American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Structural insights into collagen-binding by platelet receptor Glycoprotein VI

Tracking no: BLD-2021-013614R1

Louris Feitsma (Utrecht University, Netherlands) Harma Brondijk (Utrecht University, Netherlands) Gavin Jarvis (University of Cambridge, United Kingdom) Dominique Hagemans (Utrecht University, Netherlands) Dominique Bihan (University of Cambridge,) Natasia Jerah (University of Cambridge, United Kingdom) Marian Versteeg (Utrecht University, Netherlands) Richard Farndale (University of Cambridge, United Kingdom) Eric Huizinga (Utrecht University, Netherlands)

Abstract:
Glycoprotein VI (GPVI) plays a vital role in normal hemostasis by mediating collagen-induced platelet activation after vascular damage, but also contributes to the onset of thrombosis, heart attack and stroke. Animal models of thrombosis have identified GPVI as a promising target for antithrombotic therapy. Although for many years the crystal structure of GPVI has been known, the essential details of its interaction with collagen have remained elusive. Here, we present crystal structures of the GPVI ectodomain bound to triple-helical collagen peptides, which reveal a collagen binding site across the β-sheet of the D1-domain. Mutagenesis and binding studies confirm the observed binding site, and identify Trp76, Arg38 and Glu40 as essential residues for binding to fibrillar collagens and collagen-related peptide (CRP). GPVI binds a site on collagen comprising two collagen chains with the core formed by the sequence motif OGPOGP. Potent GPVI-binding peptides from Toolkit-III all contain OGPOGP, weaker binding peptides frequently contain a partial motif varying at either terminus. Alanine-scanning of peptide III-30 also identified two AGPOGP motifs that contribute to GPVI-binding, but steric hindrance between GPVI-molecules regulates the maximum binding capacity. We further show that no cooperative interactions could occur between two GPVI-monomers binding to a stretch of (GPO)$_5$, and that binding of two or more GPVI-molecules to a fibril-embedded helix requires non-overlapping OGPOGP-regions. Our structure confirms the previously suggested similarity in collagen binding between GPVI and Leukocyte-Associated Immunoglobulin-like Receptor 1 (LAIR-1), but also indicates significant differences that may be exploited for the development of receptor-specific therapeutics.

Conflict of interest: COI declared - see note

COI notes: RWF is Chief Scientific Officer of CambCol Laboratories Ltd. The authors declare no other competing conflict of (financial) interest.

Preprint server: No;

Author contributions and disclosures: L.J.F., T.H.C.B., G.E.J., D.H., D.B., N.J., and M.V. performed research; L.J.F., T.H.C.B., G.E.J., R.W.F., and E.G.H. designed experiments; L.J.F., R.W.F., and E.G.H. wrote the manuscript; R.W.F., and E.G.H. supervised the research.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Experimental X-Ray diffraction data and PDB-models for GPVI mutant ΔPAVS-PAPYKN and the GPVI-(GPO)5 and GPVI-(GPO)3 complexes have been deposited to the Protein DataBank with PDB-IDs 5OU7, 5OU8, and 5OU9, respectively.

Clinical trial registration information (if any):

Structural insights into collagen binding by platelet receptor Glycoprotein VI

Louris J. Feitsma[1], T. Harma C. Brondijk[1], Gavin E. Jarvis[2], Dominique Hagemans[1], Dominique Bihan[2], Natasia Jerah[2], Marian Versteeg[1], Richard W. Farndale[2§‡†], and Eric G. Huizinga[1‡]

[1]Department of Structural Biochemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

[2]Department of Biochemistry, University of Cambridge, CB2 1QW, 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.

Short title: Structural insights into collagen-binding by GPVI

Scientific category: Hemostasis, Thrombosis, and Vascular Biology

Word count Introduction: 486

Word count Methods: 608

Word count Results: 2019

Word count Discussion: 701

Word count Total (max 4000 for Introduction-Discussion): 3814 words

Word count Abstract: 241 words

Number of figures: 5

Number of tables: 2

Number of references: 52

This manuscript contains supplemental data, provided as a separate PDF-document.

Keypoints

1. Crystal structures of GPVI bound to collagen peptides reveal its primary collagen binding site across the D1-domain β-sheet. **(127 characters)**
2. GPVI binds sites in collagen formed by two of the three triple helix chains and canonical OGPOGP sequence motifs. **(116 characters)**

Abstract

Glycoprotein VI (GPVI) mediates collagen-induced platelet activation after vascular damage, and is an important contributor to the onset of thrombosis, heart attack and stroke. Animal models of thrombosis have identified GPVI as a promising target for antithrombotic therapy. Although for many years the crystal structure of GPVI has been known, the essential details of its interaction with collagen have remained elusive. Here, we present crystal structures of the GPVI ectodomain bound to triple-helical collagen peptides, which reveal a collagen-binding site across the β-sheet of the D1-domain. Mutagenesis and binding studies confirm the observed binding site, and identify Trp76, Arg38 and Glu40 as essential residues for binding to fibrillar collagens and collagen-related peptide (CRP). GPVI binds a site on collagen comprising two collagen chains with the core formed by the sequence motif OGPOGP. Potent GPVI-binding peptides from Toolkit-III all contain OGPOGP, weaker binding peptides frequently contain a partial motif varying at either terminus. Alanine-scanning of peptide III-30 also identified two AGPOGP motifs that contribute to GPVI-binding, but steric hindrance between GPVI-molecules restricts the maximum binding capacity. We further show that no cooperative interactions could occur between two GPVI-monomers binding to a stretch of $(GPO)_5$, and that binding of two or more GPVI-molecules to a fibril-embedded helix requires non-overlapping OGPOGP-motifs. Our structure confirms the previously suggested similarity in collagen binding between GPVI and Leukocyte-Associated Immunoglobulin-like Receptor 1 (LAIR-1), but also indicates significant differences that may be exploited for the development of receptor-specific therapeutics.

Introduction

Collagen exposed at sites of vascular injury contributes to thrombus formation. In response to vessel damage, platelet receptors GPIbα and integrin $\alpha_2\beta_1$ support firm platelet adhesion to collagens of the subendothelial extracellular matrix (ECM): GPIbα associates with collagen-bound von Willebrand factor (VWF) whereas integrin $\alpha_2\beta_1$ binds directly to exposed collagen[1,2]. Subsequent interaction between collagen and the glycoprotein VI–Fc-receptor γ-chain (GPVI–FcRγ) complex[3–5], triggers phosphorylation of FcRγ, ultimately resulting in full platelet activation[6–9]. GPVI itself is present in both monomeric and dimeric form in the platelet membrane, and activation by several ligands shifts the population towards the dimeric species, detected by dimer-specific antibodies[10,11]. Deficiency or inhibition of GPVI results in decreased thrombus deposition and platelet aggregation[12], emphasizing its role in platelet activation.

Fibrillar collagens provide strength to the arterial vasculature, whilst the role of exposed collagens in platelet adhesion and activation reflects its hemostatic function[13]. The triple-helical structure of collagen requires the repeating Gly-X-Y sequence, where, most frequently, X is proline (P) and Y is the post-translationally modified 4-hydroxyproline (O). Toolkits-II and -III contain homotrimeric peptides covering the full sequences of the collagen II and III triple-helical regions, respectively, and were used to identify collagenous binding motifs for several proteins[14] including integrin $\alpha_2\beta_1$[15], ECM-protein SPARC[16,17], discoidin domain receptor 2[18], the VWF A3-domain[19,20], and the Osteoclast-associated receptor (OSCAR)[21]. Binding of GPVI to Toolkit-III peptides[10,22] is notably more scattered than observed for other receptors[14]. In addition to several peptides that support only modest GPVI-binding, III-30, III-01, and III-40 exhibit high affinity for GPVI[10]. High imino acid content is not restricted to these peptides; nevertheless, the existence of sequential GPO/OGP-triplets appears a key determinant for GPVI-mediated platelet activation[22]. These conclusions are further supported by the increasing GPVI-binding capacity of model peptides containing consecutive

GPO-repeats, ranging from $(GPO)_2$ to $(GPO)_{10}$; the latter being the collagenous domain of collagen-related peptide (CRP)[23,24]. However, these studies could not unambiguously identify a GPVI-binding motif[10,22].

GP6 is located in chromosome 19 in the leukocyte receptor complex (LRC)[5] alongside many other immune receptor genes including the structurally-related but functionally separate collagen-binding receptors, OSCAR[25] and Leukocyte-Associated Immunoglobulin-like Receptors 1 and 2 (LAIR-1 and -2)[26]. Activatory receptors GPVI and OSCAR share the same extracellular domain architecture, having two C2-type Ig-like domains (D1 and D2). Collagen binds both domains of OSCAR[21], with D2 having higher affinity, whereas binding to GPVI was reported to be primarily mediated by the hinge region of its D1-domain[27–30]. LAIRs are immune-regulating receptors and have a single Ig-like domain (D1) that contains a collagen-binding site across the β-sheet formed by strands FCC'[31]. Some LAIR-1-binding Toolkit-peptides also bind either GPVI or OSCAR, but not both, suggesting discrete collagen-binding mechanisms for each receptor[10,22,25,32].

Here, we present crystal structures of GPVI in complex with collagen-like peptides that provide the structural basis for collagen binding to the canonical OGPOGP-sequence of collagen and which account for the propensity of GPVI to bind diverse sites in collagen.

Methods

Expression and purification of GPVI D1D2 WT and mutants

For crystallization, GPVI loop truncation mutants were made in a construct encoding the D1D2-domains (Uniprot-ID Q9HCN6-3; Gln21-Ser206), using a QuikChange Site Directed Mutagenesis Kit and the primers listed in Table S1. GPVI was expressed with N-terminal Cystatin-S signal peptide (Uniprot-ID P01036; Met1-Ala20) followed by TEV-cleavable His_6-$StrepII_3$-tag in HEK293-EBNA1-S cells (U-Protein Express). Proteins were purified on a StrepTactin-column (GE Healthcare) and eluted with buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl and 5 mM *d*-desthiobiotin. Purification tags were removed by overnight incubation with 1:30 mol/mol of His_6-TEV-protease/GPVI in the same buffer. Uncleaved protein and TEV-protease were removed by immobilized metal affinity chromatography. Proteins were purified further by size-exclusion chromatography on a Superdex75 column (GE Healthcare) equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl and concentrated to 10 mg/mL.

For binding studies, the wild-type GPVI fragment described above and point-mutants generated by QuikChange Site-directed mutagenesis were transiently expressed with N-terminal Cystatin-S signal peptide and C-terminal TEV-cleavable immunoglobulin Fc-tag (Uniprot ID P01857; Glu99-Lys330) in HEK293-EBNA1 cells (U-Protein Express). GPVI-Fc dimers were purified by Protein A affinity chromatography (GE Healthcare), eluted with 0.1 M glycine pH 3.0; fractions were collected into 1/10 volume of neutralization buffer (1 M Tris-HCl pH 9.0). Protein purity was determined as >95% by SDS-PAGE. Proteins were stored in neutralized elution buffer at 277 K, or, for long-term storage, at 193 K after flash-freezing in liquid nitrogen.

Peptide synthesis and melting point determination

Peptides were synthesized in an Applied Biosystems Pioneer solid-phase synthesizer, using Fmoc-chemistry as previously described[15,22], or were from CambCol Laboratories, Ely, UK. Peptides were purified by HPLC and verified using MALDI-TOF mass spectrometry. To verify the triple-helical authenticity of the peptides in solution, T_m was determined by polarimetry[15] (Table 1). T_m represents the temperature at which half of the peptide has unfolded from triple-helical to random coil conformation, and high T_m indicates accurate collagen-like structure.

Collagen binding assay

Enzyme-linked immunosorbent assays (ELISA) measuring the binding of dimeric GPVI-Fc mutants to human placental collagen type I (Sigma Aldrich) or collagen peptides were performed as described[22]. Details of fibril preparation, binding measurements, and data processing are provided in *Supplementary methods*.

Crystallization and data collection of GPVI alone and in complex with homotrimeric collagen peptides

Before crystallization, GPVI loop-truncation mutant ΔPAVS-PAPYKN (10 mg/mL in gel-filtration buffer) was deglycosylated overnight at 292 K by addition of 1:100 (v/v) EndoH_F (NEB; 1×10^6 unit/mL). GPVI crystals were obtained in approximately one month by sitting-drop vapor diffusion with well solution containing 0.1 M phosphate-citrate buffer pH 4.0 and 40% (v/v) PEG-300 at 292 K. Crystals were flash-frozen in liquid nitrogen without further cryo-protection. Diffraction data to 1.9 Å resolution were collected on the PX-beamline at the Swiss Light Source.

For co-crystallization of GPVI with collagen peptide $(GPO)_5$ or $(GPO)_3$ (Table 1), protein and peptide were mixed in a 1:1.2 molar ratio. Crystals of the GPVI-$(GPO)_5$ complex were grown at 292 K with crystallization buffer containing 0.1 M MES pH 6.0, 0.2 M $MgCl_2$, and 20% (w/v) PEG-6000. Crystals of the GPVI-$(GPO)_3$ complex were grown with buffer containing 0.1M MMT (DL-Malic acid, MES monohydrate, Tris) pH 9.0 and 25% (w/v) PEG-1500. Crystals were cryo-protected in crystallization buffer containing 20% (v/v) glycerol before flash-freezing in liquid nitrogen. Diffraction data with 2.5 Å resolution limits for both complexes were collected on the ID23-1 beamline at the European Synchrotron Radiation Facility (ESRF). Details about data processing and structure solving are provided in *Supplementary methods*. Experimental data and models for GPVI ΔPAVS-ΔPAPYKN and the GPVI-$(GPO)_5$ and GPVI-$(GPO)_3$ complexes were deposited in the Protein DataBank with PDB-IDs 5OU7, 5OU8, and 5OU9, respectively.

Results

Crystal structure determination of GPVI collagen-peptide complexes

To obtain crystals of GPVI-collagen peptide complexes, it proved essential to truncate two loops in GPVI-D2 that are flexible, judged from the published crystal structure (PDB-ID 2GI7; Figure S1a)[27]. From eight mutants tested, we selected ΔPAVS-ΔPAPYKN that lacks residues 102-105 and 131-136 (Figure S1b-d), which showed least decrease in expression yield (20%) and greatest increase in thermal stability (2.5 degrees). Its CRP- and collagen-binding affinity are both reduced about 5- to 6-fold (Figure S1d-e), measured using dimeric Fc-fusions of the mutant and wild-type protein, as in Methods.

We obtained structures of the GPVI ΔPAVS-ΔPAPYKN mutant alone (1.9 Å resolution; Figures 1a & S2a-c) and in complex with collagen peptides $(GPO)_5$ and $(GPO)_3$ (both at 2.5 Å resolution; Figures 1b-c, S3a-i & Tables 1 & S2). The complexes with collagen peptides exhibit the same crystal form with two prominent structural features (Figure S4a-c): two GPVI-molecules bind one triple-helical peptide with the same D1-domain region (Figures 1b-c & S5a), and each of these molecules additionally forms back-to-back D2-interactions with a symmetry-related receptor-peptide complex. The truncated D2-loops have recently been implicated in a dimer-switch mechanism[33], but in both situations they are located so that even the longer loops in the wild-type protein could not reach the collagen peptide. Therefore, the mildly-reduced collagen binding of the mutant used for crystallization cannot directly be explained from the crystal structures.

Similar interactions between back-to-back D2-domains have been observed in this and a previous report[27] of collagen-free dimeric structures (*see* Figure S4c); in the present collagen-bound form the angle between collagen helices is large, about 80° (Figure 1d). The rigidity of this assembly impedes the re-orientation of its D1-domains required to allow docking of the dimer onto the parallel helices that occur in fibrillar collagen[34,35]. The two GPVI molecules that bind a single peptide make no contact with each other, so that no cooperative interactions can favor the observed assembly. Instead, the absence of steric hindrance between GPVI monomers may permit this complex formation. Both structural features stabilize the observed crystal form, but neither would facilitate cooperative collagen binding *in vivo*[10,11].

The GPVI-binding site on collagen peptides

A GPVI-binding site is comprised of two of the three collagen chains and has an area of about 470 Å^2. In the GPVI-$(GPO)_5$ complex, one GPVI-molecule binds near the peptide N-terminus to a continuous surface comprising Pro2-Hyp6 and Pro2-Hyp9 residues of the middle (*M*) and trailing (*T*) chains, respectively; the other molecule binds to an equivalent site close to the C-terminus comprising residues Pro8-Hyp12 of the leading (*L*) chain and Pro8-Hyp15 of the middle chain (Figure 1d). In the GPVI-$(GPO)_3$ complex, one GPVI-molecule binds to the *L*+*M* chains; a second to the *T*+*L* chains (Figure 1e). Interestingly, the latter site has a proline originating from the C-terminal $(GPP)_2$-extension of the peptide (see Table 1 and Figure 1d) at a position that is occupied by hydroxyproline in the other $(GPO)_3$- and both $(GPO)_5$-binding sites. In these sites the hydroxyl group does not contact GPVI, and therefore the presence of proline at this position likely does not influence binding.

The collagen binding site of GPVI

The collagen-binding site of GPVI is located across the D1 β-sheet, comprising strands C, D, F, and F' (Figure 2a). Its position differs from the collagen-binding site predicted from the crystal structure of GPVI alone[27], but is consistent with mapping of the collagen binding site in LAIR1[31]. Eight central residues are either fully or highly conserved in GPVI orthologues (Figure 2b). Among them, Trp76 forms extensive hydrophobic interactions with *T*:Pro5[§] and a hydrogen bond with the backbone carbonyl of *T*:Hyp3 (Figure 2c-e). Arg38 and Arg67 form hydrogen bonds with carbonyl oxygens of the $(GPO)_5$-backbone. Gln71 and Glu40 are hydrogen-bonded to the hydroxyl groups of *M*:Hyp3 and *M*:Hyp6, respectively (two occurrences of O in P<u>O</u>GP<u>O</u>GPO), and appear essential for the hydroxyproline-specificity of collagen substrates. The Tyr47 side chain is positioned close to *M*:Pro5 and *M*:Hyp6 and buries a substantial hydrophobic surface. Ser69 and Leu42 have hydrophobic interactions with *T*:Pro5 and *T*:Pro8, respectively. Four non-conserved residues, Leu36, Asp49, Ser74, and Gln82, provide minor interactions at the periphery of the binding interface. Finally, the Asn72 side chain and attached N-acetylglucosamine remaining after EndoH cleavage of its N-linked glycan are close to, but oriented away from the collagen peptide (Figure S5b-c) and although its presence may contribute to stability of the F-F' loop, the N-glycan is unlikely to play a direct role in collagen binding.

Effect of GPVI-mutations on collagen binding

The effect of mutations in the binding site was measured in solid-state assays using three collagen substrates that bind wild type GPVI-Fc dimers with nanomolar affinity: fibrillar collagen I (K_D = 30 nM), the synthetic agonist CRP (K_D = 2.7 nM), and Toolkit peptide III-30 (K_D = 11 nM). Binding to all three ligands is drastically reduced by nearly all alanine substitutions of conserved collagen-binding residues (Figures 3a-c and Table 2). Profound reductions were observed for mutants of Trp76, Arg38 and Glu40, showing greater than 600-fold increase in K_D relative to wild type protein. Moderate reductions in affinity (17- to 360-fold) were observed for the Arg67, Tyr47, Ser69, and Asp49 mutants. Asp49 itself has little direct contact with collagen, but, by forming a salt bridge, appears to fix Arg38 for optimal interaction, thus explaining the substantial effect of Asp49 mutation. In comparison to these substantial effects, alteration of residues inside the previously suggested C'-E groove[27] affected collagen binding negligibly (Figure S6 and Table S3) confirming that the crystal structures of the GPVI-collagen complexes reveal the authentic collagen binding site.

Small but markedly substrate-dependent effects were observed upon mutation of Leu36, Gln71, and Gln82. Consistent with its peripheral location and lack of interaction in our crystal structure mutation of Gln82 had no effect on binding to CRP. It did, however, significantly reduce binding to collagen I (14-fold) and III-30 (27-fold). Apparently, binding to more complex substrates exposing residues other than hydroxyproline at the C-terminal periphery of the binding site does involve Gln82 interactions, albeit not very strong. Mutation of Gln71 and Leu36 at the opposite end of the binding site had no effect or even slightly improved binding to III-30, whereas they reduced binding to both collagen I (2- and 6-fold, respectively) and CRP (both 7-fold). The modest effect of these mutations even on CRP-binding indicates that the observed Gln71-*M*:Hyp3 hydrogen bond and Leu36-*M*:Hyp3 hydrophobic interactions contribute little to substrate binding. In combination, the substrate dependent effects of Leu36, Gln71, as well as Gln82 mutants reveal a role for sequence variation at the amino and carboxy-terminal periphery of the binding site in finetuning binding affinity.

[§] One-letter prefixes *L*:, *M*:, and *T*: are used to refer to amino acids in the leading, middle, and trailing chain, respectively.

An OGPOGP-motif with minor peripheral sequence variation forms the core GPVI-binding site

The 8-residue POGPOGPO-sequence that forms the contact site in our (GPO)$_5$-GPVI crystals ($d < 4.5$ Å; *see Figure 1d*) occurs once in the fibrillar region of collagen III[15]; nevertheless seven[22] or more[10] Toolkit-III peptides bind GPVI significantly. GPVI binding can apparently be mediated by a shorter sequence and/or sequences with one or more amino acid substitutions, as was also suggested by the differential effect of several GPVI mutants on binding to different collagen substrates. Calculations using the FoldX-suite[36] indicate that the first proline and last hydroxyproline of POGPOGPO together contribute only 6% of the binding energy (Figures 4a-b & S7a-d) and can be replaced by any other amino acid without causing steric clashes with GPVI (see *supplementary methods* for details). The shorter OGPOGP-motif occurs 11 times in collagen III and its presence in Toolkit-III peptides shows better correlation with published binding data[10] than any other shorter sequence does (Figure S8a and *supplementary methods*).

GPVI binding to sequence variants of OGPOGP with either of the terminal residues replaced by a residue equal to or smaller in size is evident (in hindsight) from previous studies[10,22] (Figure S8b) and underpinned by our alanine-scanning study of peptide of III-30 (Figure 4c & Table 1). We observed the strongest reduction of GPVI-binding for single-residue mutants of Pro11 (-31%) and Hyp12 (-56%), proving the importance of the *9*-AGPOGP-*14* sequence for GPVI-binding. Modeling of this site in the GPVI-collagen complex reveals disruption of hydrogen bonding and hydrophobic interactions with GPVI-residues Gln71 and Leu36 due to the Hyp-Ala replacement at the N-terminus of the motif (Figure 4d and *supplementary methods*), in line with our previous observation that these GPVI-residues do not contribute to III-30 binding (*see* Figure 3c). Modeling also indicates that Glu15, located directly C-terminal of the AGPOGP-motif, can form an additional hydrogen bond to GPVI-residue Gln82, mutation of which markedly reduced binding to III-30, but did not affect CRP-binding. These results demonstrate that the 6-residue OGPOGP-motif implicated in high-affinity GPVI binding can tolerate conservative substitution of a peripheral residues, whereas additional interactions with GPVI of residues just outside the core motif can further stabilize binding.

GPVI-collagen binding is modulated by steric effects

The collagen fibril architecture[34,35] and the D1-D2 inter-domain angle of GPVI impose steric constraints that restrict the ability of GPVI to bind simultaneously to closely-spaced binding motifs *in vivo* (Figure 5a). At the collagen fibril surface, binding sites arising from identical motifs in parallel triple helices occur repetitively along the circumference of the fibril. Simultaneously binding to nearby sites on parallel helices less than 34 Å apart is prohibited by the bent shape and observed binding-orientation of GPVI (Figure 5b). Binding to closely-spaced sites in the same triple helix as observed in our crystals is not feasible *in vivo* because the two GPVI molecules cover almost the entire helix circumference (Figures 5c & S9a-b) whereas in a collagen fibril only a small part of the helix circumference is accessible to GPVI. Binding to the same side of the triple helix requires a much larger distance between GPVI-molecules, since its bent shape causes steric hindrance by D2. Depending on local collagen helicity, the minimum center-to-center distance between two OGPOGP binding sites should be 32-40 Å (Figure 5d), i.e., at least two triplets (*see* Figure S10a-b and *supplementary methods*). The shortest collagen fragment allowing association with two GPVI-molecules simultaneously should therefore contain six triplets, longer than any naturally occurring GPO-repeat and thus must comprise discrete binding motifs.

Our alanine-scanning study with Toolkit peptide III-30 reveals that solid-phase experiments employing randomly-immobilized peptides closely mimic these inferred steric constraints (*see* Figure 5a). Simultaneous knock-out of the C-terminal *21*-AGPOGP-*26* and overlapping *24*-OGPOGP-*29* motif using a double Hyp24-Hyp27 substitution profoundly reduced GPVI binding (-81%) (*see* Figure 3c). Single substitutions in either of these motifs, however, had only minor effect (-4 to -22%). These results imply that exposed sites arising from both motifs can be occupied, but not simultaneously (Figure S11a-c). Apparently, the overlapping motifs, despite being identical in length to the co-crystallized $(GPO)_3$-motif cannot associate with two GPVI-molecules due to the steric constraints imposed by surface immobilization. The more N-terminal *9*-AGPOGP-*14* site binds GPVI independently, consistent with the conclusion above, that binding sites should be at least two triplets apart (Figure S11d). Therefore, steric effects observed in GPVI-binding experiments to immobilized peptide III-30 are consistent with predictions for GPVI-binding to fibrillar collagen.

Comparison of GPVI and LAIR collagen binding sites

The previously-mapped collagen-binding site of LAIR-1 is consistent with its position in GPVI, although its sequence is more variable within the LAIR family than within the GPVI family. The only residues strictly conserved within and between both groups are Trp76, Arg38 and Glu40 (GPVI-numbering; Figure S12a-b). As for GPVI, mutation of these residues in LAIR-1 diminishes collagen binding drastically[31] and therefore, the Trp-Arg-Glu triad appears to be the core element in collagen binding by both receptor families. Additionally, GPVI residues Arg67, Tyr47, and Asp49 are conserved in human LAIR-1 and -2, but not in other species. In line with the extent of sequence divergence, the collagen-binding site of human GPVI more closely resembles crystal structures of human LAIR-1 (PDB-ID: 3KGR[31]) than mouse LAIR-1 (PDB-ID: 4ESK[37]) (Figure S12c-d). These structural differences and substantial divergences in Toolkit- and CRP-binding properties of both receptors[32] underline the observation that peripheral sequence variation finetunes collagen-binding affinity and site-specificity.

Discussion

The platelet receptor GPVI binds to exposed collagens at sites of vascular injury provoking downstream signaling and platelet aggregation[12], while inhibition or deficiency of GPVI protects from thrombus formation without severe bleeding disorders[38,39], indicating its potential value as a therapeutic drug target. Our crystal structures present the structural basis for GPVI-binding to the canonical collagen sequence $(GPO)_n$. In contrast to structures available to date of proteins in complex with collagen peptides, which all involve a collagen residue with a long side chain, i.e. phenylalanine (VWF-A3[20], SPARC[17], and OSCAR[21]), glutamate ($\alpha_2\beta_1$[40]), or arginine (Hsp47[41,42]), our structures reveal a novel collagen-binding mode employing the interaction of a rather flat surface across the D1 β-sheet of GPVI with several modestly-sized hydroxyprolines in collagen.

Combined with previous Toolkit-III binding studies[10,22], our data support OGPOGP as the template motif in collagen which permits strong *in vitro* binding by GPVI, whilst allowing minor attenuating variations in its terminal positions. Such motifs, occurring in all collagens, are distributed throughout all five segments of the triple-helical domain of collagen III[34] and appear to form mixed clusters of binding sites for GPVI and other receptors (Figure S13a-b). The III-22 OGPOGP- and VWF-A3 motifs[20] and the III-30- and integrin $\alpha_2\beta_1$-motifs[15] exemplify positioning of sites sequentially on the same helix. Fibrillar stacking[34] also adds OGPOGP-motifs of III-9, III-49, III-57, and the $\alpha_2\beta_1$ GROGERmotif to these clusters. Possibly, clustering of binding sites facilitates simultaneous receptor binding and interplay *in vivo*, as e.g. proposed for $\alpha_2\beta_1$ and GPVI[43–45].

Authenticity of *in vivo* GPVI-binding sites in collagen III and heterotrimeric collagen I will be critically dependent on both the exposure of helix segments on the fiber surface at areas of vascular lesion (reviewed in Herr & Farndale[35]) and the feasibility of concerted receptor binding to clustering sites. The overlapping III-30 sites – AGPOGP and OGPOGA in the native collagen III sequence – will allow binding of one GPVI-molecule to the site supporting the strongest interactions and which maximizes the total binding capacity of this region. Simultaneous binding of GPVI to the separate AGPOGP and of $\alpha_2\beta_1$ to the nearby GMOGER[15] appears structurally possible (Figure S13c). The frequent occurrence of GPVI-binding sites complicates their authentication *in vivo*. On the other hand, the vital importance for collagen-induced signaling of the relatively sparse VWF- and $\alpha_2\beta_1$-binding sites strongly supports the need for them to be accessible along with clustering GPVI-sites, at least to the extent they are present on the same helix segment.

Collagen binding induces clustering[11] or dimerization[27] of platelet-surface GPVI, or, provided that GPVI exists partially as dimers prior to ligand binding[46,47], it provokes a shift in monomer-dimer equilibrium[44] or perhaps a conformational switch to an activated dimer[33]. Each additional platelet receptor molecule binding to collagen substantially strengthens GPVI-mediated adhesion through avidity effects: 100- to 900-fold stronger binding per molecule as estimated from monomeric GPVI and dimeric GPVI-Fc_2 binding to collagen fibrils and peptides[10,48]. In this context, fibril architecture[34] with parallel receptor sites[35] and parallel orientation of conjugated D1-D2 domains are both required; the D1-D2 angle is rigid in our and all other structural data[27,33]. Dimers demonstrated to date crystallographically, including the recently reported nanobody-induced domain-swapped dimer (PDB-ID: 7NMU)[33] (Figure S14), could not themselves bind parallel sites in collagen, unless disrupted to allow the separate monomers the freedom to bind. The latter mechanism, however, is inconsistent with observations of increasing dimer levels upon collagen-binding[10,11,44]. Our study offers no indication of ectodomain dimerization suitable for cooperative collagen-binding beyond avidity effects arising from clustering, but, since steric hindrance dominates the regulation of GPVI-collagen binding, GPVI clustering may require longer collagen helices than our peptides[49]. It remains plausible that GPVI employs independent[50] dimerization and collagen-induced activation mechanisms based on avidity alone.

Potential antiplatelet agents modulating GPVI-mediated platelet adhesion could inhibit association with collagen, downstream signaling via FcRγ[3], phosphorylation[4], or receptor dimerization[50]; the latter pending elucidation of its exact mechanism in platelet activation. Despite the substantial structural similarities with the LAIR-1-binding sequence, our structural identification of the authentic collagen binding site and complementary collagen sequence may now support *de novo* development of specific agents targeting the collagen-GPVI interaction and minimizing drug cross-reactivity between receptors, e.g., GPVI and LAIR-1, which would be important for separating anti-thrombotic and immunomodulatory activities.

Acknowledgements

We thank the Swiss Light Source (SLS), Villigen (CH) and European Synchrotron Radiation Facility (ESRF), Grenoble (FR) for providing data collection facilities and the beamline scientists for their help with data collection. This work was supported in part by ECHO grant 700.58.006 from the Council of Chemical Sciences of the Netherlands Organization for Scientific Research to E.G.H. and Medical Research Council project grant G0400701 and Wellcome Trust grant 068724/Z/02/Z to R.W.F.

Authorship contribution

L.J.F., T.H.C.B., G.E.J., D.H., D.B., N.J., and M.V. performed research; L.J.F., T.H.C.B., G.E.J., R.W.F., and E.G.H. designed experiments; L.J.F., R.W.F., and E.G.H. wrote the manuscript; R.W.F., and E.G.H. supervised the research.

Disclosure of Conflicts of Interest

RWF is Chief Scientific Officer of CambCol Laboratories Ltd. The authors declare no other competing conflict of (financial) interest.

References

1. Clemetson KJ, Clemetson JM. Platelet collagen receptors. *Thromb. Haemost.* 2001;86:189–97.
2. Ruggeri ZM. Platelets in atherothrombosis. *Nat. Med.* 2002;8(11):1227–1234.
3. Tsuji M, Ezumi Y, Arai M, Takayama H. A novel association of Fc receptor γ-chain with glycoprotein VI and their co-expression as a collagen receptor in human platelets. *J. Biol. Chem.* 1997;272(38):23528–23531.
4. Gibbins JM, Okuma M, Farndale R, Barnes M, Watson SP. Glycoprotein VI is the collagen receptor in platelets which underlies tyrosine phosphorylation of the Fc receptor γ-chain. *FEBS Lett.* 1997;413(2):255–259.
5. Clemetson JM, Polgar J, Magnenat E, Wells TN, Clemetson KJ. The platelet collagen receptor glycoprotein VI is a member of the immunoglobulin superfamily closely related to FcalphaR and the natural killer receptors. *J. Biol. Chem.* 1999;274(41):29019–24.
6. Watson SP, Asazuma N, Atkinson B, et al. The Role of ITAM- and ITIM-coupled Receptors in Platelet Activation by Collagen. *Thromb. Haemost.* 2001;86(10):276–288.
7. Watson SP, Auger JM, McCarty OJT, Pearce a C. GPVI and integrin alphaIIb beta3 signaling in platelets. *J. Thromb. Haemost.* 2005;3(8):1752–1762.
8. Nieswandt B, Watson SP. Platelet-collagen interaction: Is GPVI the central receptor? *Blood*. 2003;102(2):449–461.
9. Watson SP, Herbert JMJ, Pollitt a. Y. GPVI and CLEC-2 in hemostasis and vascular integrity. *J. Thromb. Haemost.* 2010;8(7):1456–1467.
10. 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.
11. Loyau S, Dumont B, Ollivier V, et al. Platelet glycoprotein VI dimerization, an active process inducing receptor competence, is an indicator of platelet reactivity. *Arterioscler. Thromb. Vasc. Biol.* 2012;32(3):778–785.
12. Moroi M, Jung SM. Platelet glycoprotein VI: its structure and function. *Thromb. Res.* 2004;114(4):221–33.
13. Farndale RW, Sixma JJ, Barnes MJ, de Groot PG. The role of collagen in thrombosis and hemostasis. *J. Thromb. Haemost.* 2004;2(4):561–73.
14. Farndale RW, Lisman T, Bihan D, et al. Cell-collagen interactions: the use of peptide Toolkits to investigate collagen-receptor interactions. *Biochem. Soc. Trans.* 2008;36(Pt 2):241–50.
15. Raynal N, Hamaia SW, Siljander PR-M, et al. Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J. Biol. Chem.* 2006;281(7):3821–31.
16. Giudici C, Raynal N, Wiedemann H, et al. Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J. Biol. Chem.* 2008;283(28):19551–60.
17. Hohenester E, Sasaki T, Giudici C, Farndale RW, Bächinger HP. Structural basis of sequence-specific collagen recognition by SPARC. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(47):18273–7.
18. Konitsiotis AD, Raynal N, Bihan D, et al. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* 2008;283(11):6861–8.
19. Lisman T, Raynal N, Groeneveld D, et al. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood*. 2006;108(12):3753–6.

20. Brondijk THC, Bihan D, Farndale RW, Huizinga EG. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109(14):5253–5258.

21. Zhou L, Hinerman JM, Blaszczyk M, et al. Structural basis for collagen recognition by the immune receptor OSCAR. *Blood*. 2016;127(5):529–537.

22. 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.

23. Kehrel B, Wierwille S, Clemetson KJ, et al. Glycoprotein VI is a major collagen receptor for platelet activation: it recognizes the platelet-activating quaternary structure of collagen, whereas CD36, glycoprotein IIb/IIIa, and von Willebrand factor do not. *Blood*. 1998;91(2):491–9.

24. Smethurst PA, Onley DJ, Jarvis GE, et al. Structural basis for the platelet-collagen interaction: the smallest motif within collagen that recognizes and activates platelet Glycoprotein VI contains two glycine-proline-hydroxyproline triplets. *J. Biol. Chem.* 2007;282(2):1296–304.

25. Barrow AD, Palarasah Y, Bugatti M, et al. OSCAR Is a Receptor for Surfactant Protein D That Activates TNF-alpha Release from Human CCR2+ Inflammatory Monocytes. *J. Immunol.* 2015;194(7):3317–3326.

26. Meyaard L. The inhibitory collagen receptor LAIR-1 (CD305). *J Leukoc Biol*. 2008;83(4):799–803.

27. 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.

28. Smethurst PA, Joutsi-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.

29. Lecut C, Arocas V, Ulrichts 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.

30. 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.

31. 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.

32. Lebbink RJ, Raynal N, de Ruiter T, et al. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* 2009;28(4):202–10.

33. 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.

34. 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.

35. Herr AB, Farndale RW. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J. Biol. Chem.* 2009;284(30):19781–5.

36. 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.

37. Sampathkumar, P., Bonanno, J., Fiser, A., Patskovsky, Y., Zencheck, W., Nathenson, S.G., Almo SC. Crystal structure of a strand-swapped dimer of Mouse Leukocyte-associated immunoglobulin-like receptor 1 IG-like domain (Unpublished data). *Protein data bank*. PDB-ID:4ESK.

38. Moroi M, Jung SM, Okuma M, Shinmyozu K. A patient with platelets deficient in glycoprotein VI that lack both collagen-induced aggregation and adhesion. *J. Clin. Invest.* 1989;84(5):1440–5.

39. Arai M, Yamamoto N, Moroi M, et al. Platelets with 10% of the normal amount of glycoprotein VI have an impaired response to collagen that results in a mild bleeding tendency. *Br. J. Haematol.* 1995;89(1):124–130.

40. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. Structural basis of collagen recognition by integrin alpha2beta1. *Cell*. 2000;101(1):47–56.

41. Widmer C, Gebauer JM, Brunstein E, et al. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109(33):13243–7.

42. Cai H, Sasikumar P, Little G, et al. Identification of HSP47 Binding Site on Native Collagen and Its Implications for the Development of HSP47 Inhibitors. *Biomolecules*. 2021;11(7):983.

43. Farndale RW, Slatter D a, Siljander PR-M, Jarvis GE. Platelet receptor recognition and cross-talk in collagen-induced activation of platelets. *J. Thromb. Haemost.* 2007;5 Suppl 1:220–9.

44. Poulter NS, Pollitt AY, Owen DM, et al. Clustering of glycoprotein VI (GPVI) dimers upon adhesion to collagen as a mechanism to regulate GPVI signaling in platelets. *J. Thromb. Haemost.* 2017;15(3):549–564.

45. Chow WY, Forman CJ, Bihan D, et al. Proline provides site-specific flexibility for in vivo collagen. *Sci. Rep.* 2018;8(1):13809.

46. Clark JC, Neagoe RAI, Zuidscherwoude M, et al. Evidence that GPVI is Expressed as a Mixture of Monomers and Dimers, and that the D2 Domain is not Essential for GPVI Activation. *Thromb. Haemost.* 2021;

47. Herr AB. Direct evidence of a native GPVI dimer at the platelet surface. *J. Thromb. Haemost.* 2009;7(8):1344–6.

48. Miura Y, Takahashi T, Jung SM, Moroi M. Analysis of the interaction of platelet collagen receptor glycoprotein VI (GPVI) with collagen. A dimeric form of GPVI, but not the monomeric form, shows affinity to fibrous collagen. *J. Biol. Chem.* 2002;277(48):46197–204.

49. O’Connor MN, Smethurst PA, Davies LW, et al. Selective Blockade of Glycoprotein VI Clustering on Collagen Helices. *J. Biol. Chem.* 2006;281(44):33505–33510.

50. Jiang P, Loyau S, Tchitchinadze M, et al. Inhibition of glycoprotein VI clustering by collagen as a mechanism of inhibiting collagen-induced platelet responses: The example of losartan. *PLoS One*. 2015;10(6):1–20.

51. Persikov A V., Ramshaw JAM, Kirkpatrick A, Brodsky B. Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* 2002;316(2):385–394.

52. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016;44(W1):W344–W350.

Tables

Table 1. Primary sequences of collagen model peptides

Peptide	Sequence	T$_m$ (°C)
(GPO)$_5$	(GPO)$_5$-NH$_2$	28.2
(GPO)$_3$	(GPP)$_2$(GPO)$_3$(GPP)$_2$-NH$_2$	41.1
CRP	GCO(GPO)$_{10}$GCOG-NH$_2$	82.3
III30	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	47.4
III30-A3	GPC(GPP)$_5$GAAGLRGGAGPOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	39.0
III30-A5	GPC(GPP)$_5$GAOGARGGAGPOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	46.9
III30-A6	GPC(GPP)$_5$GAOGLAGGAGPOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	44.9*
III30-A8	GPC(GPP)$_5$GAOGLRGAAGPOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	44.8*
III30-A11	GPC(GPP)$_5$GAOGLRGGAGAOGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	41.3
III30-A12	GPC(GPP)$_5$GAOGLRGGAGPAGPEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	42.2
III30-A14	GPC(GPP)$_5$GAOGLRGGAGPOGAEGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	43.3
III30-A15	GPC(GPP)$_5$GAOGLRGGAGPOGPAGGKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	47.4
III30-A17	GPC(GPP)$_5$GAOGLRGGAGPOGPEGAKGAAGPOGPO(GPP)$_5$GPC-NH$_2$	44.8*
III30-A18	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGAGAAGPOGPO(GPP)$_5$GPC-NH$_2$	45.5
III30-A23	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGAOGPO(GPP)$_5$GPC-NH$_2$	45.0
III30-A24	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGPAGPO(GPP)$_5$GPC-NH$_2$	44.4
III30-A26	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGPOGAO(GPP)$_5$GPC-NH$_2$	47.2
III30-A27	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGPOGPA(GPP)$_5$GPC-NH$_2$	45.3
III30-A24-27	GPC(GPP)$_5$GAOGLRGGAGPOGPEGGKGAAGPAGPA(GPP)$_5$GPC-NH$_2$	40.8*

* T$_m$ was estimated by the method of Persikov et al.[51]

Table 2. $K_{D,app}$ binding constants measured for wild type and mutant GPVI-Fc binding to collagenous substrates in solid-phase plate binding assays

	CRP			Fibrillar Collagen I			Toolkit-peptide III-30		
Mutant	$K_{D,app}$ (nM)	95% confidence interval (nM)	Fold increase	$K_{D,app}$ (nM)	95% confidence interval (nM)	Fold increase	$K_{D,app}$ (nM)	95% confidence interval (nM)	Fold increase
WT	2.71	2.2 - 3.2 nM		30.1	26 - 34 nM		10.9	9.7 - 12.0 nM	
E40A	2.5×10^3	$2.1 - 2.8 \times 10^3$	913	25×10^3	$22.1 - 27.5 \times 10^3$	823	74×10^3	$57 - 91 \times 10^3$	6848
R38A	2.0×10^3	$1.7 - 2.3 \times 10^3$	748	18×10^3	$16.0 - 20.3 \times 10^3$	603	42×10^3	$38 - 45 \times 10^3$	3833
W76A	2.0×10^3	$1.8 - 2.2 \times 10^3$	744	43×10^3	$37.2 - 49.6 \times 10^3$	1440	21×10^3	$14 - 27 \times 10^3$	1905
D49A	599	454 - 744	221	9.2×10^3	$8.03 - 10.4 \times 10^3$	306	3.9×10^3	$3.55 - 4.19 \times 10^3$	357
R67A	238	187 - 289	88	10×10^3	$6.63 - 12.9 \times 10^3$	324	2.0×10^3	$1.90 - 2.16 \times 10^3$	187
Y47A	160	130 - 189	59	6.0×10^3	$4.43 - 7.56 \times 10^3$	199	10×10^3	$8.5 - 12 \times 10^3$	925
S69A	45.7	24 - 67	17	1.2×10^3	$0.93 - 1.49 \times 10^3$	40	6.5×10^3	$5.66 - 7.40 \times 10^3$	602
Q71A	19.9	11 - 29	7	72.2	54.8 - 89.5	2	9.15*	6.87 - 11.4	0.84
L36A	18.8	12 - 25	7	186	150 - 222	6	7.68*	5.77 - 9.58	0.70
Q82A	3.89	2.5 - 5.3	1.4	433	329 - 538	14	289	268 - 310	27

* Maximum III-30-binding of the Q71A and L36A mutants is 109% (104 - 114%) compared to wild type protein

Figure legends

Figure 1: Crystal structures of GPVI mutant ΔPAVS-ΔPAPYKN and GPVI-collagen peptide complexes. (A) Superimposition of the crystal structures of loop truncation mutant ΔPAVS-ΔPAPYKN (dark blue loops) and previously solved wild type GPVI ectodomain (PDB-ID: 2GI7; light blue loops)[27] showing significant conformational differences limited to the region around the truncated loops (red) in its D2-domain . The insets show an overlay of the wild type loops (red) and the mutant loops (blue) in sticks representation, together with the 2*Fo-Fc* electron density map of the mutant (blue, 1.2σ contour level). *See also Figures S1 & 2.* **(B)** Crystal structure of the GPVI-$(GPO)_5$ complex showing two GPVI-molecules in cartoon and surface representations that bind the collagen peptide with their D1-domains. Chains of the collagen triple helix are shown in pink (Leading; *L*), magenta (Middle; *M*), and orange (Trailing; *T*). Molecule A (blue) binds the *M*+*T* combination of chains, molecule B (salmon) binds the *L*+*M* chain combination. Also indicated (red) are the truncated D2-loops that are situated outside the region involved in collagen-binding. *See also Figure S3.* **(C)** Crystal structure of the GPVI-$(GPO)_3$ complex showing the same molecular arrangement as observed in crystals of the GPVI-$(GPO)_5$ complex, but with shifted collagen chain stagger, such that the GPVI-molecules bind the *L*+*M* and *T*+*L* combinations of chains, respectively. The two generic GPP-repeats at either end of the $(GPO)_3$-peptide are shown in white sticks. *See also Figure S5.* **(D)** Back-to-back D2-D2 interactions between two symmetry-related GPVI-$(GPO)_5$ complexes. The inter-helix angle of about 80° is incompatible with a parallel helix orientation as in collagen fibrils[34] and the back-to-back dimer is therefore not expected to be of functional relevance for *in vivo* binding to fibrillar collagen. *See also Figure S4.* **(E-F)** Schematic representation of the GPVI contact sites on the $(GPO)_5$ **(E)** and $(GPO)_3$ **(F)** peptides, also illustrating collagen chain stagger and GPO-triplets. Collagen residues that are located within 4.5 Å of the first (blue) and second (salmon) GPVI-molecule are highlighted and the corresponding sequence motifs are depicted below the figure. Three out of four sites are formed by a POGPOGPO-sequence; the fourth, the second site on $(GPO)_3$, overlaps the generic C-terminal $(GPP)_2$-extension of the peptide and is formed by the longer POGPOGPPGPP-sequence due to the collagen chain stagger of the *T*+*L* chain combination.

Figure 2: The GPVI-collagen binding interface. (A) Surface representation of the GPVI collagen binding site (magenta) with residues situated within 4.5 Å of collagen peptide $(GPO)_5$ shown in sticks representation and β-strand labeled in green. **(B)** Representation of conservation within the GPVI-family using ConSurf[52] coloring illustrates strict conservation of six central and more variability in five surrounding residues in the collagen binding site. **(C)** The GPVI-$(GPO)_5$ interface showing hydrogen bonding interactions and salt bridges (dashed lines) and intermolecular carbon-carbon distances within 4.5 Å (thin solid lines). $(GPO)_5$-residues in the leading (pink), middle (magenta), and trailing (orange) chain are shown in sticks representations and labeled by sequence number (gray). Also displayed is the $(GPO)_5$-surface, highlighting regions within 4.5 Å of GPVI by coloring according to the nearest chain. GPVI-residues in the interface are shown as main chain ribbons and side chain sticks (blue) and labeled by residue (black). **(D-E)** Detailed views of the hydrogen bonding interactions between $(GPO)_5$ and GPVI-residues Gln71 and Trp76 **(D)** and Arg38, Glu40, and Arg67 **(E)**. Hydrogen bonds and salt bridges are indicated by dashed lines.

Figure 3: Mutational analysis of GPVI-collagen binding. Binding of GPVI-mutants (colored series) to CRP **(A)**, fibrillar collagen I **(B)** and Toolkit-peptide III-30 **(C)** as measured in solid-state binding-assays. Data measured as A_{450} is

normalized to the binding of wild type GPVI (black series). All data points represent the mean ± SD of at least three independent experiments. Binding curves are fitted to the equation *Abs* = (B_{max} × *c*) / ($K_{D,app}$ + *c*) using non-linear regression in SigmaPlot, where *c* is the GPVI concentration (µM). *See also Figure S6.*

Figure 4: Binding pattern analysis of $(GPO)_5$, $(GPO)_3$, and Toolkit-peptide III-30. (A) Graph showing the relative contribution to GPVI-binding energy by collagen-peptide residues calculated using the FoldX SequenceDetail function as in *Supplemental Methods* [36]. Binding energy values were averaged for residues occupying equivalent positions in the four crystallographically resolved complexes, values for individual complexes are provided in *Figure S7a-d*; collagen residue numbering here conforms to *Figure 2c*. The inset table lists cumulative binding energies for different length sequence patterns and shown as percentage of the total energy. [§] *Exists as Pro instead of Hyp in the 2nd chain of the carboxy-terminal binding site of $(GPO)_3$.* **(B)** Surface rendering of $(GPO)_5$ (top panel) and schematic representation of the chain stagger (bottom panel) displaying the calculated binding energy contributions per residue from panel A on a blue-white color scale. **(C)** Binding of wild type GPVI-Fc (black) and control mutants R38A (red) and E40A (blue) to alanine-scanned variants of Toolkit peptide III-30 and control peptides III-30, CRP, $(GPP)_{10}$, and BSA. For easy reference, the amino acid sequence of III-30 is shown below, highlighting presence of an OGPOGP-motif (orange) and two AGPOGP-motifs (underlined). Data measured as A_{450} is normalized to binding of wild type GPVI to CRP. All data points represent the mean ± SD of at least three independent experiments. *See also Figure S11.* **(D)** Modeled binding of GPVI to the *9*-AGPOGP-*14* region of III-30. Relative to the $(GPO)_5$ and $(GPO)_3$ complexes, Ala instead of Hyp at position 9 disrupts interactions with GPVI Gln71 and Leu36 (left panel), whereas Glu instead of Hyp at position 15 just outside the motif enables formation of an additional hydrogen bond to Gln82 (right panel). The predicted differences in interactions are in line with the divergent effect of GPVI mutations Gln71Ala, Leu36Ala and Gln82Ala on binding to CRP and III-30 *(see Figures 3a & c)*.

Figure 5: GPVI-binding to fibrillar collagen requires large distance between receptor molecules. (A) Schematic representations of GPVI-binding (blue) to a triple helix (magenta) embedded in a collagen fibril (left panel) or immobilized onto a flat surface (right panel), illustrating that the circumferential angular range available for GPVI-binding is about 200° and 250°, respectively. Additional GPVI-molecules, covering about 160°, therefore need to bind at the same side of a helix or parallel helix. Triple helical packing displayed in the left panel was taken from the *in situ* structure of collagen I, which exhibits a parallel helix packing leading to surface exposure of identical binding sites on parallel helices[34]. **(B)** Perpendicular views of two GPVI-molecules (blue and salmon) binding to identical sites on parallel helices (magenta) with minimum 34 Å spacing, showing that its bent shape and oblique orientation with respect to the collagen helix preclude binding to closer spaced helices. **(C)** Perpendicular views of the two GPVI-molecules (blue and salmon) bound to $(GPO)_5$ as observed in our crystal structure. A short center-to-center distance of only 16 Å (left panel) is accompanied by a large circumferential coverage of about 310° (right panel), which is not feasible if the helix is incorporated in the fibril or immobilized on a surface. *See also Figure S9.* **(D)** Perpendicular views of putative arrangements of two GPVI-molecules (blue and salmon) bound to a triple helix in a 7_2 (top and bottom, magenta) or 10_3 (middle, orange) conformation. Their circumferential coverage angles are within the maximum value of 200° available for binding to fibrillar collagen; their center-to-center distance ranges from 32-40 Å, at least two-fold larger

than the distance observed in crystal structures. *See also Figure S10 and Supplemental Methods for details about modeling.*

Figure 1

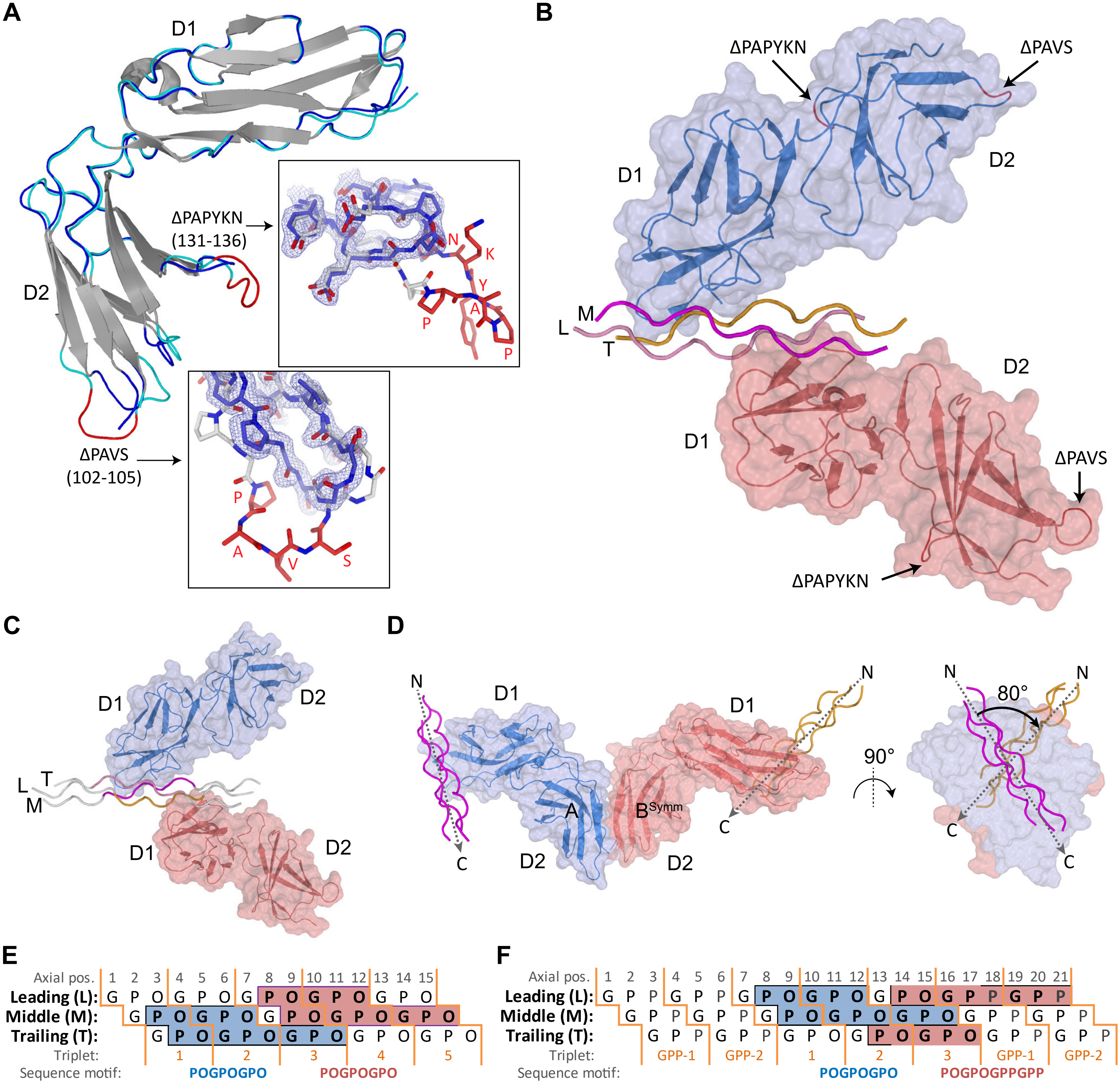

A
D1
D2
ΔPAPYKN
(131-136)
ΔPAVS
(102-105)
N
K
Y
A
P
P
P
A
V
S
B
ΔPAPYKN
ΔPAVS
D1
D2
M
L
T
D1
D2
ΔPAVS
ΔPAPYKN
C
D1
D2
T
L
M
D1
D2
D
N
D1
D1
N
C
D2
A
B^Symm
D2
C
90°
N
N
80°
C
C
E
Axial pos. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Leading (L): G P O G P O G P O G P O G P O
Middle (M): G P O G P O G P O G P O G P O
Trailing (T): G P O G P O G P O G P O G P O
Triplet: 1 2 3 4 5
Sequence motif: POGPOGPO POGPOGPO
F
Axial pos. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Leading (L): G P P G P P G P O G P O G P O G P P G P P
Middle (M): G P P G P P G P O G P O G P O G P P G P P
Trailing (T): G P P G P P G P O G P O G P O G P P G P P
Triplet: GPP-1 GPP-2 1 2 3 GPP-1 GPP-2
Sequence motif: POGPOGPO POGPOGPPGPP

Figure 2

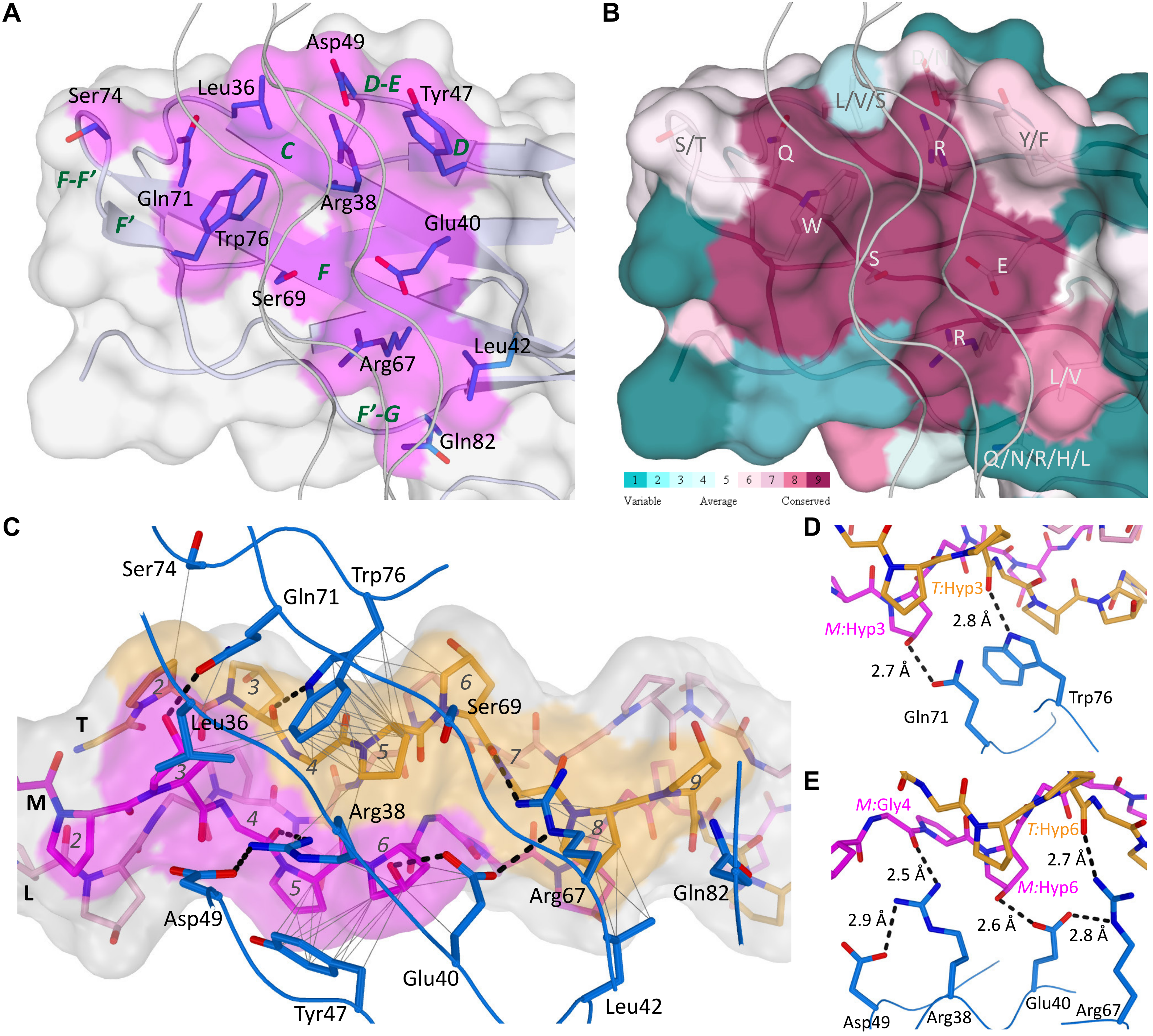

A
Asp49
D-E
Tyr47
Ser74
Leu36
C
D
F-F'
Gln71
Arg38
F'
Trp76
Glu40
F
Ser69
Leu42
Arg67
F'-G
Gln82
B
D/N
L/V/S
S/T
Q
R
Y/F
W
S
E
R
L/V
Q/N/R/H/L
1
2
3
4
5
6
7
8
9
Variable
Average
Conserved
C
Ser74
Trp76
Gln71
2
3
6
Ser69
T
Leu36
7
5
4
3
Arg38
9
M
4
8
2
6
5
L
Asp49
Arg67
Gln82
Glu40
Leu42
Tyr47
D
T:Hyp3
2.8 Å
M:Hyp3
2.7 Å
Trp76
Gln71
E
M:Gly4
T:Hyp6
2.7 Å
2.5 Å
M:Hyp6
2.9 Å
2.6 Å
2.8 Å
Asp49
Arg38
Glu40
Arg67

Figure 3

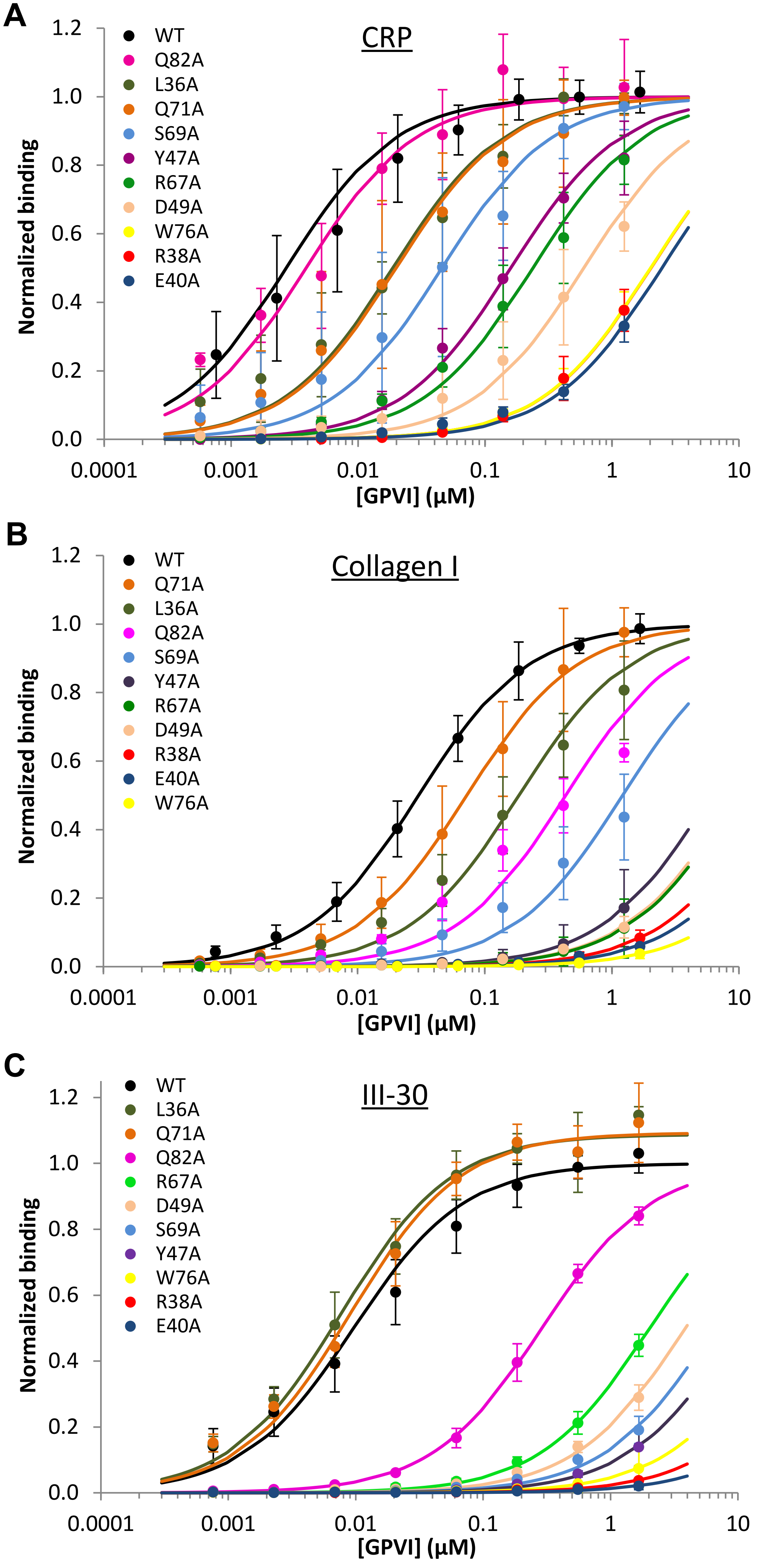
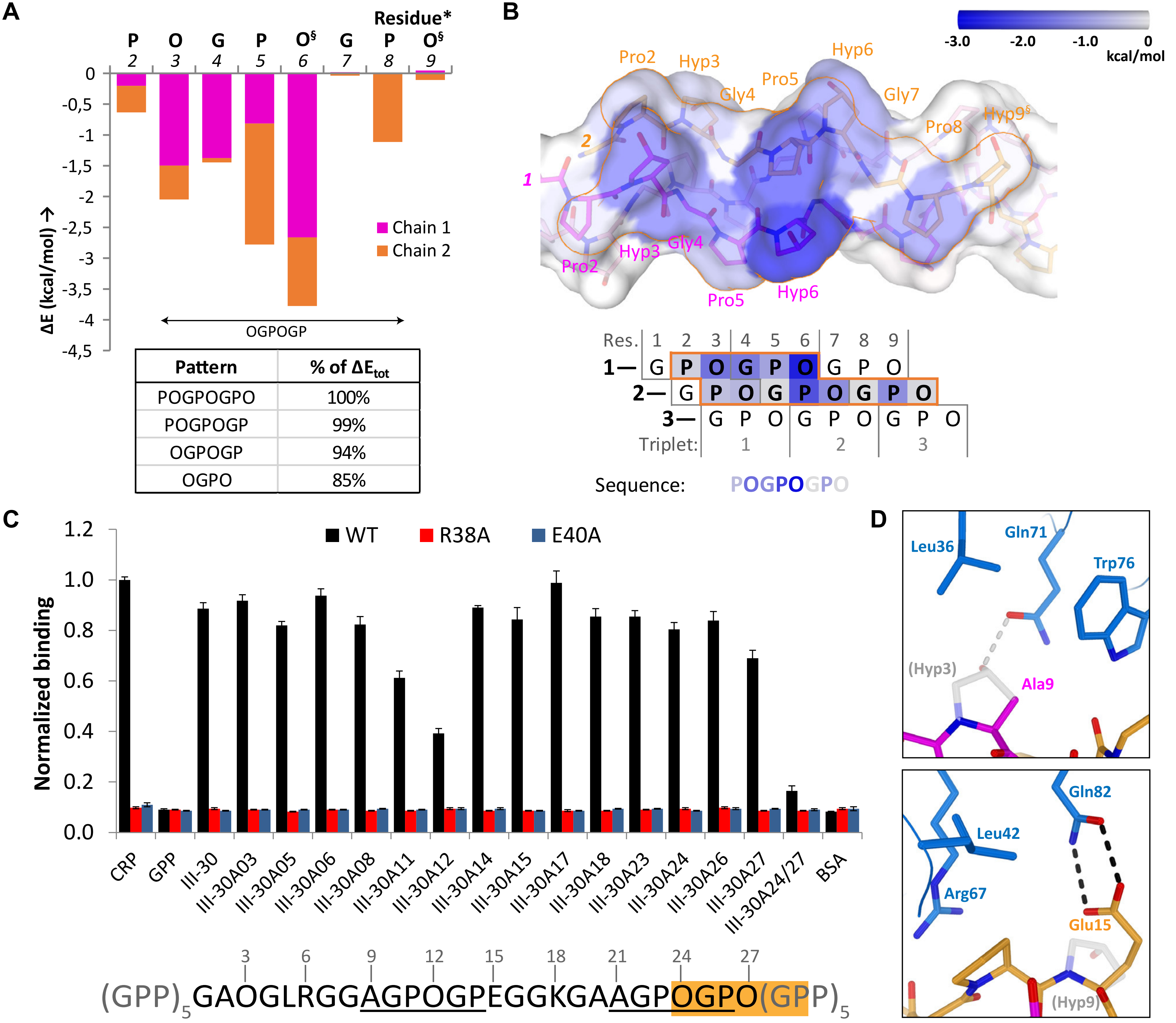
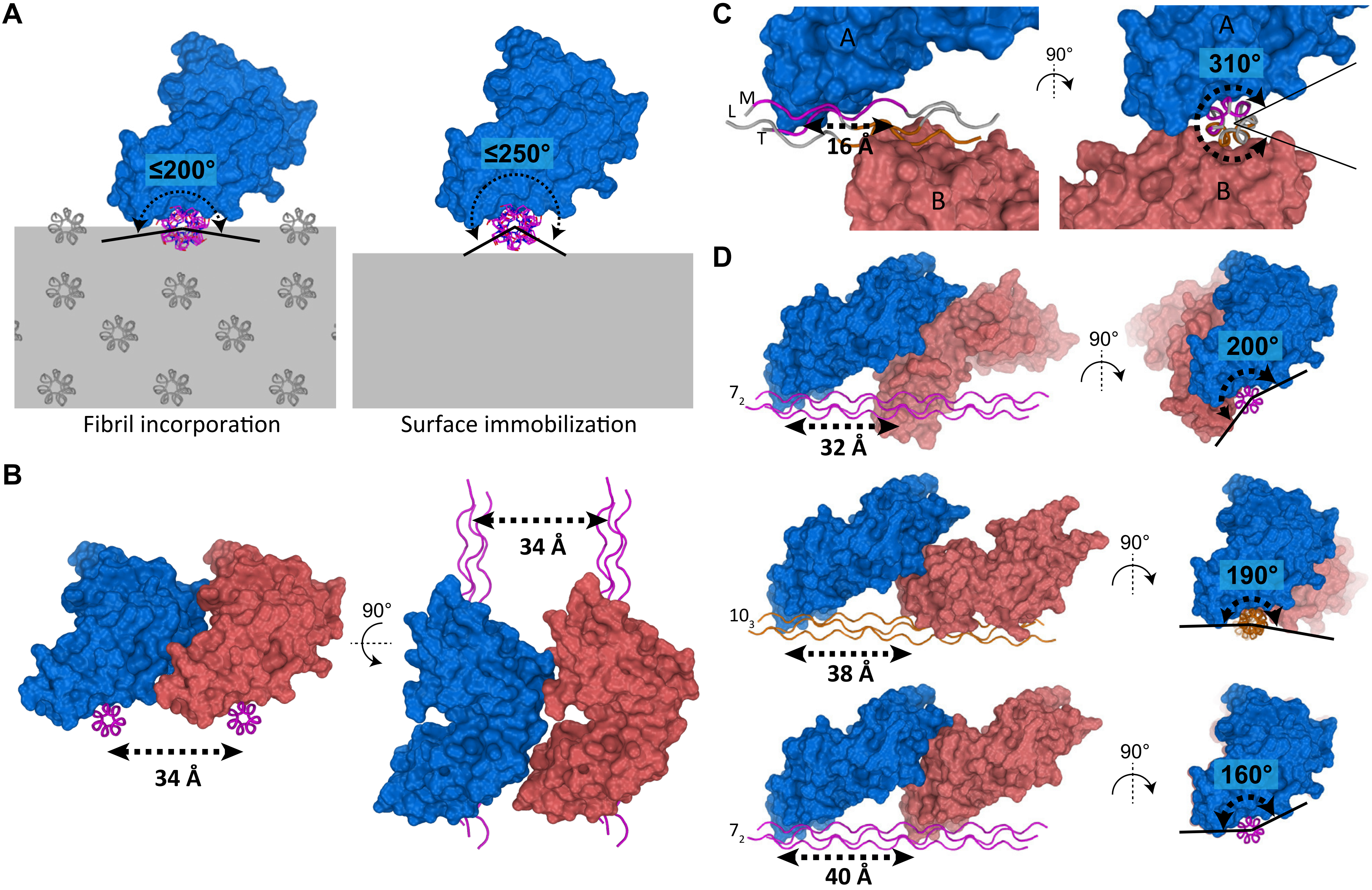

A
CRP
WT
Q82A
L36A
Q71A
S69A
Y47A
R67A
D49A
W76A
R38A
E40A
Normalized binding
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0001
0.001
0.01
0.1
1
10
[GPVI] (μM)
B
Collagen I
WT
Q71A
L36A
Q82A
S69A
Y47A
R67A
D49A
R38A
E40A
W76A
Normalized binding
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0001
0.001
0.01
0.1
1
10
[GPVI] (μM)
C
III-30
WT
L36A
Q71A
Q82A
R67A
D49A
S69A
Y47A
W76A
R38A
E40A
Normalized binding
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0001
0.001
0.01
0.1
1
10
[GPVI] (μM)

Figure 4



Pattern	% of ΔE_{tot}
POGPOGPO	100%
POGPOGP	99%
OGPOGP	94%
OGPO	85%

Figure 5


Fibril incorporation

Surface immobilization