

## Supplementary Materials

(a)

II-20 II-17 II-13, II-14 II-10, II-11

GPMGPMGPRGPPGPAGAPGPQCFQGNPGEPEGVSGPMGPRGPPGPPGKPGDDGEAGKPGKA~~GER~~GPP  
 GPQ~~GARG~~FPGTPTGLPGVK~~GHR~~YPLDGAKEAGAPGVKGESGSPGENSGPMP~~GPR~~GLP~~GER~~GRTGPA  
 GAA~~GARG~~GNDGQPGPAGPPGPVGPAGGPGFPGAPGAKGEAGPT~~GARG~~PEGAQ~~GPR~~GEPGTPGSPGPAGAS  
 GNP~~GTDGIP~~GA~~KGS~~AGAPGIAGAPGFP~~GPR~~GPPGPQGATGPLGPKGQTGEPIAGFKGEQGPKEGPPAGP  
 QGAPGPAGE~~GKRGARG~~EPGGVGP~~IGPP~~~~GER~~GAP~~GNG~~FPGQDGLAGPKGAP~~GER~~GPSGLAGPK~~GANGDP~~  
 GRPGEPLP~~GARG~~GLTGRPGDAGPQGVGPPGAP~~GER~~DGRPGPPGPQ~~GARG~~GQPGVMGFPGPKGANGEP~~GK~~  
 AGEKGLPGAP~~GLR~~GLPGKDGETGAA~~GPP~~GPAGPA~~GER~~GEQGA~~PPSGF~~QGLPGPPGPPGEG~~GK~~PGDQGV  
 GEAGAPGLV~~GPRGER~~GFP~~GER~~SGPAQGLQ~~GPR~~GLPGTPTDGPKGASGPAGPPGAQGPGLQGMP~~GER~~  
 GAAGIAGPK~~GDR~~GDVGEKGPEGAPGKDGGRLTGP~~IGPP~~GPAGANGEKGEVGP~~PP~~GPAGSAG~~GARG~~AP~~GER~~  
 GETGPPGPAGFAGPPGADGQPGAKGEQGEAGQKGDAGAPGPQGPSGAPGPQGTGVTGPK~~GARGA~~QGP  
 GATGFPGAAGRVGPPGS~~GN~~PGPPGPPSGKDGP~~KARG~~DSGPPGRAGEPLQGPAGPPGEKGEPPGDDG  
 PSGAEGPPGPQGLA~~GQR~~GIVGLP~~QQR~~GERGFPGLGPSGEPGKQAGPAS~~GDR~~GPPGPVGPGLTGPAGEP  
 GREGSPGADGPPGRDGAAGVK~~GDR~~GETGAVGAPGAPPPGSPGPAGPTGKQ~~GDR~~GEAGAQQPMGPSGP  
 A~~GARG~~IQGPQ~~GPR~~GDKGEAGEP~~GER~~GLK~~GHR~~GFTGLQGLPGPPGSGDQGASGPAGPS~~GPR~~GPPGPVGPS  
 GKDGANGIPGP~~IGPP~~~~GPR~~GRSGETGPAGPPGNPGPPGPPGP

II-39 II-24 II-26

(b)

III-14 III-5

GIPGRNGDPGIPGQPGSPGSPGPPGICESCTGPNQYSPQYDSYDVKSGVAVGGLAGYGPAGPPGPPGPGTS  
 GHPGSPGSPGYQGGPEPGQAGPSGPPGPPGAIGPSGPAKGDGESGRPGRP~~GER~~GLPGPPGKGPAGIPGFP  
 MK~~GHR~~GFDGRNGEKGETGAPGLKGENGLPGENGAPGPM~~GPR~~GAP~~GER~~GRPLPGAA~~GARG~~GND~~GARG~~SD  
 GQPGPPGPPGTAGFPSPGAKGEVGPAGSPGSNGAP~~GQR~~GEPGPQGHAGAQQPPGPPGINGSPPGKGEMGP  
 AGIPGAPGLM~~GARG~~GPPGPAGANGAP~~GLR~~GGAGEPGKNGAKGEP~~GPRGER~~GEAGIPGVPAKGEDGKDGS  
 PGEPGANGLPGAAG~~GER~~GAP~~GFR~~GAPGPNIPGEKGPA~~GER~~GAPGPA~~GPR~~GAAGEPGRDGVPPGGP~~GMR~~GM  
 PGSPGGPGSDGKPGPPGSQGESGRPGPPGPS~~GPR~~GQPGVMGFPGPKGNDGAPGKN~~GER~~GGPGGPGPQGP  
 GKNGETGPQGPPTGPGGDKGDTGPPGPQGLQGLPGTGPPGENGKPEPGPKGDAGAPGAPGKGDA  
 GAP~~GER~~GPPGLAGAP~~GLR~~GGAGPPGPEGGKGAAGPPGPPGAAGTPGLQGMPP~~GER~~GGLGSPGPKGDKGEP  
 GGPGADGVPGK~~DGPR~~GPTGPIGPPGPAGQPGDKGEGGAPGLPGIA~~GPR~~GSP~~GER~~GETGPPGPAGFPAGPGQ  
 NGEPGGK~~GER~~GAPGEKGEGGPPGVAGPPGSGPAGPPGPQGVK~~GER~~GSPGGPGAAGFP~~GARG~~LPGPPGSN  
 GNP~~GPP~~GPSGSPGKDGPAGNTGAPGSPGVSGPKGDAGQPGKEGSPGAQQPPGAPGLGIAGIT~~GARG~~L  
 AGPPGMP~~GPR~~GSPGPQGVKGESGKPGANGLS~~GER~~GPPGPQGLPGLAGTAGEPGRDGNP~~SD~~DGLPGRDGS  
 GGK~~GDR~~GENGSPGAPGAPGHPGPPGPVGPAGKS~~GDR~~GESGPAGAPGPA~~GSR~~GAPGPQ~~GPR~~GDKGET~~G~~  
 ERGAAGIK~~GHR~~GFPGNPGAPGSPGPAQQAIGSPGPA~~GPR~~GPVGPSGPPGKDGTSGHPGPIGPP~~GPRGNRG~~  
 ERGSESGPHPGQPGPPGPPGAPGPC

III-30

**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)

$G \times R^0$

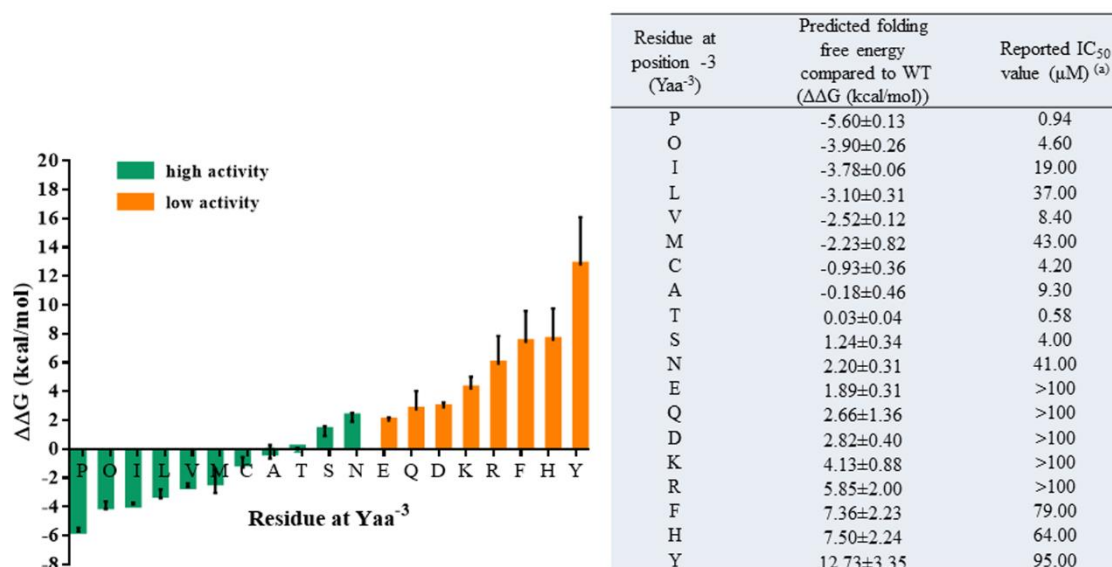
GPMGPMGPRGPOGPAGAOGPQGFQGN  
 GPQGFQGNOGEOGEOGVSGPMGPRGPO  
 GPMGPRGPOGPOGKOGDDGEAGKOGKA  
 GEAGKOGKAGERGPOGPQGARGFOTGP  
 GARGFOTGTOGLOGVKGHRGYOGLDGAK  
 GESGSOGENSGOGPMGPRGLOGERGRT  
 GLOGERGRTGPAGAAGARGNDGQOGPA  
 GEOGPAGPQGAOGPAGEEGKRGARGEO  
 GKRGARGEOGGVGPPIGPOGERGAOGR  
 GERGAOGRGFOGQDGLAGPKGAOGER  
 GPKGAOGERGPSGLAGPKGANGDOGRO  
 GAOGEDGROGPOGPQARGQOQVMGFO  
 GKDGETGAAGPOGPAGPAERGEQGAO  
 GEAGAOLVGPREGRGFOERGSOGAQ  
 ERGSOGAQGLQGPRLGLOGTOGTDGPK  
 GAQGPQGLQGMOGERGEAAGIAGPKGDR  
 GIAGPKGDRGDVGEKGPEGAOGKDGR  
 GAOGKDGRLTGPIGPOGPAGANGEK  
 GPAGANGEKGEVGPPOGPAGSARGAO  
 GSAGARGAOGERGGETGPAGFAGPO  
 GPSGAOGPQPTGVTGPKARGAQGP  
 GPSGKDGPKARGDSGPOGRAGEOGLQ  
 GLAGQRGIVGLOGQRERGFLOGPS  
 GFOGLOGPSGEOGKQGAOGASGDRGPO  
 GASGDRGPOGPVGPGLTGOAGEGRE  
 GRDGAAGVKGDRGETGAVGAOGAOGPO  
 GAOGAOGPOGSPGAGPTGKQGDREGEA  
 GKQGDREGEAAGQPMGPSGPAGARGIQ  
 GPAGARGIQGPQPRGDKGEAGEOGER  
 GEAGEOERGLKGHRGFTGLQLOGPO  
 GLQGLOGPOGPSGDQASGPAGPSGPR  
 GPAGPSGPRGPOGPVGPSKDGANGIO  
 GKDGANGIOGPPIGPOGPRGRSGETGPA  
 GPRGRSGETGPAGPOGNOGPOGPOGPO

(b)

$G \times R^0$

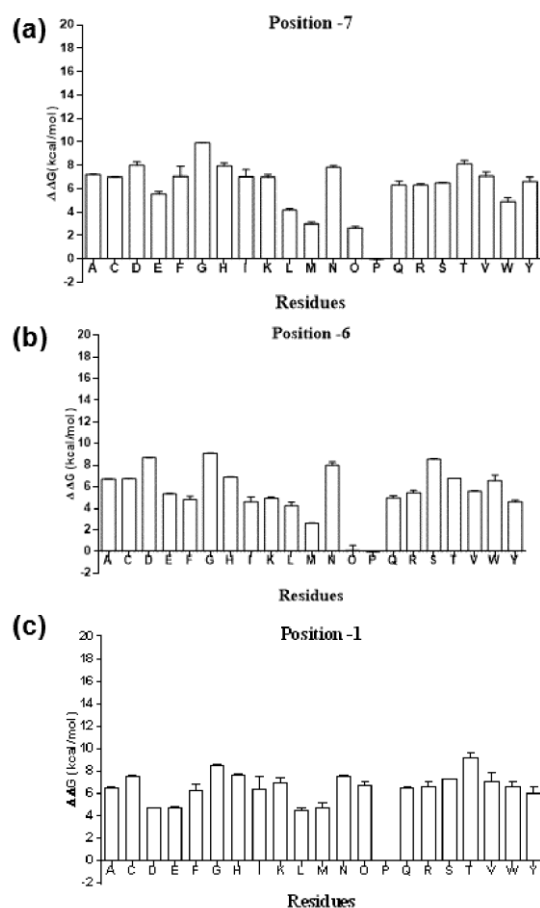
GPSGPAGKDGESGROGROERGLOGPO  
 GFOGMKGHRGFDGRNGEKGETGAOGLK  
 GETGAOGLKGENGLOGENGAOGPMGPR  
 GAOGPMGPRGAOGERGROGLOGAAGAR  
 GLOGAARGARGNDGARGSDGQOGPOGPO  
 GAKGEVGPAGSOGSNGAOGORGEOGPO  
 GORGEOGPQGHAGAOGPOGPOGINGSO  
 GANGAOGLRGGAGEOGKNGAKGEOGPR  
 GAKGEOGPRGERGEAGIOGVOGAKGED  
 GANGLOGAARGERGGAOGFRGPAGPNGIO  
 GPAGPNGIOGEKGPAGERGAOGPARGPR  
 GAOGPARGPRGAAGEOGRDGVGGOGMR  
 GVGGOGMRGMOGSOGGOGSDGKOGPO  
 GSDGKOGPOGSQGESGROGPOGPSGPR  
 GPOGPSGPRGQOQVMGFOGPKGNDGAO  
 GPKGNDGAOGKNERGGOGGOGPOGPO  
 GGKGDAGAOGERGPOGLAGAOGLRGGA  
 GAAGPOGPOGAAGTOGLQGMORGGL  
 GMOERGLGSGOPKGDKEOGGOGAD  
 GEOGGOGADGVGKDGPRGPTGPIGPO  
 GEGGAOGLGAGPRGSGOERGETGPO  
 ERGETGPAGFOGAOGQNGEOGK  
 QONGEOGKGERGGAOGEKGEGPOGVA  
 GPOGPQGVKERGSGOGGAAGFOGAR  
 GAAGFOGARGLGPOGSNGNOGPOGPS  
 GAOGPLGIAGITGARGLAGPOGMOGPR  
 GPOGMOGPRGSOGPQGVKGESGKOGAN  
 GESGKOGANGLSERGPOGPQGLOLA  
 GNOGSDGLOGRDGSOGGKGDRGENGSO  
 GDRGENGSOGAOGAOGHOGPOGPVGP  
 GPOGPVGPAGKSGDRGESGPAGPAO  
 GPAGPAGAOGPAGSRGAOGPARGDK  
 GPARGDKGETGERGAAGIKGHRGFO  
 GIKGHRGFOGNOGAOGSOGPAGQQGAI  
 GPAGQQGAIGSOGPARGPVGPSGPO  
 GPVGPSGPOGKDGTSGHOGPIGPOGPR  
 GPIGPOGPRGNRERGSEGSOGHOGQO  
 GERSEGSOGHOGQOGPOGPOGAOGPC

**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).

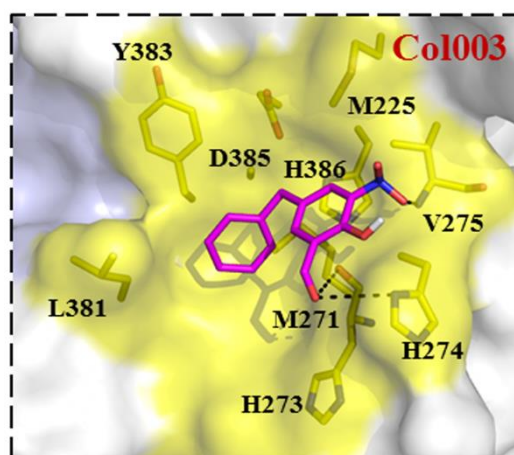


**Figure S3.** The predicted folding free energy of a serial of Yaa<sup>-3</sup>-substituted synthetic collagen-model peptides compared to peptide 8 (WT: POGP<sup>-7</sup>O<sup>-6</sup>GP<sup>-4</sup>T<sup>-3</sup>GP<sup>-1</sup>R<sup>0</sup>GPOGPO) binding of HSP47. Peptides with reported IC<sub>50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

<sup>a</sup> Reference. 31



**Figure S4.** Predicted folding free energy changes for the substitutions at positions -7, -6 and -1 in CMP (PPGP-<sup>7</sup>P-<sup>6</sup>GP-<sup>4</sup>T-<sup>3</sup>GP-<sup>1</sup>R<sup>0</sup>GPPGPPG) on binding of HSP47. Data were shown as  $\Delta\Delta G \pm SD$  (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.