: II. The HSP47-binding structural motif in collagens and related proteins. *J. Biol. Chem*. **2006**, 281, 11177-11185.
25. Widmer, C.; Gebauer, J.M.; Brunstein, E.; Rosenbaum, S.; Zaucke, F.; Drogemuller, C.; Leeb, T.; Baumann, U. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci*. **2012**, 109, 13243–13247.
26. 26. Farndale, R.W.; Lisman, T.; Bihan, D.; Hamaia, S.; Smerling, C.S.; Pugh, N.; Konitsiotis, A.; Leitinger. B.; de. Groot. P.G.; Jarvis, G.E.; Raynal, N. Cell-collagen interactions: The use of peptide toolkits to investigate collagen-receptor interactions. *Biochem. Soc. Trans*. **2008**, 36, 241-250.
27. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; De Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic. Acids. Res*. **2018**, 46(W1), W296-W303.
28. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: An online force field. *Nucleic. Acids. Res*. **2005**, 33(Web Server issue),W382-W388.
29. Morris, G.; Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem*. **2009**, 30, 2785-2791.
30. Hamaia, S.W.; Pugh, N.; Raynal, N.; Némoz, B.; Stone, R.; Gullberg, D.; Bihan, D.; Farndale, R.W. Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin 1$\alpha\beta$1 using collagen Toolkits II and III. *J. Biol. Chem*. **2012**, 287, 26019-26028.
31. Nishikawa, Y.; Takahara, Y.; Asada, S.; Shigenaga, A.; Otaka, A.; Kitagawa, K.; Koide, T. A structure-activity relationship study elucidating the mechanism of sequence-specific collagen recognition by the chaperone HSP47. *Bioorg. Med. Chem*. **2010**, 18, 3767-3775.
32. 32. Farndale, R.W. Collagen-binding proteins: Insights from the Collagen toolkits. Essays Biochem. **2019**.;63(3):337-348.
33. 33. Knight, C.G; Morton, L.F.; Peachey, A.R.; Tuckwell, D.S.; Farndale, R.W.; Barnes, M.J. The collagen-binding A-domains of integrin α1β1 and α2β1 recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J. Biol. Chem*. **2000**, 275, 35-40.
34. 34. Siljander, P.R.; Hamaia, S.; Peachey, A.R.; Slatter, D.A.; Smethurst, P.A.; Ouwehand, W.H.; Knight, C.G.; Farndale, R.W. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J. Biol. Chem*. **2004**, 279, 47763-72.
35. 35. Raynal, N,; Hamaia, S.W.; Siljander, P.R.; Maddox, B.; Peachey, A.R.; Fernandez, R.; Foley, L.J.; Slatter, D.A.; Jarvis, G.E.; Farndale, R.W. Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J. Biol. Chem*. **2006** 281, 3821-31.
36. Williams, B.R.; Gelman, R.A.; Poppke, D.C.; Piez, K. Collagen fibril formation. Optimal in vitro conditions and preliminary kinetic results. *J. Biol. Chem*. **1978**, 253, 6578-6585.
37. Yoshida M, Saito M, Ito S, et al. Structure-Activity Relationship Study on Col-003, a Protein-Protein Interaction Inhibitor between Collagen and Hsp47. *Chem. Pharm .Bull (Tokyo*). **2020**, 68, 220-226.

Supplementary Materials

(a)



(b)

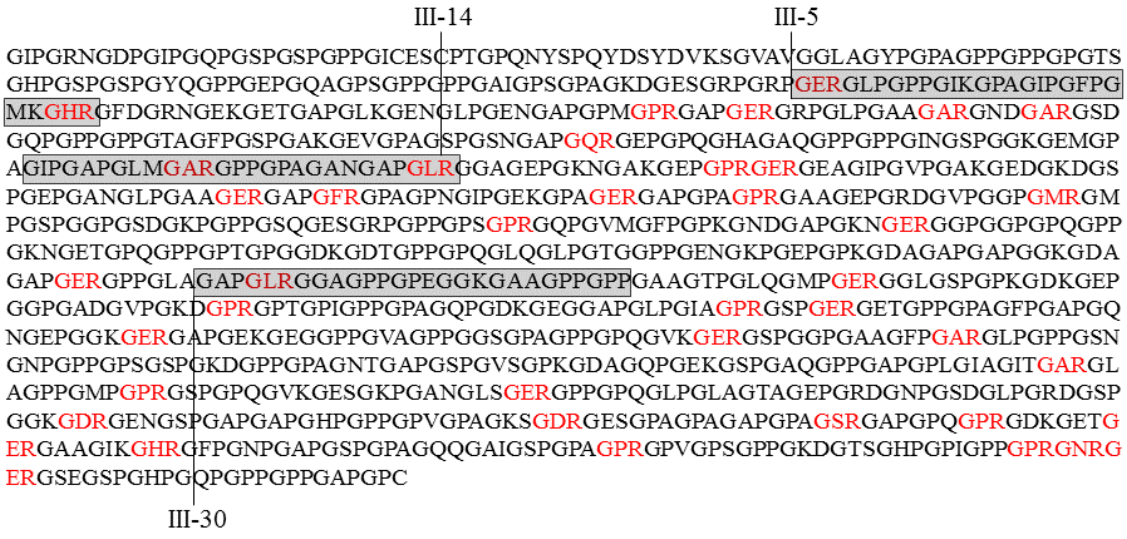
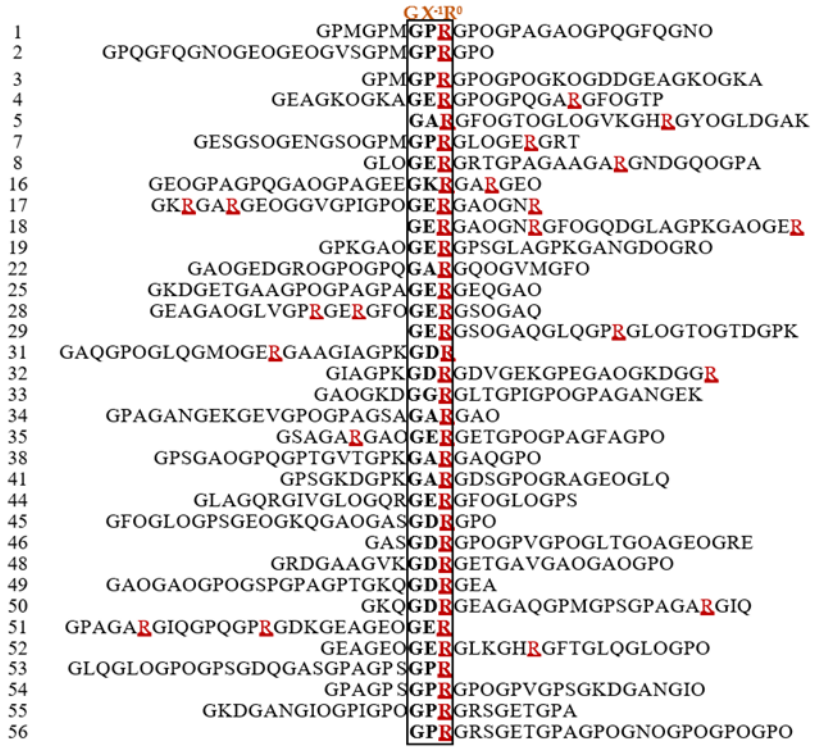


Figure S1. Binding sites of HSP47 on collagen II (**a**) and collagen III (**b**) with the corresponding collagen sequences highlighted. All the GXR sequences are colored red. The sequences do not include post-translational modification of proline (P).

(a)


G X-1R0
1 GPMGPMGPRGPOGPAGAOGPQGFQGNO
2 GPQGFQGNOGEOGEOGVSGPMGPRGPO
3 GPMGPRGPOGPOGKOGDDGEAGKOGKA
4 GEAGKOGKAGERGPOGPQGARGFOGTP
5 GARGFOGTOGLOGVKGHRGYOGLDGAK
7 GESGSOGENGSOGPMGPRGLOGERGRT
8 GLOGERGRTGPAGAAGARGNDGQOGPA
16 GEOGPAGPQGAOGPAGEEGKRGARGEO
17 GKRGARGEOGGVGPIGPOGERGAOGNR
18 GERGAOGNRGFOGQDGLAGPKGAOGER
19 GPKGAOGERGPSGLAGPKGANGDOGRO
22 GAOGEDGROGPOGPQGARGQOGVMGFO
25 GKDGETGAAGPOGPAGPAGERGEQGAO
28 GEAGAOGLVGPRGERGFOGERGSOGAQ
29 GERGSOGAQGLQGPRGLOGTOGTDGPK
31 GAQGPOGLQGMOGERGAAGIAGPKGDR
32 GIAGPKGDRGDVGEKGPEGAOGKDGGR
33 GAOGKDGGRGLTGPIGPOGPAGANGEK
34 GPAGANGEKGEVGPOGPAGSAGARGAO
35 GSAGARGAOGERGETGPOGPAGFAGPO
38 GPSGAOGPQGPTGVTGPKGARGAQGPO
41 GPSGKDGPKGARGDSGPOGRAGEOGLQ
44 GLAGQRGIVGLOGQRGERGFOGLOGPS
45 GFOGLOGPSGEOGKQGAOGASGDRGPO
46 GASGDRGPOGPVGPOGLTGOAGEOGRE
48 GRDGAAGVKGDRGETGAVGAOGAOGPO
49 GAOGAOGPOGSPGPAGPTGKQGDRGEA
50 GKQGDRGEAGAQGPMGPSGPAGARGIQ
51 GPAGARGIQGPQGPRGDKGEAGEOGER
52 GEAGEOGERGLKGHRGFTGLQGLOGPO
53 GLQGLOGPOGPSGDQGASGPAGPSGPR
54 GPAGPSGPRGPOGPVGPSGKDGANGIO
55 GKDGANGIOGPIGPOGPRGRSGETGPA
56 GPRGRSGETGPAGPOGNOGPOGPOGPO

(b)

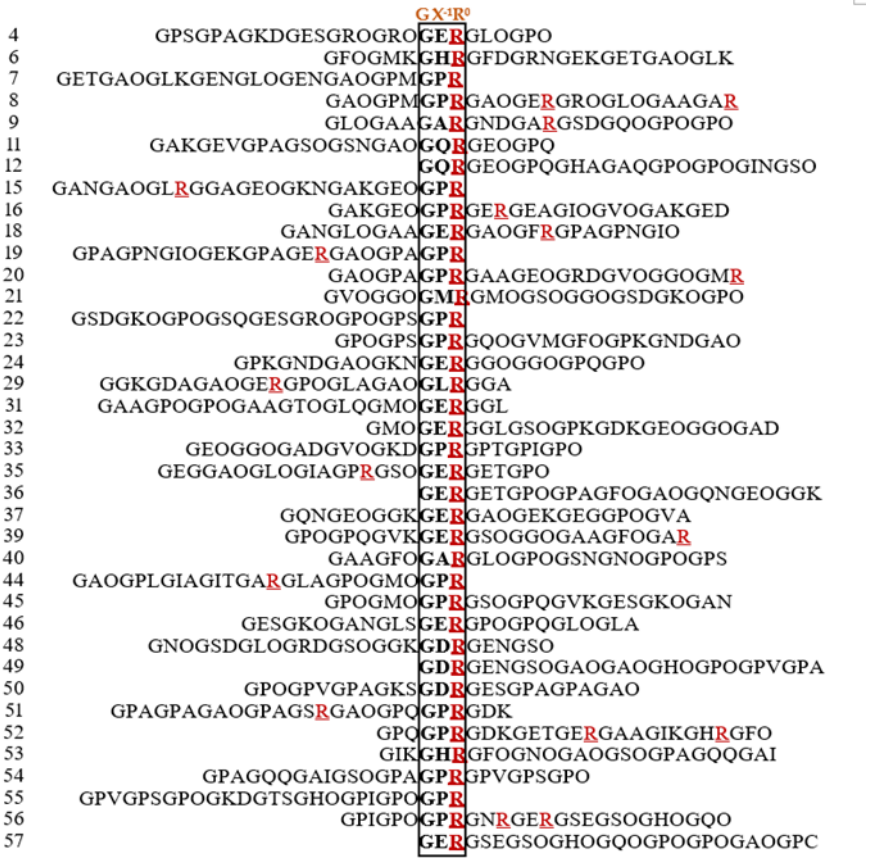
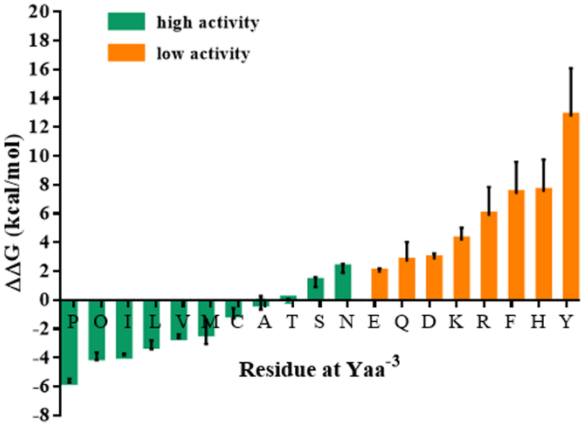

G X-1R0
4 GPSGPAGKDGESGROGROGERGLOGPO
6 GFOGMKGHRGFDGRNGEKGETGAOGLK
7 GETGAOGLKGENGLOGENGAOGPMGPR
8 GAOGPMGPRGAOGERGROGLOGAAGAR
9 GLOGAAGARGNDGARGSDGQOGPOGPO
11 GAKGEVGPAGSOGSNGAOGQRGEOGPQ
12 GQRGEOGPQGHAGAQGPOGPOGINGSO
15 GANGAOGLRGGAGEOGKNGAKGEOGPR
16 GAKGEOGPRGERGEAGIOGVOGAKGED
18 GANGLOGAAGERGAOGFRGPAGPNGIO
19 GPAGPNGIOGEKGPAGERGAOGPAGPR
20 GAOGPAGPRGAAGEOGRDGVOGGOGMR
21 GVOGGOGMRGMOGSOGGOGSDGKOGPO
22 GSDGKOGPOGSQGESGROGPOGPSGPR
23 GPOGPSGPRGQOGVMGFOGPKGNDGAO
24 GPKGNDGAOGKNGERGGOGGOGPQGPO
29 GGKGDAGAOGERGPOGLAGAOGLRGGA
31 GAAGPOGPOGAAGTOGLQGMOGERGGL
32 GMOGERGGLGSOGPKGDKGEOGGOGAD
33 GEOGGOGADGVOGKDGPRGPTGPIGPO
35 GEGGAOGLOGIAGPRGSOGERGETGPO
36 GERGETGPOGPAGFOGAOGQNGEOGGK
37 GQNGEOGGKGERGAOGEKGEGGPOGVA
39 GPOGPQGVKGERGSOGGOGAAGFOGAR
40 GAAGFOGARGLOGPOGSNGNOGPOGPS
44 GAOGPLGIAGITGARGLAGPOGMOGPR
45 GPOGMOGPRGSOGPQGVKGESGKOGAN
46 GESGKOGANGLSGERGPOGPQGLOGLA
48 GNOGSDGLOGRDGSOGGKGDRGENGSO
49 GDRGENGSOGAOGAOGHOGPOGPVGPA
50 GPOGPVGPAGKSGDRGESGPAGPAGAO
51 GPAGPAGAOGPAGSRGAOGPQGPRGDK
52 GPQGPRGDKGETGERGAAGIKGHRGFO
53 GIKGHRGFOGNOGAOGSOGPAGQQGAI
54 GPAGQQGAIGSOGPAGPRGPVGPSGPO
55 GPVGPSGPOGKDGTSGHOGPIGPOGPR
56 GPIGPOGPRGNRGERGSEGSOGHOGQO
57 GERGSEGSOGHOGQOGPOGPOGAOGPC

Figure S2. Sequences of 'GXR' containing collagen peptides from Toolkits II (a**) and III (**b**). O represents hydroxyproline.** For the sequences that contain more than one 'GXR' motif, we highlighted the most plausible motif according to the calculated results from of substitution at position -1 (Figure S4c).



Residue at position -3 (Yaa^{-3})	Predicted folding free energy compared to WT (ΔΔG (kcal/mol))	Reported IC_{50} value (μM) [a]
P	-5.60±0.13	0.94
O	-3.90±0.26	4.60
I	-3.78±0.06	19.00
L	-3.10±0.31	37.00
V	-2.52±0.12	8.40
M	-2.23±0.82	43.00
C	-0.93±0.36	4.20
A	-0.18±0.46	9.30
T	0.03±0.04	0.58
S	1.24±0.34	4.00
N	2.20±0.31	41.00
E	1.89±0.31	>100
Q	2.66±1.36	>100
D	2.82±0.40	>100
K	4.13±0.88	>100
R	5.85±2.00	>100
F	7.36±2.23	79.00
H	7.50±2.24	64.00
Y	12.73±3.35	95.00

Figure S3. The predicted folding free energy of a serial of Yaa^{-3}-substituted synthetic collagen-model peptides compared to peptide 8 (WT: $POGP^{-7}O^{-6}GP^{-4}T^{-3}GP^{-1}R^{0}GPOGPO$) binding of HSP47. Peptides with reported IC_{50} values lower than 50 μM are defined as high activity peptides (green), those higher than 50 μM are defined as low activity peptides (orange).

[a] Reference. 31

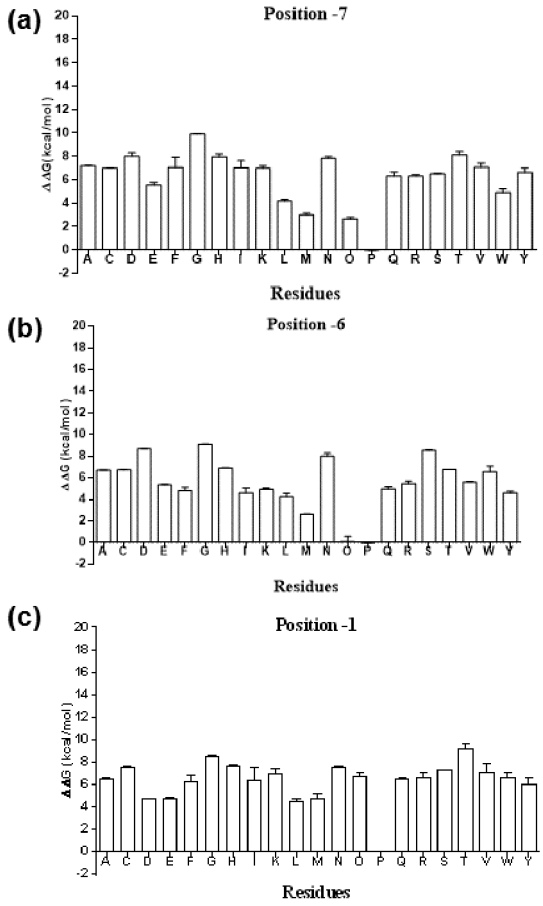


Figure S4. Predicted folding free energy changes for the substitutions at positions -7, -6 and -1 in CMP (PPGP^{-7}P^{-6}GP^{-4}T^{-3}GP^{-1}R^{0}GPPGPPG) on binding of HSP47. Data were shown as ΔΔG ± SD (kcal/mol) (n=5).

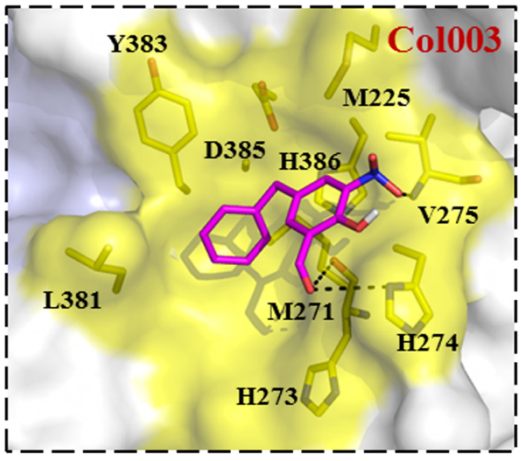


Figure S5. Predicted binding mode of Col003 to HSP47. HSP47 is shown as a colored surface. Residues of HSP47 that involved in interacting with the compounds are shown in sticks and labeled. Compound Col003 is shown in magenta sticks. The black dashed lines present hydrogen bonds.