

Structural insights into collagen-binding by platelet receptor Glycoprotein VI

Louris J. Feitsma¹, T. Harma C. Brondijk¹, Gavin E. Jarvis², Dominique Hagemans¹, Dominique Bihan², Natasia Jerah², Marian Versteeg¹, Richard W. Farndale^{2§†}, and Eric G. Huizinga^{†1}

¹Department of Structural Biochemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands. ²Department of Biochemistry, University of Cambridge, United Kingdom. [§]Present address: CambCol Laboratories Ltd, Littleport, Ely, Cambridgeshire, CB6 1RS, UK [†]These authors contributed equally. [†]Corresponding author; correspondence should be addressed to Richard W. Farndale, Department of Biochemistry, University of Cambridge, CB2 1QW, United Kingdom; e-mail: rwf10@cam.ac.uk.

Methods

Processing of diffraction data and structure solution

Diffraction data were integrated with MOSFLM¹ and scaled using Aimless² in the CCP4 suite³. Phasing of the GPVI dataset by molecular replacement was performed with PHASER⁴ and 4 copies of GPVI obtained from the previously solved structure of GPVI (PDB-ID: 2GI7)⁵. Because of an arrangement in two equal pairs of dimers related by non-crystallographic symmetry, chains B and D were restrained to chain A and C during refinement of the structure, for which tight restraints were applied on main chain atoms and medium restraints on side chain atoms.

Molecular replacement to solve the GPVI-(GPO)₅ complex involved initial placement of two copies of the refined model of GPVI truncation mutant and rigid body refinement using PHASER. The resulting *Fo*-*Fc* difference density map (Figure S1a) allowed unambiguous placement of one (GPO)₅-trimer, generated from the crystal structure of (GPO)₁₀ (PDB-ID: 1V4F⁶; Figure S1b-d). The structure of the GPVI-(GPO)₃ complex in the same crystal form was obtained by rigid body refinement of the GPVI-(GPO)₅ complex (Figure S1e) and manual rebuilding of the collagen peptide.

All structures were refined with REFMAC⁷ alternated with model improvement using COOT⁸ (Figure S1f-i). Refinement included TLS-refinement with 1 TLS-domain per GPVI chain and per collagen-trimer. In all cases, sufficient density was available to model the truncated loops (Figure S3b-c). Statistics of data processing and refinement for all structures are listed in Table S2.

Thermostability assay

GPVI loop truncation mutants were diluted 3- to 10-fold in PBS to 0.25 mg/mL. 18.75 µL of diluted protein solution was mixed with 6.25 µL 40x Sypro Orange (Sigma-Aldrich) in a 96-well plate (Greiner 652201). Thermal-unfolding curves were measured in sealed plates, heated from 289 to 368K with increments of 0.5K in an iCycler iQ Real Time PCR Detection System (Bio-Rad). The change in fluorescence was measured as described⁹. Melting temperatures (*T_m*) were defined as the inflection point of the unfolding curves.

Collagen-binding assay and data processing

To obtain fibrillar collagen, human placenta collagen type I (Sigma-Aldrich cat. no. C7774) was dissolved in 50 mM acetic acid and dialyzed overnight against 10 mM sodium phosphate buffer pH 7.4¹⁰. Collagen peptides were dissolved in 10 mM acetic acid and coated directly from acidic solution. Coating of peptides (5 µg/mL) and collagen fibrils (50 µg/mL) on 96-well plates (Costar 2595; Corning)

was performed overnight at RT. Wells were blocked with 3% (w/v) BSA in PBS for 2 hours at RT and then washed 4x with PBST (PBS + 0.1% Tween-20).

Serial 3-fold dilutions with concentrations ranging from 2.0 to 0.002 µM GPVI-Fc in PBS + 0.1% (w/v) BSA were incubated for 2 hours at RT. After discarding excess GPVI-Fc and 4x washing, 100 µL of 2000-fold diluted polyclonal rabbit anti-human IgG-HRP (DAKO, P0214) in PBS + 0.1% BSA was added to each well and incubated for 1 hour at RT. After 4x washing, color reactions were started by addition of 100 µL HRP-substrates *ortho*-phenylenediamine (0.4 mg/mL) and H₂O₂ (0.08%, v/v) in phosphate-citrate buffer pH 5.0. Reactions were stopped by adding 50 µL 1 M H₂SO₄ to each well, and absorbance was measured at 450 nm on a Model 680 Microplate Reader (Bio-Rad). Binding of each GPVI-Fc mutant was measured at least nine times in three independent sets of three measurements.

The average absorbance of three series of blank measurements without GPVI was subtracted from each measurement. Binding curves are displayed relative to the maximum binding of GPVI-Fc WT. For each mutant, the apparent dissociation constant *K_{D,app}* and relative *B_{max}* (the latter only if evidently different from 1) were obtained by fitting the binding data to formula 1 using non-linear regression in SigmaPlot (Systat software).

$$Abs = \frac{B_{max} \times c}{K_{D,app} + c} \quad (1)$$

where *c* is the GPVI concentration (µM).

Binding of wild type GPVI-Fc and the E40A and R38A mutants to CRP, GPP, Toolkit III-30 and III-30 derivatives (Table 1) was measured as previously described¹¹. Binding to all peptides was measured at a fixed GPVI-Fc concentration of 20 µg/mL and colored using anti-human IgG-HRP and HRP-substrate KPL TMB Sure Blue Reserve (Insight Biotechnology). Absorbance at 450 nm was measured on a Fluostar plate reader. Binding to these peptides was measured in at least three different experiments, and scaled to binding of wild type GPVI-Fc to CRP.

Analysis of GPVI-collagen complexes using FoldX

Contributions of individual (GPO)₅- and (GPO)₃-residues to the total GPVI binding energy were calculated with the FoldX Suite¹² in two steps: energy optimization using the *RepairPDB* function and subsequent energy calculation using the *SequenceDetail* function. The FoldX Suite¹² was obtained from <http://foldxsuite.crg.eu/> (64-bit) under academic license and run on a desktop machine equipped with 48 processing cores and Debian 7 operating system.

Energy optimization using the *RepairPDB* function was applied on all residues of each complex. Default values were used for ionic strength, pH, and temperature and all calculations were performed using the 'Ignore waters' option. Changes in coordinates of residues in the GPVI-collagen interface appeared minimal, and calculations using other options for water modeling did not lead to substantially different results and conclusions.

Calculation of the contributions of individual collagen peptide residues to GPVI-binding was performed by running the *SequenceDetail* function twice per GPVI-collagen interaction. In the first run, the energy optimized complexes including the target GPVI-molecule were used, by manual deleting the other molecule. In the second run, the collagen peptide only was used, deleting also the target molecule. Subsequent comparison of the results with and without GPVI allowed calculation of the ΔE_i -values for the binding energy per (GPO)₅- and (GPO)₃-residue *i* using formula 2.

$$\Delta E_i = E_{\text{complex},i} - E_{\text{peptide},i} \quad (2)$$

To calculate residue contributions averaged over the four crystallographically resolved complexes, (GPO)₅-residues involved in formation of the complex with the first GPVI-molecule, i.e. M:Pro2-Hyp6 and T:Pro2-Hyp9, were used as a template for residue numbering and other residue numbers were transposed down. The two collagen chains involved in binding were assigned *chain 1* and *chain 2*, respectively, in order of the chain stagger. Differences in numbering between the staggered chains arising from the T+L combination in the GPVI-(GPO)₃ complex were also transposed down.

Analysis of the correlation between GPVI binding to Toolkit-III peptides and the presence of various GPO-containing motifs

Previous studies^{11,13} employing Toolkit-III peptides¹⁴ to scan the entire fibrillar region of homotrimeric collagen III identified 19 peptides that bind GPVI-Fc dimers better than GPP-peptide does¹³, whereas only one of these (III-01) comprises the full POGPOGPO-sequence. To determine the sequence motif in Toolkit-III peptides that best correlates with previously obtained binding data for GPVI-Fc dimers¹³ we calculated for all 57 Toolkit-III peptides the number of occurrences of various non-overlapping sequence motifs comprising either the POGPOGPO sequence that constitutes the GPVI-contact site on (GPO)₅ in our crystals, or a shorter sequence contained within it starting with either P or O (G exists by default). Pearson's correlation coefficients (*r*) between the number of occurrences of a particular sequence pattern in peptide *i* (*p_i*) and the normalized *A*₄₅₀-value for GPVI-binding to peptide *i* (*A*_{450,*i*}) (taken from Jung et al.¹³) were calculated using formula 3.

$$r = \frac{\sum (p_i - \bar{p})(A_{450,i} - \bar{A}_{450})}{\sqrt{\sum (p_i - \bar{p})^2 \sum (A_{450,i} - \bar{A}_{450})^2}} \quad (3)$$

where \bar{p} is the mean number of occurrences of the pattern and \bar{A}_{450} is the mean *A*₄₅₀-value for all 57 peptides.

In silico analysis of binding site distributions on triple helices and peptide model generation

The repetitive composition of the GPVI-binding pattern OGPOGP, i.e. (OGP)₂, results in a large repertoire of partially overlapping GPVI binding sites on (GPO)_{*n*}-peptides and -fragments. To determine structural constraints for (dual) GPVI-binding to crystallization peptides (GPO)₅ and (GPO)₃, and models of longer collagen fragments including Toolkit-peptide III-30, we performed an extensive *in silico* analysis assessing frequency, position, and overlap of binding sites as shown in Figure S9a. We also analyzed the feasibility of dual receptor binding without steric hindrance.

We performed a binding site analysis of (GPO)₅ and (GPO)₃ by assessing all possible combinations of staggered collagen chains and positions on the triple helix. As shown in Figure S9a, eight structurally identical, but overlapping binding sites are present on (GPO)₅: three sites formed by L+M, M+T, and T+L combinations of collagen chains, respectively, start from Pro2 of the first GPO-triplet, three from Pro5 of the second triplet, and, limited by the length of the GPO-stretch, only two (L+M & M+T) from Pro8 of the third triplet. The shorter (GPO)₃ yielded only two sites within the GPO-stretch itself, but, as observed in the crystal structure of the GPVI-(GPO)₃ complex, sites can also partially overlap the (GPP)₂-tail included for peptide stability at the C-terminus, which added another four sites in this particular situation.

To screen for all possibilities of (GPO)₅ to bind two GPVI-molecules simultaneously, *in silico* analysis was performed employing a template fragment that contained the first GPVI-molecule of the GPVI-(GPO)₅ complex and (GPO)₅-residues involved in its binding (M:Pro2-Pro6 & T:Pro2-Hyp9), which was subsequently superimposed onto all available sites. Output models were filtered for solutions evidently precluded by steric hindrance. This analysis exhibited two structurally distinct molecular configurations as shown in Figure S9b: GPVI-molecules binding to sites with a relative difference of 2 or 5 axial positions.

Extension of this analysis to include longer collagen fragments with different triple helical conformations required *in silico* generation of model peptides. A model of (GPO)₁₀ with a 7₂-helix conformation was adapted from the infinite (GPO)_{*n*}-structure (PDB-ID: 1V4F⁶); a model with 10₃-helix conformation was obtained from the structure of the G982-G1023-region of human type III collagen (PDB-ID: 3DMW¹⁵, residues 6-17) and manually modified to (GPO)₁₀. To determine possible sites on collagen that can be simultaneously occupied by two GPVI-molecules, we generated hypothetical models of dual binding by positioning our GPVI-(GPO)₅ structure onto the helix as described above.

Different models of Toolkit-peptide III-30 were used depending on the type of analysis. When analyzing interactions with GPVI as shown in Figures 4d-e, binding site models were generated by manual alteration of the (GPO)₅ amino acid sequence without modifying the collagen backbone structure. For *in silico* analysis of the full length peptide as shown in Figure S13c, the amino acid sequence of (GPO)₁₀-10₃-helix was manually modified. Positioning of GPVI-molecules onto the helix was performed as described above.

REFERENCES

1. Leslie AGW, Powell HR. *Evolving Methods for Macromolecular Crystallography*. Dordrecht: Springer Netherlands; 2007.
2. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* 2013;69(Pt 7):1204–14.
3. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 2011;67(Pt 4):235–42.
4. McCoy AJ, Grosse-Kunstleve RW, Adams PD, et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 2007;40(Pt 4):658–674.
5. Horii K, Kahn ML, Herr AB. Structural basis for platelet collagen responses by the immune-type receptor glycoprotein VI. *Blood.* 2006;108(3):936–42.
6. Okuyama K, Hongo C, Fukushima R, et al. Crystal Structures of Collagen Model Peptides with Pro-Hyp-Gly Repeating Sequence at 1.26 Å Resolution: Implications for Proline Ring Puckering. *Biopolymers.* 2004;76:367–377.
7. Vagin A a, Steiner R a, Lebedev A a, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D. Biol. Crystallogr.* 2004;60(Pt 12 Pt 1):2184–95.
8. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 2010;66(Pt 4):486–501.
9. Ericsson UB, Hallberg BM, Detitta GT. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* 2006;357(2):289–298.
10. van der Plas RM, Gomes L, Marquart J a, et al. Binding of von Willebrand factor to collagen type III: role of specific amino acids in the collagen binding domain of vWF and effects of neighboring domains. *Thromb. Haemost.* 2000;84(6):1005–11.
11. Jarvis GE, Raynal N, Langford JP, et al. Identification of a major GpVI-binding locus in human type III collagen. *Blood.* 2008;111(10):4986–4996.
12. Schymkowitz JWH, Rousseau F, Martins IC, et al. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102(29):10147–10152.
13. Jung SM, Moroi M, Soejima K, et al. Constitutive dimerization of glycoprotein VI (GPVI) in resting platelets is essential for binding to collagen and activation in flowing blood. *J. Biol. Chem.* 2012;287(35):30000–30013.
14. Raynal N, Hamaia SW, Siljander PR-M, et al. Use of synthetic peptides to locate novel integrin $\alpha 2\beta 1$ -binding motifs in human collagen III. *J. Biol. Chem.* 2006;281(7):3821–31.
15. Boudko SP, Engel J, Okuyama K, et al. Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J. Biol. Chem.* 2008;283(47):32580–9.
16. Smethurst PA, Joutsu-Korhonen L, O'Connor MN, et al. Identification of the primary collagen-binding surface on human glycoprotein VI by site-directed mutagenesis and by a blocking phage antibody. *Blood.* 2004;103(3):903–11.
17. Lecut C, Arocas V, Ulrichs H, et al. Identification of residues within human glycoprotein VI involved in the binding to collagen: evidence for the existence of distinct binding sites. *J. Biol. Chem.* 2004;279(50):52293–9.
18. O'Connor MN, Smethurst P a, Farndale RW, Ouwehand WH. Gain- and loss-of-function mutants confirm the importance of apical residues to the primary interaction of human glycoprotein VI with collagen. *J. Thromb. Haemost.* 2006;4(4):869–73.
19. Brondijk THC, de Ruiter T, Ballering J, et al. Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood.* 2010;115(7):1364–73.
20. Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A.* 2006;103(24):9001.
21. Slater A, Di Y, Clark JC, et al. Structural characterization of a novel GPVI-nanobody complex reveals a biologically active domain-swapped GPVI dimer. *Blood.* 2021;137(24):3443–3453.

Table S1. Primer sequences for construction of GPVI loop-truncation mutants

Mutant	+/-	Deleted residues	Primer sequence
ΔPAVS	Fw	Pro122-Ser125	5'-CTCTCAGCCCAGCCCGGCTCAGGAGGGGACGTAACCC-3'
ΔPAVS	Rv	Pro122-Ser125	5'-GGGTTACGTCCCCTCCTGAGCCGGGCTGGGCTGAGAG-3'
ΔPAVSG	Fw	Pro122-Ser125, Gly127	5'-CTCTCAGCCCAGCCCGGCTCAGGGGACGTAACCCTACAGTGTC-3'
ΔPAVSG	Rv	Pro122-Ser125, Gly127	5'-GACACTGTAGGGTTACGTCCCCTGAGCCGGGCTGGGCTGAGAG-3'
ΔPAPYK	Fw	Pro151-Lys155	5'-GCTCTGTACAAGGAAGGGGACAATCCCGAGAGATGGTACCGGG-3'
ΔPAPYK	Rv	Pro151-Lys155	5'-CCCGGTACCATCTCTCGGGATTGTCCCCTTCCTGTACAGAGC-3'
ΔPAPYKN	Fw	Pro151-Asn156	5'-GCTCTGTACAAGGAAGGGGACCCCGAGAGATGGTACCGGGCTAG-3'
ΔPAPYKN	Rv	Pro151-Asn156	5'-CTAGCCCGGTACCATCTCTCGGGGTCCCCTTCCTGTACAGAGC-3'

Table S2. Data collection and refinement statistics

Data Collection	GPVI ΔPAVS-PAPYKN	GPVI-(GPO) ₅	GPVI-(GPO) ₃
Synchrotron	Swiss Light Source (SLS)	European Synchrotron Radiation Facility (ESRF)	European Synchrotron Radiation Facility (ESRF)
Beamline	PX	ID23-1	ID23-1
Wavelength (Å)	0.9999	0.9724	0.9724
Space group	P 2 ₁	P 4 ₁ 2 ₁ 2	P 4 ₁ 2 ₁ 2
Cell dimensions a, b, c (Å) α, β, γ (°)	78.65, 44.05, 117.7 90, 104.7, 90	59.97, 59.97, 313.2 90, 90, 90	59.72, 59.72, 319.7 90, 90, 90
Resolution range (Å)*	52.4 - 1.90 (1.94 - 1.90)	60.0 2.50 (2.61 - 2.50)	58.7 - 2.50 (2.61 - 2.50)
No. unique reflections	58873 (3656)	20736 (2463)	21273 (2503)
Redundancy	2.1 (2.1)	7.9 (7.4)	6.4 (6.8)
R _{merge}	0.081 (1.08)	0.121 (1.03)	0.113 (0.971)
I/σI	5.0 (1.1)	9.7 (1.8)	8.5 (1.9)
Completeness (%)	95.1 (93.3)	99.5 (99.8)	99.9 (100.0)
CC(1/2)	0.990 (0.620)	0.995 (0.535)	0.984 (0.646)
Refinement R _{work} / R _{free}	0.221 / 0.259	0.208 / 0.257	0.227 / 0.260
Content of asymmetric unit	4× GPVI, 14× Cl ⁻ , 2× PO ₄ ³⁻ , 2× PEG ₆ , 344× H ₂ O	2× GPVI, 1× (GPO) ₅ , 2× Cl ⁻ , 2× PEG ₄ , 88× H ₂ O	2× GPVI, 1× (GPO) ₃ , 2× Cl ⁻ , 2× PEG ₄ , 41× H ₂ O
No. atoms Protein Water / other ligands	6101 5645 344 / 112	3114 2973 88 / 53	3210 3117 41 / 52
Average B / Wilson B (Å ²)	33.2 / 25.6	58.4 / 34.2	69.5 / 56.4
RMS deviations Bond lengths (Å) Bond angles (°)	0.011 1.48	0.009 1.63	0.008 1.56
Ramachandran Plot Favored (%) Allowed (%) Outliers (%)	98.5 1.37 0.14	97.4 2.63 0	96.0 3.71 0.25
MolProbity Clashscore (percentile)	1.52 (95 th)	1.29 (100 th)	1.64 (99 th)

* Numbers between brackets refer to the highest resolution shell.

Table S3. $K_{D,app}$ binding constants measured for fibrillar collagen I-binding in solid-phase plate assays by GPVI-Fc mutations outside the primary binding site

Mutant	$K_{D,app}$	95% confidence interval	Increase	Max. binding*
WT	30.1 nM	$26.8 \leq K_D \leq 33.5$ nM		
K59A	72.4 nM	$39.0 \leq K_D \leq 106$ nM	2-fold	85% (76 - 95%)
K41A	35.2 nM	$28.3 \leq K_D \leq 42.1$ nM		119% (113 - 124%)
Q48A	19.3 nM	$13.5 \leq K_D \leq 20.5$ nM		109% (102 - 115%)
L62E	29.3 nM	$22.4 \leq K_D \leq 36.2$ nM		108% (103 - 114%)
R46E	47.7 nM	$36.3 \leq K_D \leq 59.2$ nM		105% (99 - 110%)

* Values represent the max. binding (B_{max}) and 95% confidence intervals, if other than 100%

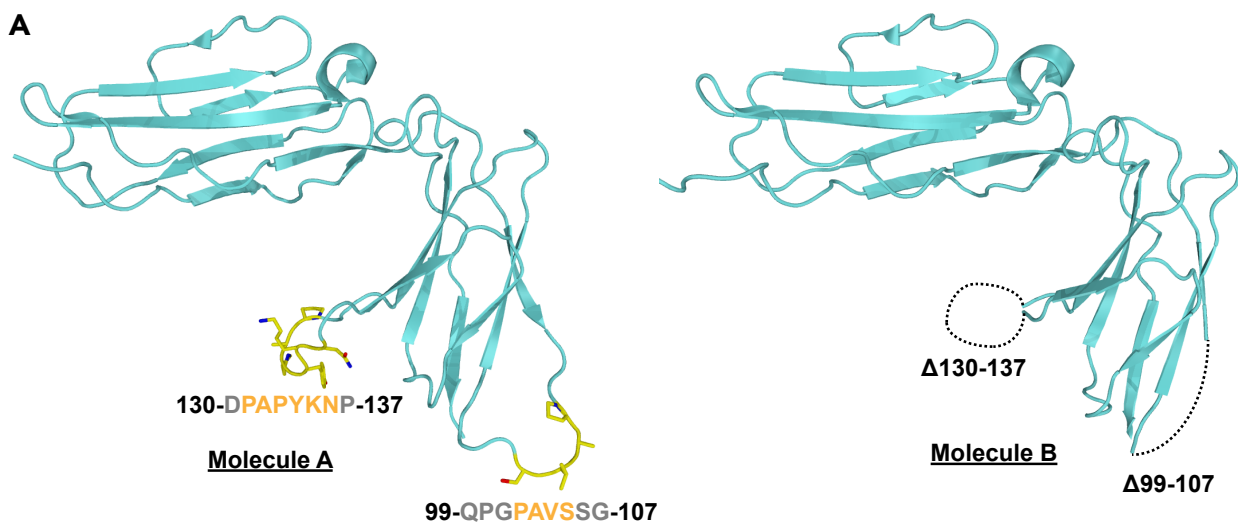
Video S1. Rotating video of the GPVI-(GPO)₅ binding interface.

See uploaded file Video1.mov. Sticks representation of collagen peptide (GPO)₅ colored by chain, i.e pink (L), magenta (M), and orange (T), and cartoon representation of the first GPVI-molecule of the GPVI-(GPO)₅ complex also showing key residues situated within 4.5 Å of (GPO)₅ as sticks. Hydrogen bonding interactions and salt bridges are displayed as black dashed lines.

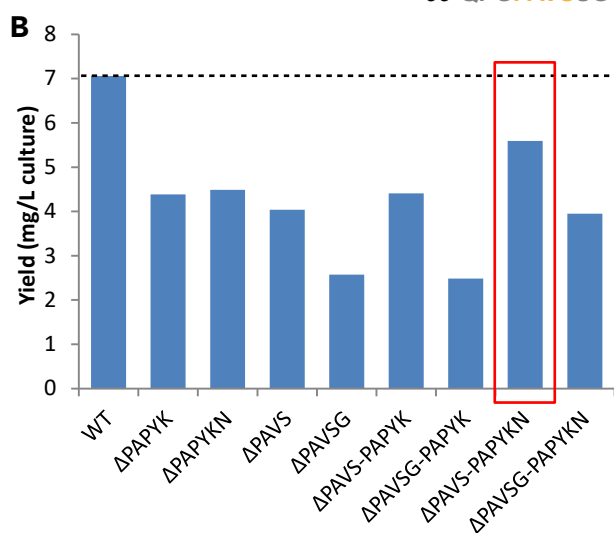
Supplemental Figures

Figure S1

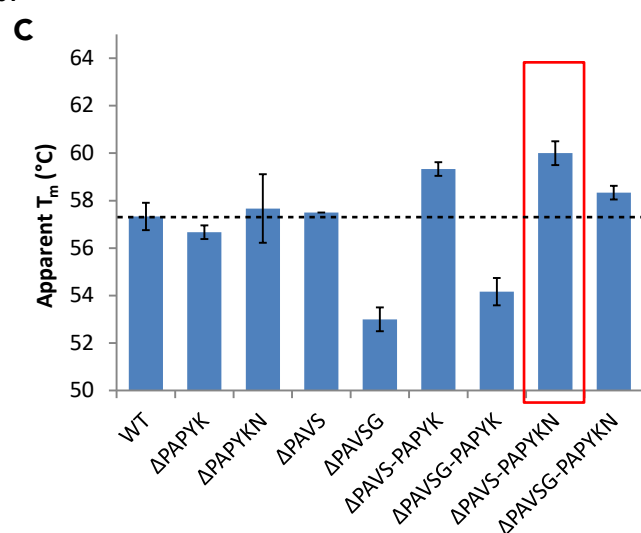
A



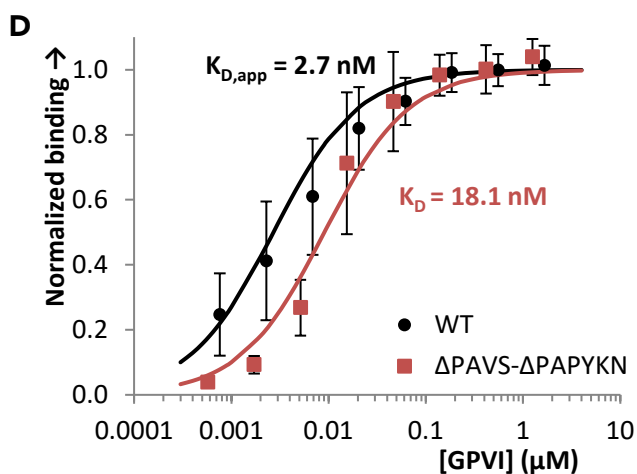
B



C



D



E

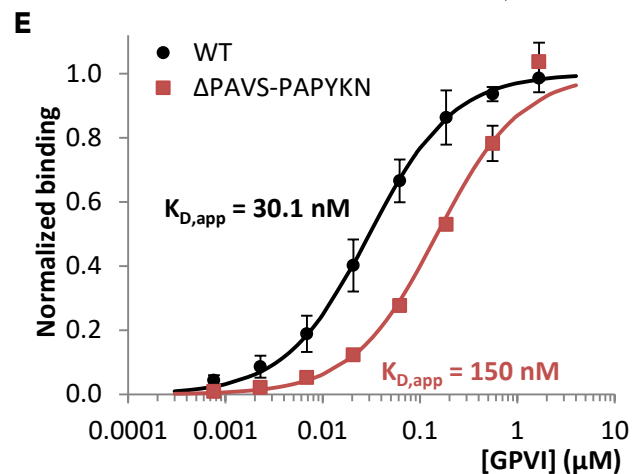


Figure S1. Construction and characterization of GPVI loop truncation mutants. (A) Previously solved GPVI-structure (PDB ID code: 2GI7⁵; cyan cartoon representation) highlighting two loops in the D2-domain (res. 99-107 and 130-137) that are visible in molecule A (left, yellow sticks), but invisible in molecule B due to high mobility (right, dashed lines). (B) Expression yields of wild type GPVI and eight loop truncation mutants showing moderate reductions for all mutants with respect to wild type protein (7.1 mg/L culture). The highest yield (5.6 mg/L culture) was obtained for the ΔPAVS-ΔPAPYKN mutant (red box). (C) Unfolding temperatures (T_m) of wild type GPVI and loop truncation mutants. The highest unfolding temperature was obtained for the ΔPAVS-ΔPAPYKN mutant (red box), which is about 2.5 °C more thermostable than wild type GPVI. (D-E) CRP- (D) and Collagen I-binding (E) of wild type GPVI (black) and the ΔPAVS-ΔPAPYKN mutant (red) as measured in a solid-state assay. Data measured as A450 is normalized to the binding of wild type protein. All data points represent the mean \pm SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{max} \times c) / (K_{D,app} + c)$ for 1 to 1 binding using non-linear regression in SigmaPlot, where c is the GPVI concentration (μM).

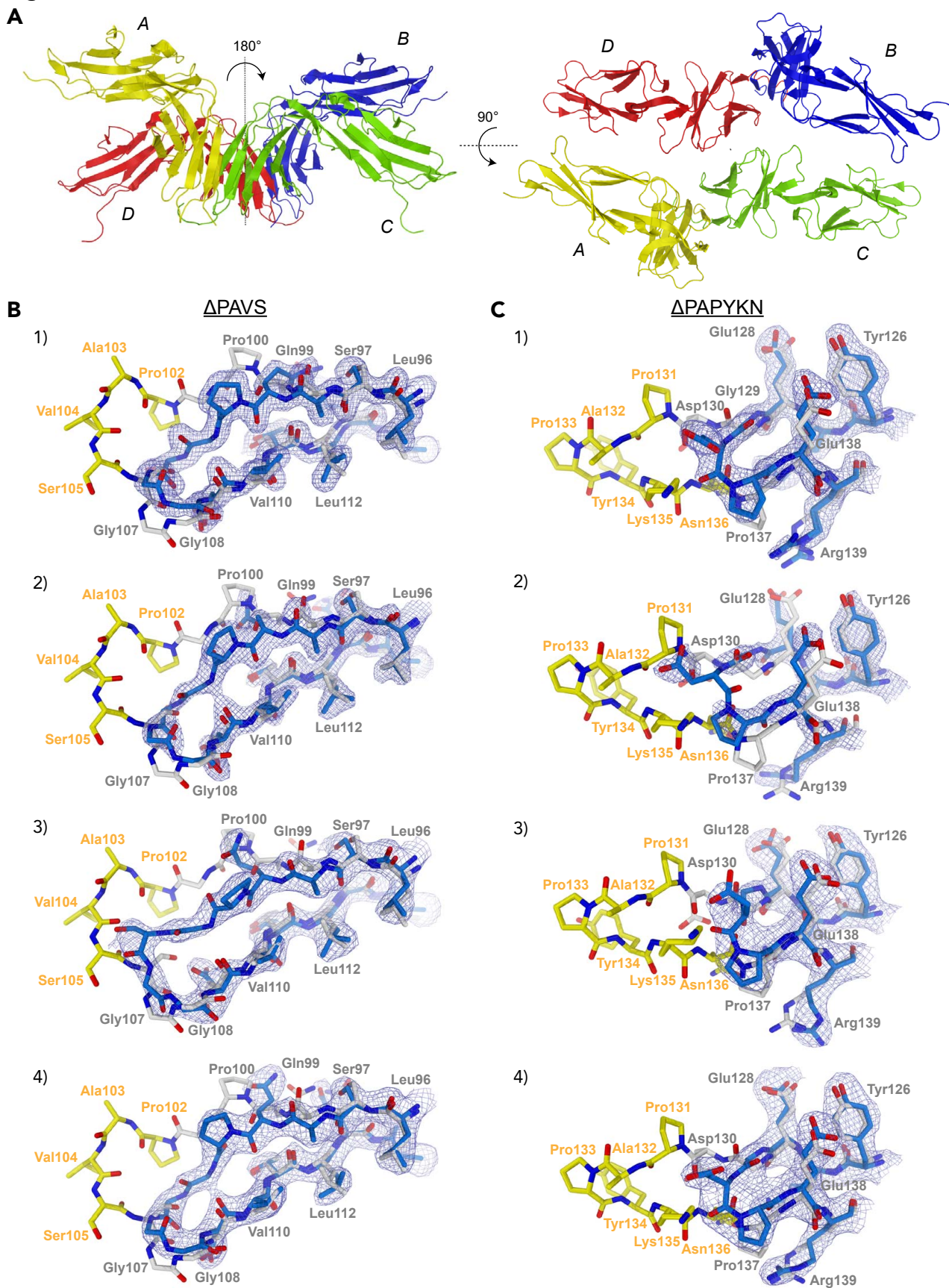
Figure S2

Figure S2. Crystal structure of GPVI Δ PAVS- Δ PAPYKN. (A) 1.9 Å-resolution structure of the GPVI Δ PAVS- Δ PAPYKN mutant showing a dimer-of-dimers arrangement with two GPVI-dimers (yellow-green and blue-red) that are related by a non-crystallographic 2-fold rotation. (B-C) Conformation of the truncated Δ PAVS- (B) and Δ PAPYKN-loops (C) as observed in GPVI-monomers A (panel 1) and C (panel 2) of the unbound structure, as well as in GPVI-monomers A (panel 3) and B (panel 4) of the GPVI-(GPO)₅ complex. Truncated loops are depicted as blue sticks together with the 2Fo-Fc electron density map contoured at 1.2 σ . Each panel shows for comparison the structure of the non-truncated loop of wild type GPVI (PDB-ID: 2GI7⁵) in yellow and gray sticks.

Figure S3

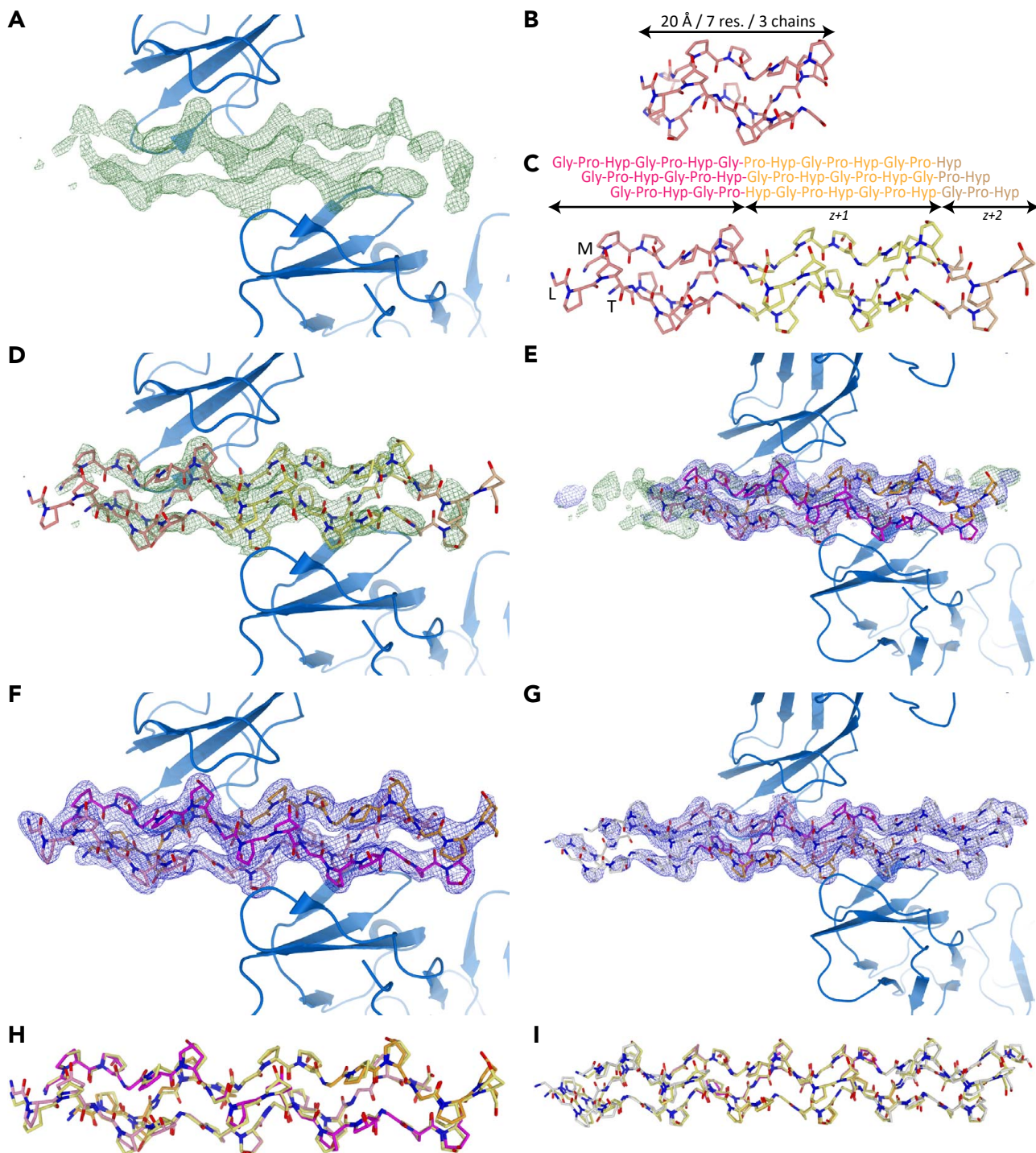


Figure S3. Model building of collagen peptides in the GPVI-(GPO)₅ and GPVI-(GPO)₃ complexes. (A) *Fo-Fc* difference electron density map contoured at 2σ (green) obtained after positioning of two GPVI molecules (cartoon representation; marine-blue) in the asymmetric unit of the GPVI-(GPO)₅ crystals by molecular replacement. (B) Structure of one 7₂-helix turn of the infinite structure of (GPO)_n (PDB- ID: 1V4F⁶). (C) Initial model of the one-residue staggered (GPO)₅ trimer constructed from the (GPO)_n-asymmetric unit in panel B. (D) (GPO)₅ triple helix (panel C) manually positioned in the difference electron density map (panel A). (E) 2*Fo-Fc* (blue, 1.2σ contour) and *Fo-Fc* (green, 2σ contour) electron density maps after rigid body refinement of the GPVI-(GPO)₅ complex against the diffraction data of the GPVI-(GPO)₃ complex. The positive (green) difference density at either end of the triple helix is consistent with the presence of a (GPP)₂ sequence at either end of the (GPO)₃ sequence, causing the length of what we refer to as the (GPO)₃ peptide to exceed the length of the (GPO)₅ peptide (also see Table 1). (F-G) Final model and 2*Fo-Fc* electron density map (1.2σ contour) obtained after crystallographic refinement of the GPVI-(GPO)₅ (F) and GPVI-(GPO)₃ complexes (G). The electron density map is only drawn within a distance of 1.6 Å from the collagen peptide. Each peptide chain is shown in a different color for the GPO containing regions; in panel G the (GPP)₂ sequences at either end of the (GPO)₃ peptides are shown in white. (H) Superposition of the initial 7₂-helix model (yellow) and refined model (tricolor) of (GPO)₅. (I) Superposition of a (GPO)₇ trimer with 7₂ helical symmetry (yellow) and refined model of (GPO)₃ (tricolor; white (GPP)₂-termini).

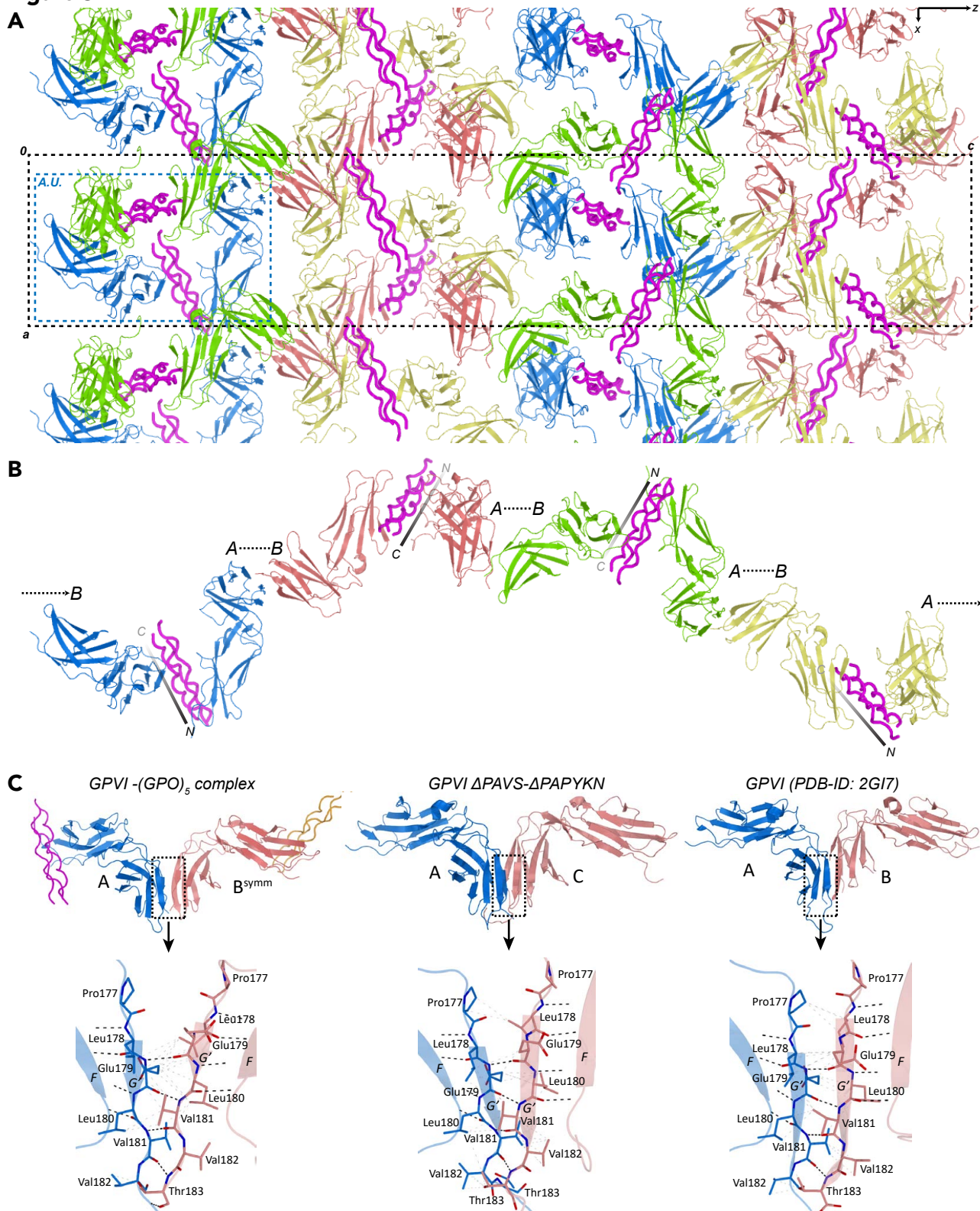
Figure S4

Figure S4. Crystallographic back-to-back interactions of GPVI D2-domains. (A) Crystal packing of the GPVI-(GPO)₅ crystal structure shown in the xz-plane of the crystal. The asymmetric unit (A.U., blue dashes) and unit cell (black dashes) are indicated. GPVI-molecules are shown in ribbon representation differently colored to facilitate discrimination between overlapping molecules; the collagen peptide is magenta. (B) GPVI- and (GPO)₅-trimers in the z-direction show an alternating arrangement of two GPVI-molecules bound to the same collagen trimer through their D1-domain and two GPVI-molecules contacting each other through their D2-domain that we refer to as back-to-back interaction. The amino to carboxy-terminal direction of the collagen peptides is indicated by black lines labeled the ends of which are labeled N and C. Coloring as in panel A. (C) Molecular details of the back-to-back D2-interactions as observed in the structures of the GPVI-(GPO)₅ complex (left panel), the GPVI Δ PAVS- Δ PAPYKN mutant (middle panel), and the previously solved structure of GPVI (PDB-ID: 2GI7⁵; right panel). Back-to-back interaction appears to be stabilized by the formation of an extended intermolecular β -sheet arising from the interaction between strand G' of each molecule. GPVI-molecules (blue and salmon) are shown as cartoon representations. Strand G' of each molecule is also shown in sticks representation in the bottom zoom-in panels. Inter- and intramolecular hydrogen bonds between backbone atoms are indicated by dashed lines.

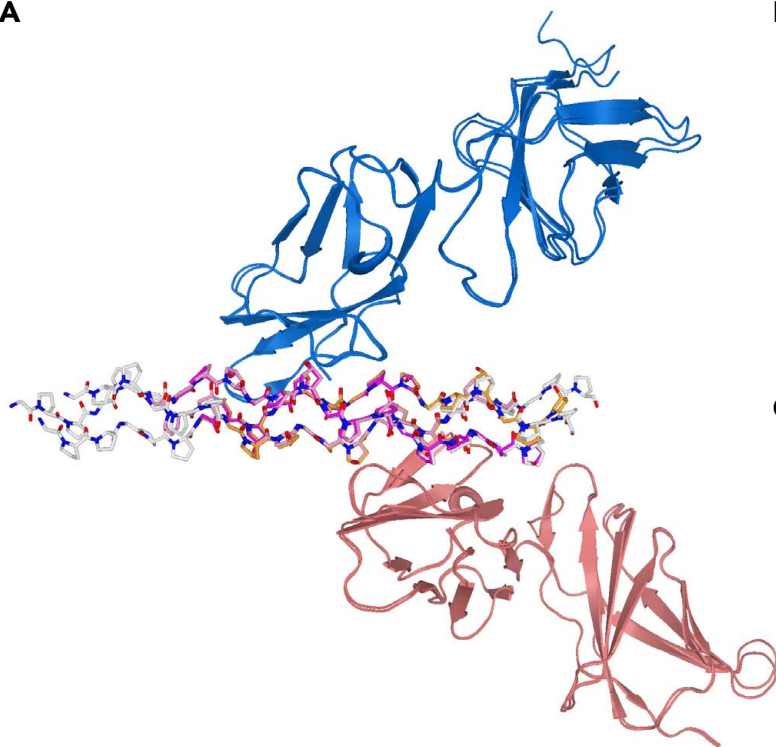
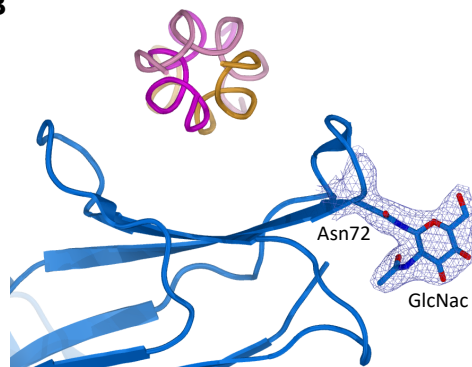
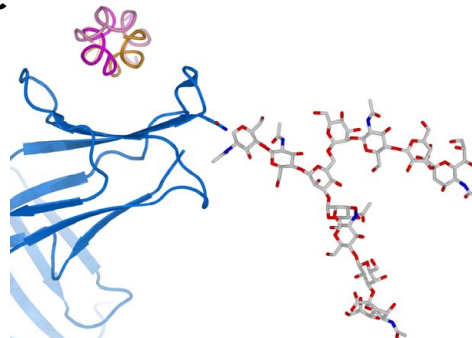
Figure S5**A****B****C**

Figure S5. Structural comparison of GPVI-(GPO)₅ and GPVI-(GPO)₃ complexes and the position of the truncated N-linked glycan. (A) Superposition of the GPVI-(GPO)₅ and GPVI-(GPO)₃ complexes. The two GPVI molecules in each complex are shown as ribbons (blue and salmon); overlaid molecules, shown in the same color, have nearly identical structures except for a very small difference in the orientation of the D2-domains with respect to the D1-domains that does not affect the interaction of D1 with the collagen peptide. The (GPO)₅ triple helix is shown in sticks representation with the *L*, *M* and *T* chains colored pink, magenta, and orange, respectively and presents its *M*+*T* chains to the first (blue) and *L*+*M* chains to the second (salmon) GPVI-molecule. Also shown in sticks representation is the (GPO)₃ triple helix. Coloring of its three GPO-repeats is identical to (GPO)₅, whereas the two (GPP)₂-sequences included for peptide stability at either end of (GPO)₃ sequence are displayed as white sticks. The (GPO)₃ triple helix is rotated like a corkscrew (counter-clockwise) relative to (GPO)₅ and presents its *L*+*M* chains to the first and *T*+*L* chains to the second GPVI-molecule. (B) Position and orientation of the EndoH-truncated N-linked glycan attached to GPVI-residue Asn72 relative to the collagen binding site in the GPVI-(GPO)₅ complex. Asn72 and attached GlcNac monosaccharide are shown in blue sticks together with their 2Fo-Fc electron density contoured at 1.2σ and part of the GPVI D1-domain (blue ribbon). The viewing direction is along the (GPO)₅ triple helix depicted as pink (*L*), magenta (*M*), and orange (*T*) ribbons and interacting with the collagen-binding site underneath (blue cartoon). (C) Modelling of an oligo-mannose glycan antenna on Asn72 (gray sticks) illustrates the size of an intact glycan. Although Asn72 is in the F-F' loop not far from the collagen binding site its side chain and attached glycan point away from the (GPO)₅-peptide site suggesting it may not play a direct role in collagen binding.

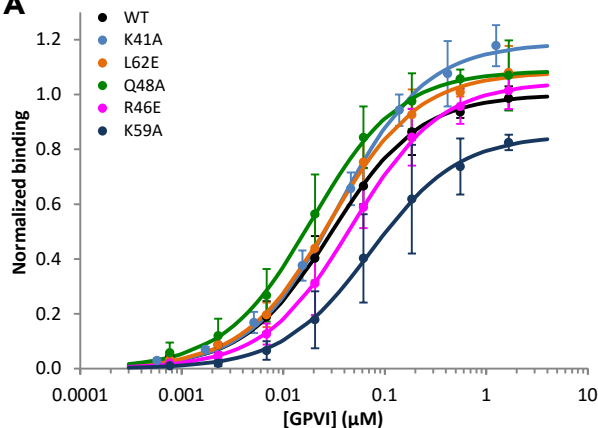
Figure S6**A**

Figure S6. Mutational analysis of GPVI-residues outside the collagen-binding site. Collagen I-binding of wild type GPVI (black) and point mutants of residues located in the groove between β-strands C'-E (colored series) as measured in a solid-state assay. Data measured as A450 was normalized to the binding of wild type protein. Of all tested mutants, only Lys59Ala showed a significant but small decrease in binding, about 2.5-fold, which is a similar decrease as previously measured for the Lys59Glu mutant¹⁷. Because of the small effect of Lys59 mutation on collagen binding combined with similar effects reported for other residues, including Arg60 and Arg166^{18,19} that are located outside the binding site observed in our crystals, we cannot exclude the presence of a secondary low affinity binding site altogether. Data points represent the mean ± SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{max} \times c) / (K_{D,app} + c)$ using non-linear regression in SigmaPlot, where *c* is the GPVI concentration (μM).

Figure S7

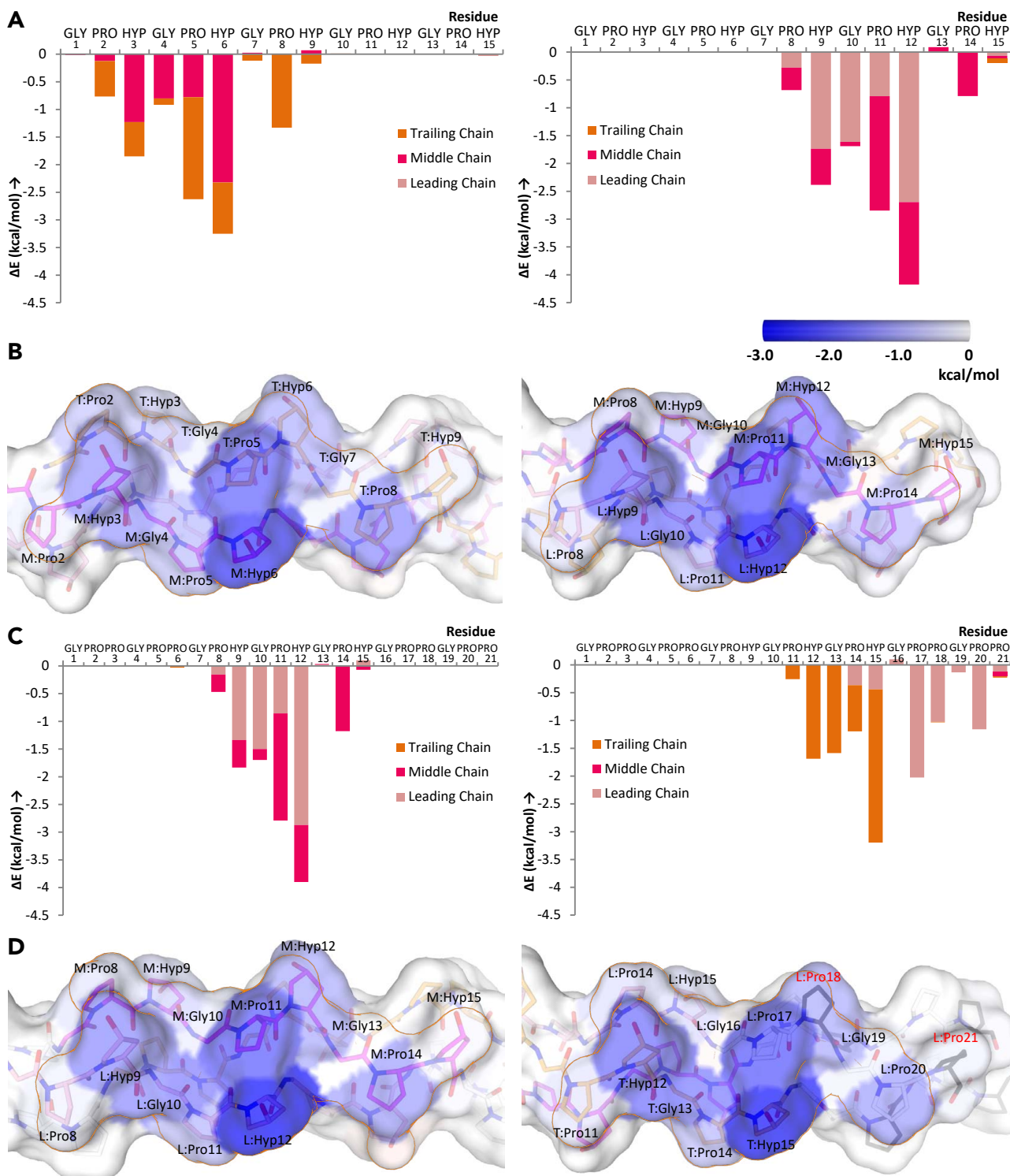
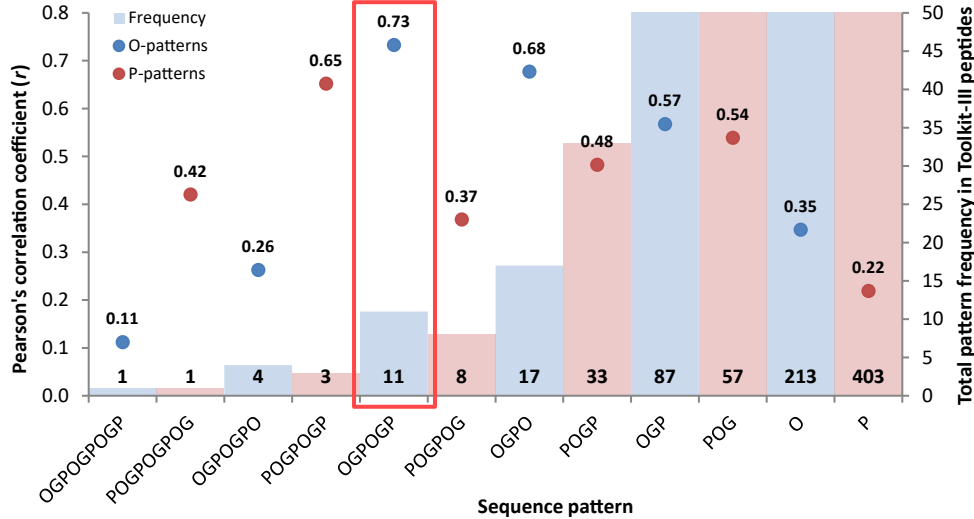


Figure S7. Analysis of interaction energies in the GPVI-collagen complexes using FoldX¹². (A-B) Binding energy contribution of individual collagen peptide residues located in the N-terminal (left) and C-terminal (right) GPVI binding site on (GPO)₅ depicted as a graph (A) and on a (GPO)₅-surface rendering (B) in which surface patches of residues are colored according to the size of their contribution to binding energy from blue (large contribution) to white (no contribution). The C-terminal GPVI binding site on (GPO)₃ is unique in the sense that it contains two Pro residues (red residue labels) at positions occupied by Hyp in the other three binding sites and arising from the presence of a (GPP)₂-sequence at its C-terminus. In all complexes a large contribution to binding energy is provided by a central Pro-Hyp-patch (T:Pro5-M:Hyp6 in (GPO)₅), followed by somewhat smaller contribution of surrounding residues (M:Hyp3, T:Hyp6 and T:Pro8 in (GPO)₅). Residues at the extremes of the binding site (M:Pro2 and T:Hyp9 in (GPO)₅) contribute minimally.

Figure S8

A



B

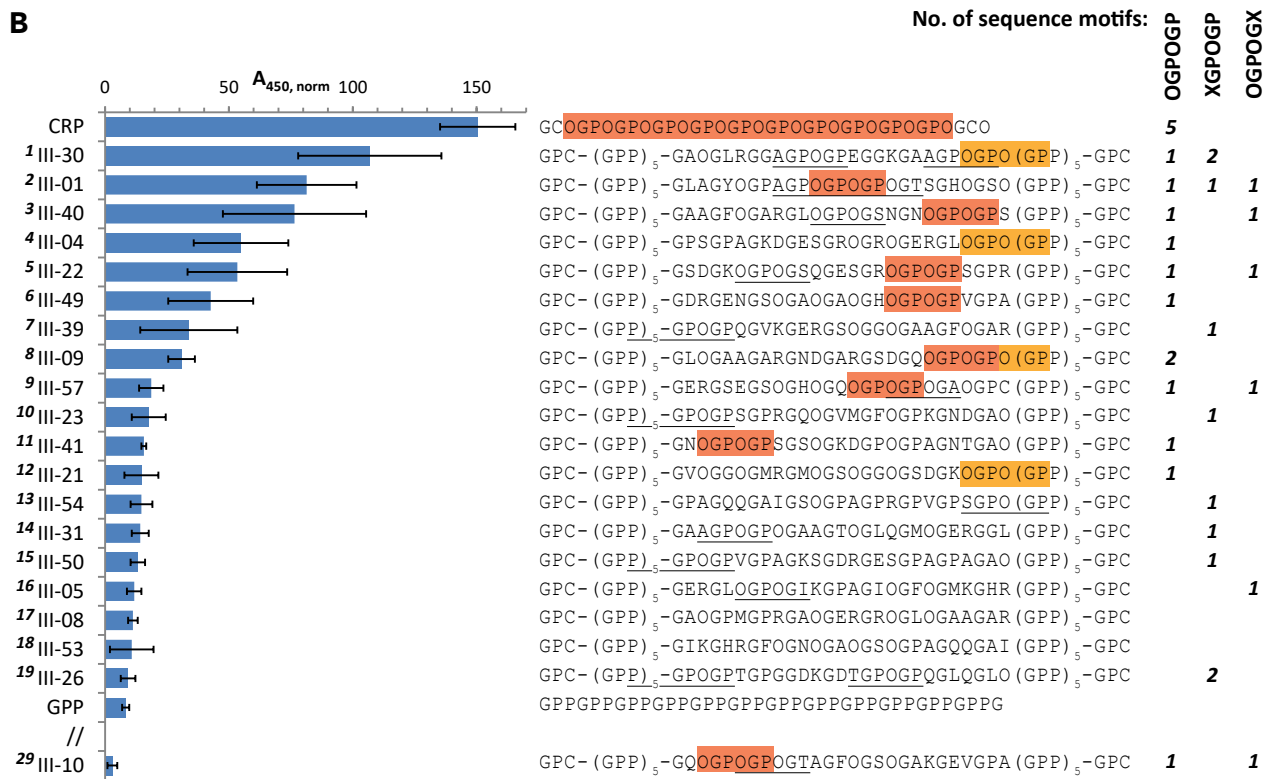


Figure S8. Analysis GPVI-binding sites in Toolkit-III peptides. (A) Pearson's correlation coefficients (r) calculated for sequence patterns of various length starting with hydroxyproline (O; blue) or proline (P; red) as in *supplemental methods*, specifying the correlation between (non-overlapping) pattern frequency in Toolkit-III peptides and GPVI-Fc binding to these peptides (data taken from Jung et al.¹¹). For each sequence pattern, also the total occurrence in Toolkit-III peptides is indicated on the right axis. The correlation obtained for the full POGPOGPO-sequence forming the GPVI contact site in our crystals is only 0.42, indicating that not all residues in the observed contact site are required for binding. The highest correlation (0.73) was found for an OGPOGP-pattern, suggesting it to be a primary recognition sequence in Toolkit-III peptides. **(B)** Previously obtained data on GPVI-binding to Toolkit-III peptides, CRP and (GPP)₁₀¹¹, which was used for the analysis shown in panel A and arranged in order of decreasing binding capacity (left). Shown on the right are the corresponding peptide sequences with OGPOGP-sequences highlighted in orange if present in the collagen III derived region and in light-orange if partially overlapping the generic (GPP)₅-terminus of the peptide. Ten of the eleven OGPOGP-containing peptides rank among the twelve peptides that best support GPVI-binding; the eleventh, III-10, does not display significant GPVI-binding, which is unexpected judged from its amino acid sequence alone. Underlined are partial motifs XGPOGP and OGPOGX, with X being Ala, Ser, Thr, Ile, or Pro. Such a partial motif is present in seven out of 9 peptides lacking the OGPOGP-motif but supporting GPVI-binding better than (GPP)₁₀ does. Also four of the five peptides that best support binding contain one or more partial motifs in addition to (and sometimes overlapping) an OGPOGP motif. Presence of the full and/or partial OGPOGP-motif together explains to considerable extent GPVI-binding capacity of the sixteen top-ranking peptides.

Figure S9

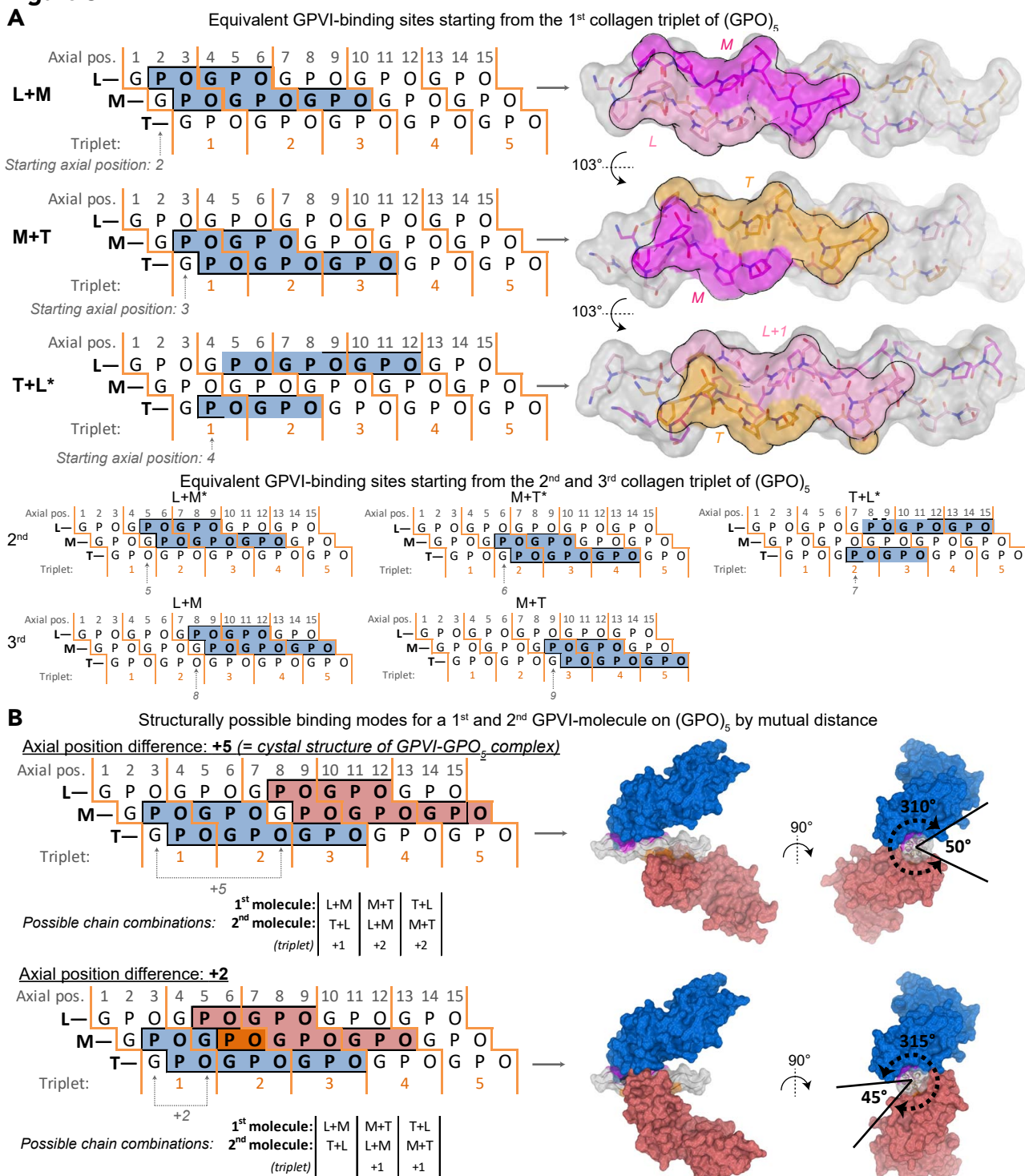
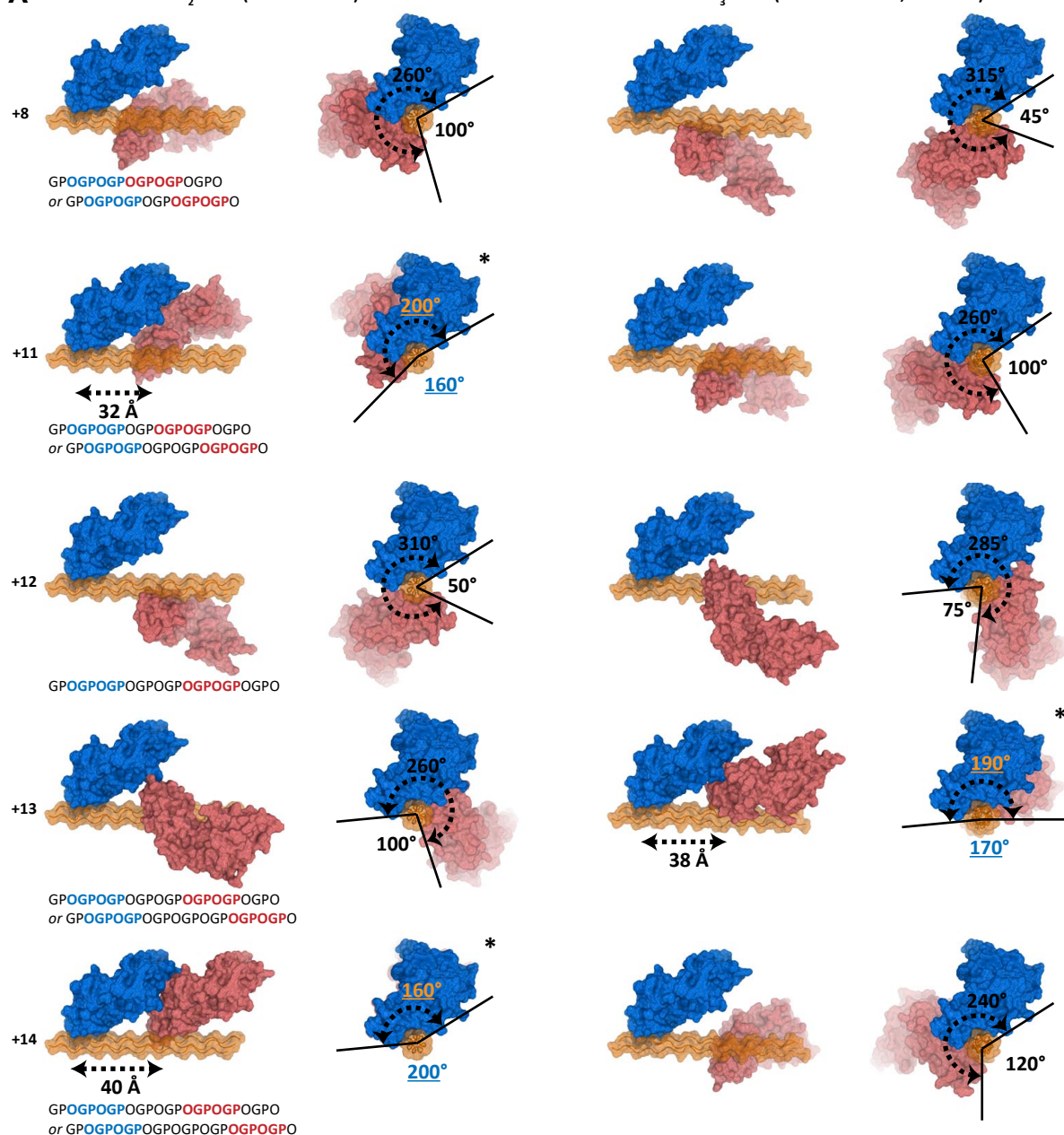


Figure S9. In silico analysis of all possible binding sites on collagen peptide (GPO)₅ and pairs thereof that can bind GPVI simultaneously. **(A)** Distribution of eight equivalent GPVI-binding sites arising from the triple helical and repetitive nature of (GPO)_n peptides. The top panels provide a schematic representation of triple helix (GPO)₅ (left) showing the regions of the collagen chains involved in formation of three binding sites starting from the first GPO-triplet and a surface representation of (GPO)₅ as adapted from our crystal structure (right), highlighting the corresponding regions involved in formation of these binding sites colored by chain: L (pink), M (magenta), and T (orange). The schematic representations also indicate the one residue chain stagger and the axial position that corresponds to the residue numbering of the leading chain. The starting position of each GPVI-binding site is assigned as the axial position of the first proline involved in binding and indicated by an arrow. Successive sites are positioned by a translation of one axial position along the helix in combination with a rotation of 103° around the helix axis. The bottom panels show the schematic representation of binding sites starting from the 2nd and 3rd GPO-triplet of (GPO)₅. On (GPO)₃, the binding site distribution follows the same pattern, but includes in four out of six binding sites one chain of the (GPP)₂-region, indicated with a *. **(B)** Visualization of pairs of GPVI-binding sites on (GPO)₅ that can bind, or likely can bind, GPVI simultaneously without causing steric hindrance. Only two relative arrangements of binding sites appear possible and are depicted schematically (left) and in surface representation (right) with (GPO)₅ (gray) and two bound GPVI-molecules (blue and pink) viewed along and perpendicular to the helix axis. The two arrangements are characterized by a difference in axial starting position of the two binding sites that is +5 or +2, respectively. The first arrangement (+5) is identical to the organization in our crystal structures, the second (+2) shows some steric hindrance between GPVI molecules, which could be relieved by minor rearrangement of the Ser43-Arg46 and Ser74-Leu75 containing loops of GPVI, which seems feasible judging from the mobility of these regions in the crystal structure of GPVI (data not shown). Chain combinations of first and second site that would also lead to the corresponding molecular arrangements are specified in the table below.

A

10₃-helix (PDB-ID: 3DMW; res. 6-17)



B	Mutual distance	1 st site:			Extra triplets	Covered	Covered
		L+M	M+T	T+L		7/2-helix	10/3-helix
		2 nd site*:					
+2	T+L	L+M (+1)	M+T (+1)	0-1	310°	-	
+5	T+L (+1)	L+M (+2)	M+T (+2)	1-2	315°	330°	
+8	T+L (+2)	L+M (+3)	M+T (+3)	2-3	260°	315°	
+11	T+L (+3)	L+M (+4)	M+T (+4)	3-4	200°	260°	
+12	L+M (+4)	M+T (+4)	T+L (+4)	4	310°	285°	
+13	M+T (+4)	T+L (+4)	L+M (+5)	4-5	300°	190°	
+14	T+L (+4)	L+M (+5)	M+T (+5)	4-5	160°	240°	

* numbers between brackets indicate the shift in triplets between sites

Figure S10. Dual binding modes of GPVI-molecules to a single 7₂- or 10₃-helix. (A) Models of two GPVI-molecules binding to a 7₂- (left) or 10₃-type (right) collagen-helix, as obtained by superimposition of GPVI on existing models of collagen peptides having PDB-IDs 1V4F⁶ (7₂-helix) and 3DMW¹⁵ (10₃-helix; residues 6-17) as in supplementary methods. Only presented are models that are sterically feasible, with the GPVI-molecules shown in blue (static molecule) and salmon (mobile molecule) surface representation, and the collagen-helix in orange Ca-traces and surfaces. Each model is displayed in both a side-projection and a projection along the triple helix, of which the latter also indicates the circumferential coverage and accessibility (°). The distance in axial position between two sites on the triple helix is specified per model (left; see also Figure S9). (B) Table summarizing for each sterically feasible model the mutual distance between sites, possible chain combinations to form the 1st and 2nd site, separation in GPO-triplets, and circumferential helix coverage (°) in case of a 7₂ or 10₃ model. Consistent with panel A, models marked with * exhibit circumferential coverage of 200° or less (orange numbers), indicating that a particular binding mode could occur in the context of a collagen fibril.

Figure S11

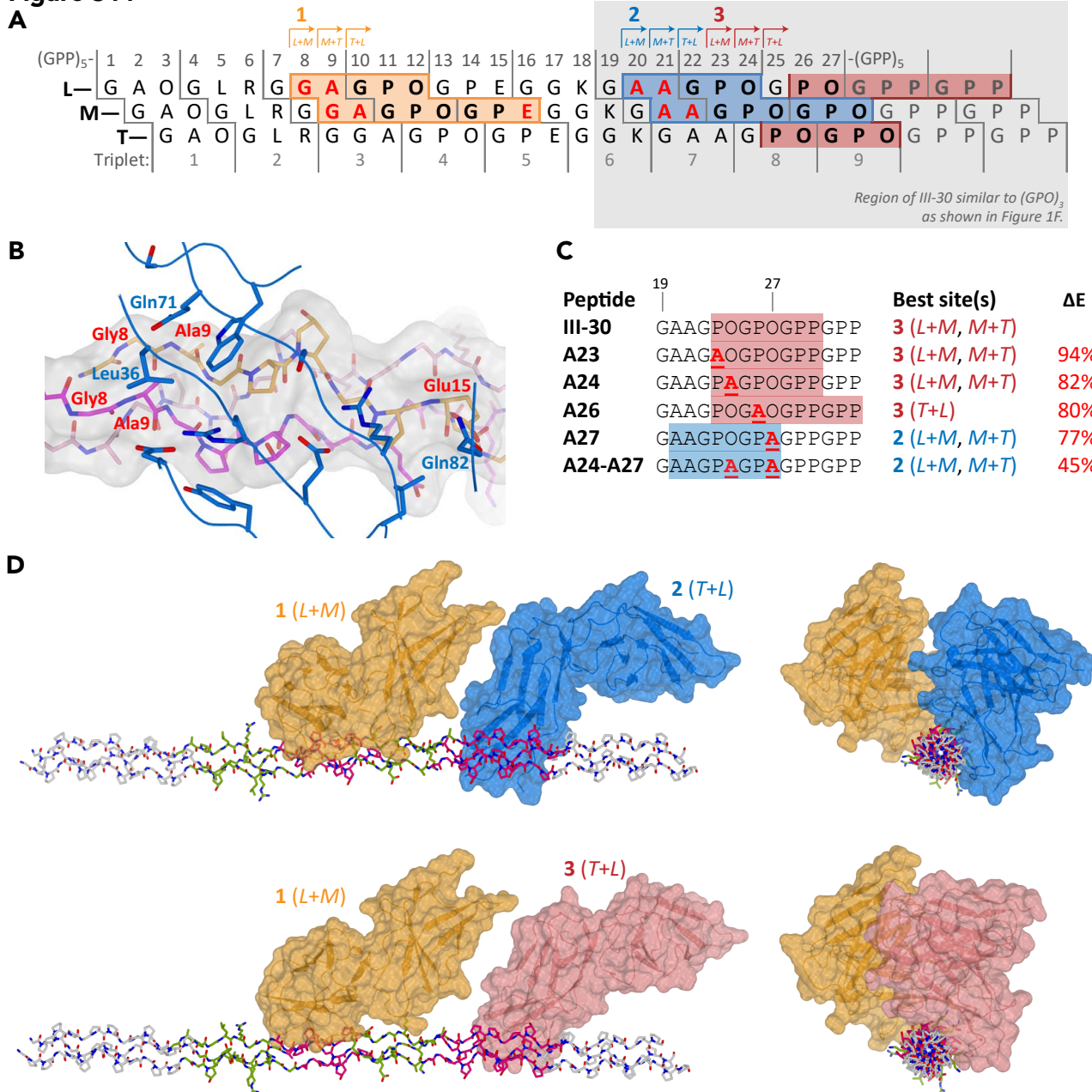


Figure S11. Binding site locations in Toolkit III-30 and effects of mutations and steric hindrance on GPVI-binding. (A) Schematic view of three copies of the amino acid sequence of the collagen III-derived region of III-30 (black) and C-terminal GPP-triplets (gray) arranged with a one-residue stagger as present in the collagen triple-helix. Indicated above are the possible starting positions and chain combinations of GPVI-binding sites constituted by the 9-AGPOGP-14 (#1, orange), 21-AGPOGP-26 (#2, blue), and 24-OGPOGP-29 (#3, salmon) motifs. Each motif gives rise to three overlapping binding sites, only one of which is highlighted in the scheme. Colored in red are residues in these sites that differ from (hydroxy)prolines occupying equivalent positions in the GPVI-binding site in our crystals. Of note, the sequence of the C-terminal region of the peptide (gray box; residues 19-27 and GPP-tail) with the exception of Ala20-Ala21 is identical to (GPO)₃-residues 7-21 (see Figure 1F) and therefore simultaneous binding to sites #2 and #3 is likely possible in situations where the full circumference of the triple helix is accessible, but not in situations when the peptide is incorporated in a fibril or immobilized on a surface as is the case in our solid state binding experiments to III-30 peptides and variants thereof. (B) Modeling of GPVI-binding by the III-30 region comprising the 9-AGPOGP-14 motif. Shown in red are the sequence differences between III-30 and (GPO)₅ at either end of the binding site, which mainly affect interactions with GPVI-residues Leu36, Gln71 at the amino-terminal end and with Gln82 at the carboxyterminal end, while preserving core interactions with GPVI. (C) C-terminal sequence (residues 19-27 and GPP-tail) of III-30 and alanine-scanned variants, indicating per variant the strongest binding site and its binding energy expressed as percentage of the binding energy of III-30 as calculated using FoldX, by omitting amino acids that deviate from (GPO)₅ from the calculations. These calculations indicate that the reduction in binding energy achieved by a single alanine substitution is expected to be moderate ranging from about 6 to 23%, whereas the double alanine variant is predicted to achieve a more substantial two-fold reduction; findings that are in-line with the results from our solid-state binding experiments (see Figure 4c). (D) Models of III-30, generated as in Supplemental Methods, showing the possibility of sites constituted by 9-AGPOGP-14 and either 21-AGPOGP-26 (top panels) or 24-OGPOGP-29 (bottom panels) motifs to independently associate with a GPVI-molecule (colors consistent with panel A). Models are presented in sideview (left panels) and viewed along the triple-helix axis (right panels), showing in both cases a circumferential coverage sterically compatible with helix incorporation in the fibril and immobilization on a surface.

Figure S12

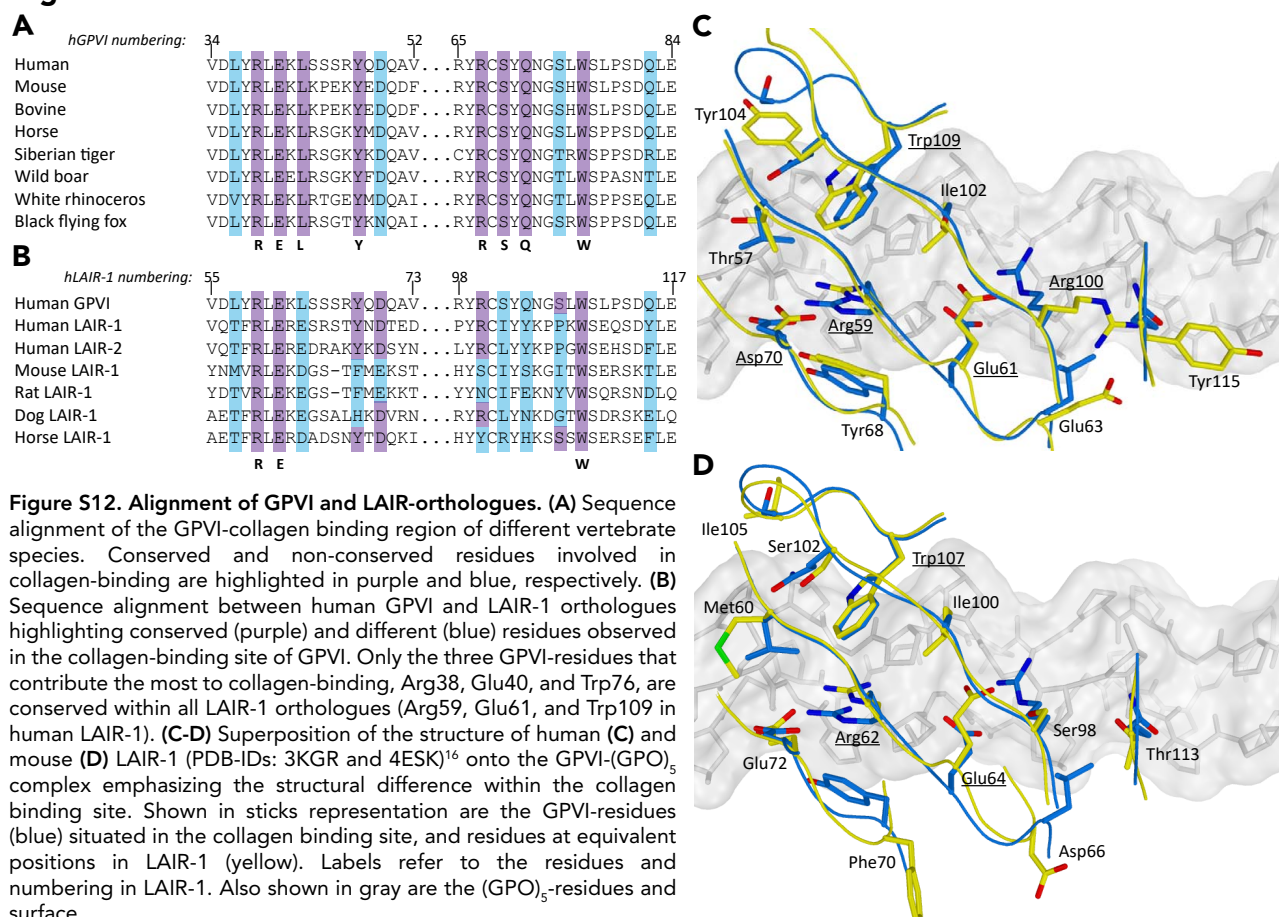


Figure S12. Alignment of GPVI and LAIR-orthologues. (A) Sequence alignment of the GPVI-collagen binding region of different vertebrate species. Conserved and non-conserved residues involved in collagen-binding are highlighted in purple and blue, respectively. (B) Sequence alignment between human GPVI and LAIR-1 orthologues highlighting conserved (purple) and different (blue) residues observed in the collagen-binding site of GPVI. Only the three GPVI-residues that contribute the most to collagen-binding, Arg38, Glu40, and Trp76, are conserved within all LAIR-1 orthologues (Arg59, Glu61, and Trp109 in human LAIR-1). (C-D) Superposition of the structure of human (C) and mouse (D) LAIR-1 (PDB-IDs: 3KGR and 4ESK)¹⁶ onto the GPVI-(GPO)₅ complex emphasizing the structural difference within the collagen binding site. Shown in sticks representation are the GPVI-residues (blue) situated in the collagen binding site, and residues at equivalent positions in LAIR-1 (yellow). Labels refer to the residues and numbering in LAIR-1. Also shown in gray are the (GPO)₅-residues and surface.

Figure S13

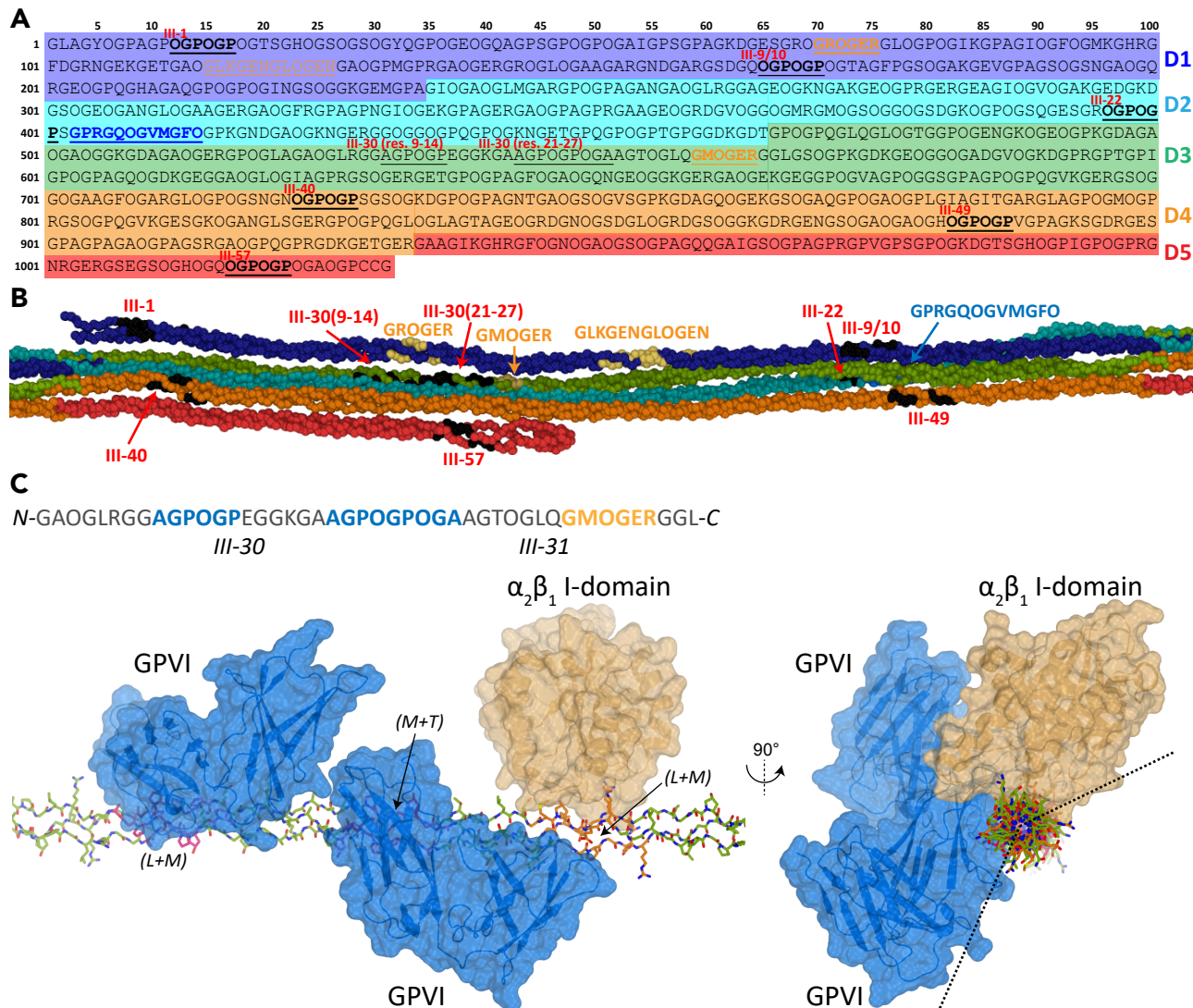
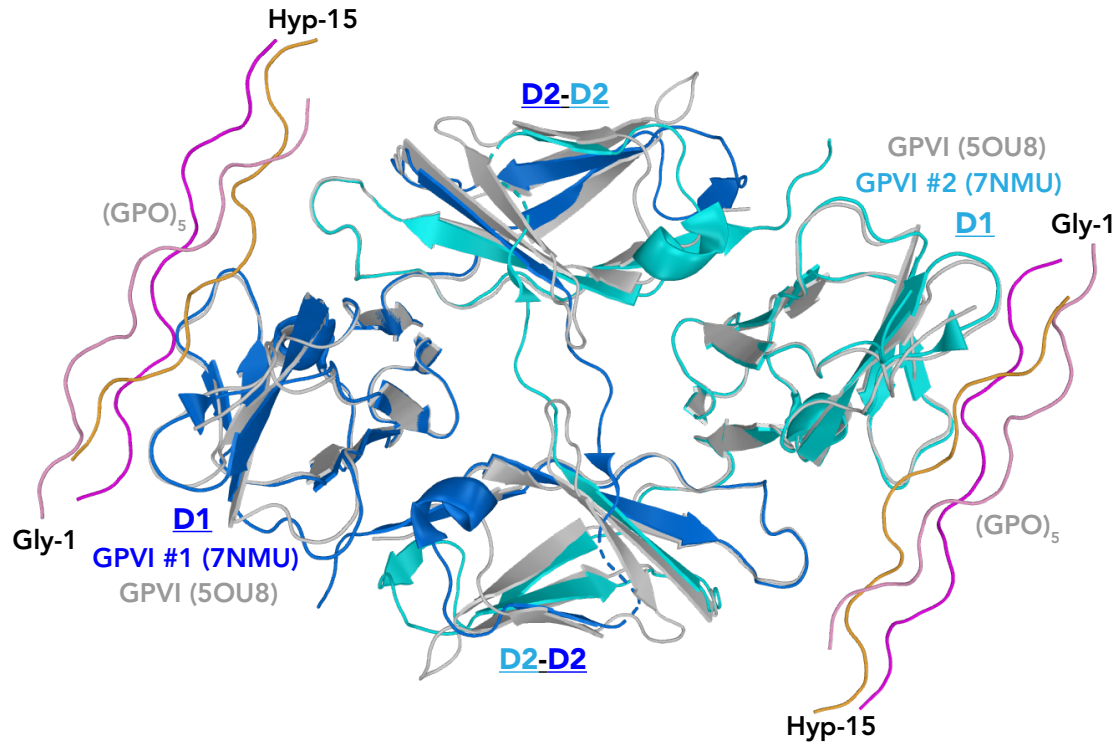


Figure S13. Relative positions of binding sites for GPVI, integrin $\alpha_2\beta_1$ I-domain, and VWF A3-domain in fibrillar collagen III using the *in situ* structure of collagen I²⁰. Full length sequence of the triple helical region of collagen III (A) and side view of one D-segment of the full triple helical region of collagen I²⁰ (B), colored according to the 5 parallel D-segments (blue, light blue, green, orange, and red, respectively). Shown are positions of the OGPOGP-motifs and regions of Toolkit-peptide III-30 involved in GPVI-binding (black, underlined), sequence motifs involved in strong integrin $\alpha_2\beta_1$ I-domain binding (light orange), and the single GPRGQOGVMGFO-motif required for binding of the VWF A3-domain (blue). GPVI-binding sites scatter through all segments of the triple helical region of collagen III, but, the helix packing in the fibril provokes small clusters, in which (different) receptor-sites are positioned in proximity. The OGPOGP-regions of peptide III-9, -22, and -49 are situated close to the VWF-site, and the regions of III-30 and the integrin GMOGER-site (same helix) are situated close to the other strong integrin site, GROGER, and OGPOGP in terminal peptide III-57 (different helices).⁵ Of note, the triple helical regions of heterotrimeric collagen I and homotrimeric collagen III are equal in sequence length, and have equal positions of the VWF-binding site, but vary in the positions of the other receptor binding sites. To map sites on collagen III according to the Toolkit-III binding data, we used the collagen I structure and residue numbering, but transposed the collagen III sequence. (C) Sequence and 10₃ triple helical model (constructed using PDB-ID 3DMW¹⁵, residues 6-17, as in supplementary methods) of collagen III residues 523-567 covered by Toolkit-peptides III-30 and -31, that include GPVI-binding regions AGPOGP and AGPOGPOGA of III-30 (blue) and $\alpha_2\beta_1$ -site GMOGER of III-31 (light orange). Collagen-complexes of GPVI (GPVI-(GPO)₅) and the $\alpha_2\beta_1$ I-domain (PDB-ID: 1DZI) have been superimposed onto the triple helical model by using the L+M, M+T, and L+M combinations of collagen chains for the first and second GPVI-molecule and $\alpha_2\beta_1$ I-domain, respectively. This model shows the feasibility of simultaneous binding of these receptors to one side of the triple helix without steric hindrance and seems compatible with binding to fibrillar collagen under the assumption that the helical twist is 10₃ along the complete length of the peptide. However, though the helical twist of this region is estimated to be 10₃ on average, substantial local deviations may occur due to the presence of GPO-regions, heterogenous triplets, as well as several GG-, GA-, and AA-repeats, which would affect the relative orientation of these sites and prevent accurate modeling of the situation *in vivo*.

Figure S14

A



B

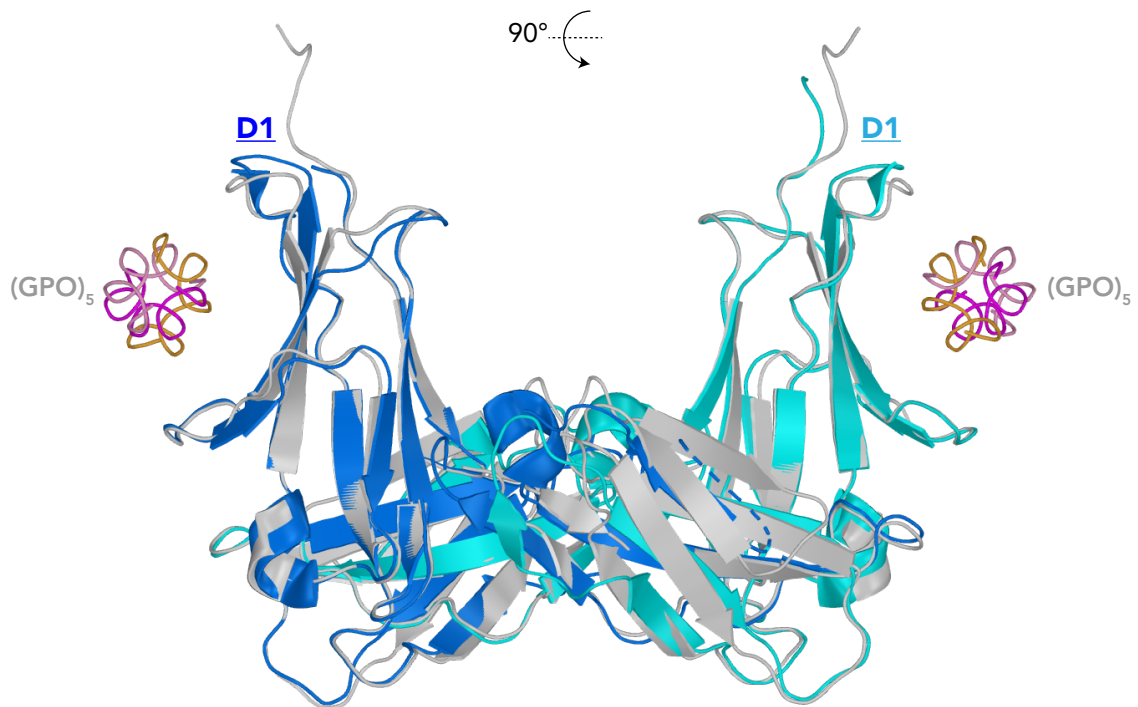


Figure S14. Structural analysis of collagen binding by a domain-swapped GPVI-dimer. Superimposition of a GPVI-molecule (gray) with bound collagen-peptide (pink, magenta, orange) from the GPVI-(GPO)₅ complex onto each of the two GPVI-molecules (blue & cyan) from the D2-domain swapped (GPVI-Nb2)₂-dimer structure (PDB-ID: 7NMU)²¹, using only the D1-domains (res. 1-90) for spatial alignment. The Nb2-molecules were not taken into account. The putative (GPVI-(GPO)₅)₂-dimer viewed perpendicular to the triple helix axis (**A**) and along the triple helix axis (**B**) shows an anti-parallel orientation of bound triple helices, which likely arises from crystallographic packing of the domain-swapped dimer that results in an anti-parallel orientation of D1-domains and collagen-binding sites. The domain swapped dimer observed in the crystal can therefore not represent a situation in which two GPVI-molecules cooperatively bind to parallel helices in fibrillar collagen.