Adverse effects of Alport syndrome-related Gly missense mutations on collagen type IV: Insights from molecular simulations and experiments

HHS Public Access

Author manuscript

Biomaterials. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Biomaterials. 2020 May ; 240: 119857. doi:10.1016/j.biomaterials.2020.119857.

Adverse effects of Alport syndrome-related Gly missense mutations on collagen type IV: Insights from molecular simulations and experiments

Jingjie Yeo[1,2,3,4,†], **Yimin Qiu**[1,5,†], **Gang Seob Jung**[2], **Yong-Wei Zhang**[3], **Markus J. Buehler**[2,*], **David L. Kaplan**[1,*]

[1]Department of Biomedical Engineering, Tufts University, 4 Colby Street, Medford, MA 02155, USA

[2]Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Institute of High Performance Computing, A*STAR, 1 Fusionopolis Way, Singapore 138632, Singapore

[4]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853, USA

[5]National Biopesticide Engineering Technology Research Center, Hubei Biopesticide Engineering Research Center, Hubei Academy of Agricultural Sciences, Biopesticide Branch of Hubei Innovation Centre of Agricultural Science and Technology, Wuhan, 430064, P. R. China

Abstract

Patients with Alport syndrome (AS) exhibit blood and elevated protein levels in their urine, inflamed kidneys, and many other abnormalities. AS is attributed to mutations in type IV collagen genes, particularly glycine missense mutations in the collagenous domain of *COL4A5* that disrupt common structural motifs in collagen from the repeat $(Gly–Xaa–Yaa)_n$ amino acid sequence. To characterize and elucidate the molecular mechanisms underlying how AS-related mutations perturb the structure and function of type IV collagen, experimental studies and molecular

[*]**Corresponding Authors**: Address: Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, mbuehler@mit.edu, Address: Department of Biomedical Engineering, Tufts University, 4 Colby Street, Medford, MA 02155, USA, david.kaplan@tufts.edu.
[†]These authors contributed equally.

Author Contributions

DLK and MJB conceptualized and supervised this research. JY, YQ, GSJ, and YWZ developed the methodologies. JY and GSJ developed the scripts and software. JY, YQ, and GSJ conducted the research, analyzed the data, and created the visuals. YQ performed the experiments. The manuscript was written by all the authors and they have given approval to the final version of the manuscript.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The authors declare that all the relevant data supporting the findings of this study are available within the paper and from the corresponding authors upon reasonable request.

simulations were integrated to investigate the structure, stability, protease sensitivity, and integrin binding affinity of collagen-like proteins containing amino acid sequences from the α5(IV) chain and AS-related Gly missense mutations. We show adverse effects where (i) three AS-related Gly missense mutations significantly reduced the structural stability of the collagen in terms of decreased melting temperatures and calorimetric enthalpies, in conjunction with a collective drop in the external work needed to unfold the peptides containing mutation sequences; (ii) due to local unwinding around the sites of mutations, these triple helical peptides were also degraded more rapidly by trypsin and chymotrypsin, as these enzymes could access the collagenous triple helix more easily and increase the number of contacts; (iii) the mutations further abolished the ability of the recombinant collagens to bind to integrins and greatly reduced the binding affinities between collagen and integrins, thus preventing cells from adhering to these mutants. Our unified experimental and computational approach provided underlying insights needed to guide potential therapies for AS that ameliorate the adverse effects from AS disease onset and progression.

Keywords

Alport syndrome; Collagen type IV; molecular dynamics simulation; experimental assays; integrin binding; enzyme digestion; structural stability

Introduction

Alport syndrome (AS) is a rare, debilitating genetic disorder [1]. Patients with AS manifest symptoms of blood and high levels of protein in their urine, inflamed kidneys and progressive loss of kidney functions that lead to end-stage renal disease (ESRD), loss of hearing, and visual abnormalities [2]. AS is attributed to mutations in type IV collagen genes of *COL4A3* and *COL4A4* that reside on chromosome 2, and *COL4A5* that resides on the X chromosome [3]. Mutations in the *COL4A5* gene account for an estimated 80% of patients diagnosed with AS [4], resulting in more than 300 mutations in the *COL4A5* gene with single amino acid substitutions accounting for an estimated 40% of these mutations [5]. These single amino acid substitutions are typically glycine missense mutations in the collagenous domain of *COL4A5* that disrupt the common structural motif: a repeating $(Gly–Xaa–Yaa)_n$ amino acid sequence that is found in all types of collagen [5]. Moreover, collagen type IV is the main constituent and the structural foundation for all basement membranes. Collagen type IV consists of heterotrimers that are coiled into triple helices, where the primary constituents are the collagenous α3(IV), α4(IV), and α5(IV) proteins in the sub-epithelial layer of the mature mammalian glomerular basement membranes (GBM) [6]. Unique to collagen type IV, the collagenous domains naturally contain approximately 21 to 26 glycine interruptions [7], while proline and hydroxyproline usually occupy the Xaa and Yaa positions, respectively. These interruptions are surmised to confer collagen type IV with molecular flexibility to form random networks [7,8]. Collagen type IV networks serve as molecular scaffolds that interact with other proteins in the extracellular matrix (ECM), such as laminins, perlecans, and proteoglycans. Critically, collagen type IV scaffolds in the basement membrane also interact with cells for biological processes such as cellular adhesion, migration, and differentiation. In particular, integrins on cellular surfaces bind

with collagens that possess native triple-helical conformation and show a high affinity to the GFOGER motif in the major collagens (O denotes hydroxyproline) [7].

Due to the intricate structure of collagen type IV networks, AS-related glycine missense mutations can cause the loss of multiple protein functions in the GBM [4]. These mutations primarily disrupt the normal folding of collagen α3(IV)/α4(IV)/α5(IV) triple helices and consequently lead to partial or the complete absence of these collagens within mature GBM [4]. In a previous study, the effects of single-residue glycine substitution mutations of varying clinical severity were examined with molecular dynamics simulations of short type IV tropocollagen molecules [9,10]. They observed local kinking located at the site of mutation in the tropocollagen structures after equilibration. Under tensile loads, there was also significant changes in tropocollagens' stress–strain responses and stiffnesses [9,10]. Interestingly, chemical chaperones are found to facilitate the re-folding of defective collagen strands into native triple helices [4]. Like the wild-type α5(IV) monomers, the mutant versions are stable within the cells, sharing similarities in degradation mechanisms and intracellular localization by the regulation of endoplasmic reticulum chaperones [11]. More crucially, some molecules from a class of chemical chaperones, known as osmolytes, show the capability of restoring the trimerization in α5(IV) with different types of glycine missense mutations *in vitro*, although it is still unclear whether this therapeutic effect can be reproduced *in vivo*. Moreover, even though the hierarchical structures of wild-type collagen IV were extensively characterized, the molecular mechanisms underlying how AS-related mutations perturb the structure and function of type IV collagen are not fully elucidated yet [4].

In this paper we provide molecular insight into these mechanisms by expressing recombinant collagen-like proteins, containing amino acid sequences from the α5(IV) chain and reported AS-related Gly missense mutations (Figure 1). The structure and stability of these recombinants were investigated with experimental studies and molecular dynamics simulations, showing that the three AS-related Gly missense mutations greatly reduced the structural stability of the collagen. The protease sensitivity was probed with trypsin and chymotrypsin digestion assays, showing that the mutations accelerated the rate of degradation. Computational simulations attributed this acceleration to the easier access of the enzymes to the collagen, as the mutations locally unwound the triple helix. The ability of these bacterial collagens to bind to integrin was examined by solid-state binding assays and their binding energies were characterized computationally. Mutations abolished binding, having greatly reduced binding affinities between collagen and integrin. Hence, by integrating experimental and computational approaches, we provide deeper molecular insights into the potential consequences of structural perturbations in AS symptoms. Our results will guide potential molecular therapies for AS, especially the molecular design of chemical chaperones for re-folding defective collagen type IV triple helices to ameliorate adverse effects from AS disease onset and progression.

Results

Design and expression of bacterial collagens containing type IV collagen sequence

In previous research, the effects of natural interruptions and Gly substitutions on the structure and stability of collagenous triple helices were studied with model peptides [12–16]. The structural consequences of inserting a 4-residue interruption sequence (AAVM) between two tandem CL domains were also characterized [17]. Here, we explored the influence of natural interruptions and pathological Gly substitutions on structural and functional properties of recombinant collagen-like proteins.

To avoid the low amounts of smaller triple-helical contaminants in purified protein, which is a known phenomenon for proteins with repetitive sequences produced in bacterial expression system [17], all mutants were constructed based on Scl2 collagen-like protein VCL from *Streptococcus pyogenes*. Protein VCL contains an N-terminal trimerization domain (V domain) and a core collagen-like domain (CL domain) $(Gly\text{-}Xaa\text{-}Yaa)_{79}$ [18,19]. Although prolyl 4-hydroxylase and hydroxyproline (Hyp) are absent in *Escherichia coli*, the recombinant collagen expressed in *E. coli* still possesses a triple helical structure with similar thermal stability to mammalian collagens (T_m ~ 36–37°C). VCL protein has no inherent biological activity. Introducing the GFPGER-containing sequence (human α1(I) chain residues 496–513) confers VCL-Int protein the ability to bind to the integrin α2 I-domain [20]. GFOGER was identified as one of the major integrin recognition sequences in type IV collagen, and ^{400}GFPGER^{405} was a predicted integrin binding site in α5 chain of type IV collagen [21]. On this basis, a 29-residue sequence from human α5(IV) chain (residues 386–414, GPPGAAVMGPPGPPGFPGERGQKGDEGPP) was inserted within the bacterial CL triple helix domain, thereby generating protein VCL-X_{29}. This sequence includes the potential integrin-binding motif GFPGER and a natural 4-residue interruption AAVM whose conformation was determined from NMR studies on a peptide model [22]. Additionally, the highly charged residue pair ^{408}KGD^{410} would play a role in stabilizing the triple helix [16], and repeated GPP triplets flanking the break ensure that the peptides adopt the triple-helical conformation. Constructs were then generated respectively with Gly→Glu, Gly→Val, and Gly→Asp mutations at the G400, G406, and G409 sites of VCL-X_{29}. These Gly missense mutations were commonly observed in Alport syndrome carriers. VCL-X_{29} derivatives with Gly substitutions were designated as G400E, G406V, and G409D. To investigate the consequences of the interruption, comparisons were made with a homologous protein VCL-X_{30}, generated based on VCL-X_{29} with the Gly introduced between GAA^{391} and ^{392}VM to recreate the Gly-X-Y pattern, thereby having the sequence GPPGAAGVMGPPGPPGFPGERGQKGDEGPP (Figure 1). The purity and identity of the proteins were confirmed by SDS-PAGE and mass spectrometry (data not shown). Utilizing these experimental designs, similar computational molecular models were generated for detailed analysis of the structures, stability, and binding functions as well, specified in the Materials and Methods section.

Structure and stability of recombinant collagens

The biophysical results for recombinant proteins constructed in this work were compared with a homologous recombinant collagen constructed previously, VCL-Int, that contained

six triplets from human collagen (residues 496–513 from the α1 chain of type I collagen, GARGERGFPGERGVQGPP) around the GFPGER integrin-binding site [20]. The conformation and thermal stability of these collagen-like proteins were examined by circular dichroism (CD) spectroscopy and differential scanning calorimetry (DSC). CD spectra of all constructs showed typical triple-helical features, similar to that of the VCL-Int control, with a positive peak near 220 nm and a large negative peak near 198 nm (Figure S1A). Mean residue ellipticity (MRE) at 220 nm was monitored as a function of temperature, and the first derivative of the transition was used to obtain the melting temperature, T_m. The CD melting curves for VCL-Int and the recombinant proteins each showed a single sharp thermal transition (Figure S1B, C), suggesting fully folded triple-helical structures. Introducing the sequence from the human α5 (IV) chain into CL led to a small but reproducible decrease in T_m compared with the control VCL-Int (T_m = 35.1 °C), in all cases of the natural Gly break or pathological Gly substitutions. The VCL-X_{29} and VCL-X_{30} proteins displayed comparatively similar thermal stabilities, whereas the introduction of Gly missense mutation resulted in a decrease in stability. T_m values for the three Gly substitution constructs ranged from 32.7 to 33.3°C, compared with 34.2°C for the parent protein VCL-X_{29}. Thermal transitions obtained by DSC exhibited the same trends (Figure S1D) with slightly higher T_m values than those measured by CD because of the faster heating rate. The DSC profile of VCL-Int showed a sharp transition near 37.1°C, with a calorimetric enthalpy of 3660 kJ/mol (Figure 2A, B). The introduction of the COL4A5 sequence also lowered the calorimetric enthalpy (ΔH_{cal}): 2566 kJ/mol for VCL-X_{29}, 2643 kJ/mol for VCL-X_{30}. The Gly substitution mutants had significantly lower values of ΔH_{cal}, ranging from 1799 to 1982 kJ/mol for G400E, G406V, and G409D (Figure 2B).

From fully atomistic MD simulations, the destabilized triple-helical structures of the mutant peptides were also characterized qualitatively with uniaxial tensile and rotational tests, as well as through replica exchange simulations after binding to integrin (see Materials and Methods). For each tensile test, we assumed that the crossover points of the two fitted linear functions corresponded to the critical strain where the triple helix was unfolded (Figure S4C). The estimated critical strain for each case was approximately in the range between 0.25 to 0.3 where hydrogen bonds ruptured (Figure S4C), as shown by a previous study [23]. The collective drop in the external work needed to unfold the mutant peptides was in good qualitative agreement with our measurements of the calorimetric enthalpies and melting points. Similarly, by untwisting the peptides, we also found that the mutant peptides must be unwound for structural relaxation and to achieve lower energetic minima (Figure 2C, Figure S5). The large amount of unwinding and decreases in energy needed to relax the mutant structures were in qualitative correlation with their decrease in calorimetric enthalpies and melting points, suggesting that the stability of the mutant peptides differed from the wild-type due to differing equilibrium helical structures. Finally, replica exchange simulations of the bound structures of collagen with integrin also showed extensive structural kinking and unfolding of the mutant peptides near the sites of mutation (Figure 2D), in contrast to the largely undisturbed triple helical structures of VCL-X_{29} and VCL-X_{30}. These large deformations were also observed in the greater spread of backbone dihedral angles in Ramachandran plots away from the canonical collagen triple helical angles (Figure S6), suggesting that the bound structures were unable to accommodate mutations in the collagen-

like peptides. These kinks and misfolds likely led to the subsequent increase in protease susceptibility and impaired binding functions as discussed in the following sections.

Protease sensitivity of collagens

The standard triple-helical conformation is known to be resistant to non-specific proteases. Therefore, trypsin digestion was performed on the recombinant constructs VCL-X_{29} and VCL-X_{30} to ascertain if the inserted human α5 (IV) sequences maintained a tight triple helix. Like the VCL-Int control, VCL-X_{29} and VCL-X_{30} remained resistant to trypsin after a 15-min trypsin treatment at 20°C (Figure 3A). Mutants with Gly substitutions within the inserted human sequence in VCL-X_{29} were digested with trypsin to assess structural perturbations to the native triple helix. In contrast to the native collagens, introducing the Gly to Glu/Val/Asp mutations resulted in dramatically enhanced susceptibility of recombinants to trypsin digestion at 20 °C after a relatively rapid 2-min digestion time. These conditions were typically used to test for native triple helix in animal collagens [19] (Figure 3A). Surprisingly, Gly substitutions at positions 406 and 409 were more susceptible to trypsin digestion than the mutant with Gly substitution at position 400 within the ^{400}GFPGER^{405} motif. G406V and G409D were completely degraded within 15 min (Figure 3A). Additionally, a more general enzyme chymotrypsin was applied to confirm the consequences of interruptions and Gly substitutions on structure of triple helix. Consistent with the trypsin digestion results, VCL-X_{29} and VCL-X_{30} were resistant to digestion, while G406V and G409D were more susceptible to chymotrypsin than G400E (Figure S2).

Molecular docking of bovine pancreatic trypsin to each collagen construct showed that the increased enzymatic susceptibility correlated with a large increase in the number of contacts between the enzyme and the mutated collagen peptides (Figure 3B). Interestingly, VCL-X_{29} had the lowest number of contacts with the enzyme (700 vs. 818 in VCL-X_{30}), possibly suggesting that the presence of the natural interruption in collagen type IV might be beneficial for their structural longevity due to lower accessibility to degrading enzymes. This observation reinforced our finding of much higher structural and thermal stability in VCL-X_{29} than mutant peptides. The number of contacts increased by more than 20% when the mutations were introduced, especially for G409D which rose by almost 41% in good correlation with experimental observations. The increased numbers of contacts might be a result of enhanced accessibility of the degrading enzymes as they docked onto the structural distortions that were localized around the mutated regions and unwound the triple helices (Figure 3B).

Integrin binding to recombinant constructs

Recombinant I-domain from the integrin α2 subunit was used as a probe in binding assays for characterizing collagen-like peptides and proteins *in vitro* [20,24,25]. Therefore, solid-state binding assays were performed to determine the binding affinity of recombinant proteins with Gly interruptions and Gly substitutions to integrin α2 I-domain. Consistent with previous studies, the original VCL protein without the GFPGER insertion showed no binding affinity to the integrin α2 I-domain, whereas the positive control VCL-Int, with the GARGERGFPGERGVQGPP insertion from human α1(I) chain, showed a strong and specific binding to the I-domain (Figure 4A) [25]. Introducing residues 386–414 from

human α5(IV), GPPGAAVMGPPGPPGFPGERGQKGDEGPP, into VCL endowed the protein VCL-X_{29} with integrin binding activity, even though the binding signals were slightly decreased compared to VCL-Int. Similar results were found for VCL-X_{30}. Gly to Glu substitution at position Gly^{400} within the essential ^{400}GFPGER^{405} binding site, G400E, abolished integrin binding activity, showing values similar to the VCL negative control (Figure 4A). Noticeably, replacing the Gly residues C-terminal to the ^{400}GFPGER^{405} sequence, G406V and G409D, showed essentially no integrin binding activity (Figure 4A).

To assess the energetics of the binding affinity, the binding energy of collagen complexed with integrin was rapidly sampled with the method of multiple-walker adaptive biasing force [26,27] (Figure 4B), where the collective variable sampled was the distance between the center-of-masses (COM) of each collagen triple-helix and the integrin α2 I-domain. The COM distances were normalized such that all energetic minima coincided at 1. The energy curves were strongly divergent as binding energies of VCL-X_{30} and VCL-X_{29} were in excess of 1000 kJ/mol above the normalized COM distance of 1.6, whereas G400E, G406V, and G409D greatly reduced binding energies by between 200 and 400 kJ/mol. This loss of binding affinities paralleled the disruptions of the triple-helical structures after introducing these Gly mutations (Figure 2D), suggesting that the protein structures strongly influenced the binding of type IV collagen to integrin α2(I) in conjunction with having specific integrin-binding domains.

Also, the human fibrosarcoma cell line HT1080 is known to express integrin α2β1 as its sole functional collagen-binding integrin receptor. We further examined the adhesion of HT1080 cells to the mutants to evaluate the cellular function of these constructs. Cell adhesion profiles were analyzed by live-cell imaging and cell counting. Results of *in vitro* cell adhesion studies agreed well with the integrin binding assay. Consistent with the previous report, HT1080 cells showed very low adhesion to VCL, as well as the negative control PBS, but showed a strong binding to wells coated with VCL-Int [20]. Insertion of GFPGER containing sequence from human α5(IV) conferred VCL-X_{29} adhesion capacity for HT1080 cells. Similar to VCL-X_{29}, VCL-X_{30} showed lower cell adhesion than that of VCL-Int. The proteins with the Gly substitutions in VCL-X_{29}, namely G400E, G406V, and G409D, showed almost no cell adhesion. This was similar to the VCL construct, thereby confirming poor binding (Figure 5).

Discussion

In contrast with the observation that fibrillar collagens have Gly as every third residue within the repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence, multiple interruptions are observed in the triple-helix domains of all non-fibrillar collagens. Despite the natural presence of these interruptions in native type IV collagen, more than 100 Gly substitution mutations have been defined in the X-linked α5(IV) chain leading to the genetic disorder, Alport Syndrome [4]. It is interesting to note that the natural interruptions and pathological Gly substitution mutations in type IV collagen present a seemingly conflicting view where pathological mutations are highly disruptive to structure and function of collagens. To explore the adverse effects, recombinant bacterial system was applied here for proteins with a natural break found in the α5(IV) chain of type IV collagen (residues 386–414), and a set of homologous

proteins that either restored the Gly interruption or introduced Gly substitutions. Structural and functional characterizations of these recombinant constructs elucidated the underlying molecular mechanisms and etiology of collagen diseases.

Analyses of non-fibrillar collagens showed that the 4-residue interruption (G4G) is the most common type of interruptions in type IV collagen, with a total of 28 G4G found in all chains [28]. Previous studies on model peptides provide an approach to assessing the effects of interruptions on triple helix structure and stability [29,30]. As supported by NMR studies, the 30-mer peptide containing a GAAVM break forms a stable trimeric structure, with the Val residues occupying the central position where Gly residues are expected and creating a small hydrophobic core in the middle of the triple helix [22,29]. However, the GAAVM interruption resulted in a decrease in the triple helix content, a drop of 10°C in thermal stability, and a loss of calorimetric enthalpy, compared to the homologous peptide with Gly as every third residue (GAAGVM) [29]. Additionally, peptide studies showed that interruptions from 1 to 9 residues in length could be successfully incorporated within a stable triple helical structure, respectively, but caused a substantial destabilization of the triple helix [12]. It is known that the short length of the model peptide would magnify the perturbation of the triple helix at the interruption site.

According to the alignment of protein sequences, the G4G break (GAAVM) in the α5 chain of type IV collagen is opposite to the G4G breaks respectively in the α3 (GSSRP) and α4 (GEACA) chains [21], conferring a structure similar to that seen in the homotrimers used in this work. Biophysical characterization of recombinant protein VCL-X4-CL, which was generated with introduction of a GAAVM break between two CL domains, showed that the interruption had little effect on the melting temperature [17]. As expected, both the homologous VCL proteins, VCL-X_{29} and VCL-X_{30}, studied here exhibited sharp thermal transitions and had similar decreases in T_m, indicating the negligible effect of interruption on the global stability of the protein. In contrast, pathological Gly substitutions at the 400, 406, and 409 positions perturbed the triple helix in comparison with the VCL-X_{29} protein, including lower thermal stability and reduced calorimetric enthalpy. Tensile and rotational tests with computational MD simulations also confirmed that the stability of the mutant peptides differed from the wild-type due to differing equilibrium helical structures, while Replica Exchange MD (REMD) simulations clearly showed extensive structural kinking and unfolding of the triple helices near the mutation sites. Similar results were observed for the collagen mimetic peptides with a similar sequence environment, one modeling a natural G5G interruption (POALO) and the other one mimicking a pathological Gly→Ala substitution (LOAPO). It was shown that the Gly→Ala substitution caused a large reduction in global protein stability and triple-helical content, indicating that Gly substitutions may have much larger destabilizing effects than natural interruptions, therefore resulting in pathological conditions [13]. High-frequency occurrence of highly charged R(K)GE(D) motifs in both fibrillar and nonfibrillar collagens might also play a key role in stabilizing the triple helices [31]. Introduction of a *Osteogenesis Imperfecta* (OI) pathological Gly→Ala/Ser substitution in the similar highly charged sequence context resulted in the reduction of thermal stability and the disruption of hydrogen bonding [16,32]. Consistent with previous studies, more destabilization of the triple helices, in terms of reduced melting temperature, was observed when replacing Gly409 by Arg within the charged KGD motif, in

comparison with the substitutions at Gly400 and Gly406 positions. Data from our MD simulations also showed qualitatively comparable results, where destabilized structures had a lower energetic barrier to stretching or unfolding.

G1G and G4G interruptions were also incorporated into straight triple helices with highly localized perturbations [17,33]. The resistance of VCL-X_{29} to trypsin and chymotrypsin suggested a tightly packed structure, in agreement with the protease resistance observed in recombinant collagen VCL-X4-CL [17]. In contrast to VCL-X_{29}, all Gly substitutions increased both trypsin and chymotrypsin susceptibilities, indicating that Gly replacements caused significant local unfolding of the triple helices even if the substituted residue was incorporated very deeply within the triple helix. This unfolding might enhance accessibility of digestive enzymes to the collagen, as demonstrated by our molecular docking simulations. These results were also consistent with previous observations that VCL(G-S)-X4-CL were susceptible to trypsin and thermolysin digestion [17]. VCL(G-S)-X4-CL was generated based on VCL-X4-CL with a Gly→Ser mutation three triplets N-terminal to the 4-residue interruption. These results suggested that replacing Gly by a larger residue leads to local disruptions in the triple helix in general, but the sensitivity to protease hydrolysis was modulated by the local sequence environment of the Gly, and its location relative to the natural interruptions.

In previous studies, recombinant Scl2.28 proteins with inserted integrin binding sequences (GFPGER, GFPGEN, and GLPGER) showed the abilities to bind to integrins *in vitro* and to mediate cell binding [24,34]. Here, we found that the GFPGER containing sequence from type IV collagen α5 chain (residues 400–405) also supported integrin binding and cell attachment. Due to similar binding activities in VCL-X_{29} and VCL-X_{30}, recombinant proteins VCL-Int_{GAAVM} and VCL-Int_{GAAGVM} were further generated by introducing the 4-residue interruption GAAVM and the reference sequence GAAGVM in front of Gly496 within the insertion (GARGERGFPGERGVQGPP) in VCL-Int, respectively. Comparable integrin and cell binding activities were observed in proteins VCL-Int_{GAAVM} and VCL-Int_{GAAGVM}, providing further evidence that the natural Gly interruptions would not affect the binding activity (Figure S7). These results also supported the suggestion that natural interruptions in the $(Gly\text{-}X\text{-}Y)_n$ sequence pattern in non-fibrillar collagens are likely to play a role in molecular flexibility, higher order assembly, and ligand binding [12]. Interestingly, the proteins with the insertion from human α1(I) chain (VCL-Int, VCL-Int_{GAAVM}, and VCL-Int_{GAAGVM}) exhibited higher binding affinity to integrin α2β1 than the proteins with the insertion from human α5(IV) chain (VCL-X_{29} and VCL-X_{30}) (Figure S7). This result was in agreement with the previous work that α2β1 prefers fibril forming collagens over basement membrane type IV collagen, opposite to the α1β1 preference [35,36].

Patients with AS experience the partial, or even complete, inability to form dense collagen networks in the basement membranes, thereby manifesting symptoms that include progressive loss of kidney functions, loss of hearing, and visual abnormalities [2]. From our results, we may infer that the inability to form collagen networks can be associated with the loss in collagen structure and function from the molecular to the cellular level. Protein mutations related to AS destabilizes the structural integrity of collagen as shown by the significant drop in the mutant collagen peptides' T_m, calorimetric enthalpy, and the work to

unfold, as well as displaying kinking and unfolding in regions with mutations. The loss of collagen's molecular structure is further exacerbated by the fact that these structural instabilities enhance the ease of collagen digestion by proteases, such as trypsin and chymotrypsin, and this is likely to be due to increased substrate-enzyme contacts at the unfolded regions in mutant collagen. Moreover, at the cellular level, mutations abolish the ability of collagen to bind to integrin due to diminished binding energy, directly causing cells to be unable to adhere. Taken together, we can now identify the multiscale cascade of events that inhibits the formation of dense collagen networks and clearly ties in with the disease states of patients with AS.

Conclusions

AS-related glycine missense mutations can cause the loss of multiple protein functions in human GBM by disrupting the normal folding of collagen α3(IV)/α4(IV)/α5(IV) triple helices. Consequently, this disruption lead to partial or complete absence of these collagens within mature GBM. Currently, there is a dearth of therapies available to ameliorate these adverse effects, although certain chemical chaperones may restore the trimerization in α5(IV). To provide the underlying insights needed to guide further development of potential therapies for AS, especially the molecular design of chemical chaperones for re-folding defective collagen type IV triple helices, we comprehensively characterized and elucidated the molecular mechanisms underlying how AS-related mutations perturbed the structure and function of type IV collagen.

We combined experimental studies and molecular simulations to investigate the structure, stability, protease sensitivity, and integrin-binding affinity of collagen-like proteins containing amino acid sequences from the α5(IV) chain and AS-related Gly missense mutations reported previously. From CD and DSC data, we showed that three AS-related Gly missense mutations significantly reduced the structural stability of the collagen, decreasing the melting temperatures by more than 3% and significantly reducing calorimetric enthalpies by more than 30%. In qualitative agreement, we found a collective drop in the external work needed to unfold the mutant peptides in computational simulations of uniaxial tensile and rotational unwinding of these mutants. These mutations also accelerated the rate of degradation of the collagen by trypsin and chymotrypsin, compared to the digestion resistance of VCL-X_{29} and VCL-X_{30}. The enzymes could easily access, and increase their number of contacts, with the collagenous triple helix, due to local unwinding around the mutated sites demonstrated by REMD and molecular docking simulations. In solid-state binding assays, mutations further abolished the ability of the recombinant collagens to bind to integrin. Binding affinities between mutant collagen and integrin were greatly reduced by between 200 and 400 kJ/mol, determined computationally with the method of multiple-walker adaptive biasing force. Abolishing integrin-binding abilities also prevented human fibrosarcoma HT1080 cells from adhesions to the mutants. Through our study, we highlight the fact that potential therapies being developed for AS will need to account for complex interactions between the adverse consequences on the molecular structure, stability, protease sensitivity, and integrin-binding affinity in mutant collagen type IV.

Materials and Methods

Design and cloning of constructs

DNA constructs were derived from the previously constructed plasmid pColdIII-VCL [20]. The VCL construct harbored the DNA sequence for the V trimerization domain and the $(Gly\text{-}Xaa\text{-}Yaa)_{79}$ CL triple-helix domain with 6-His tag at N-terminus for purification purpose, while its derivative VCL-Int was constructed previously by insertion of the integrin-binding region in the human α1(I) chain (^{496}GARGERGFPGERGVQGPP^{513}) [20]. The introduction of GAAVM and GAAGVM in front of Gly496 within the inserted sequence in VCL-Int were performed using Q5 Site-Directed Mutagenesis Kit, thus yielding proteins $VCL\text{-}Int_{GAAVM}$ and $VCL\text{-}Int_{GAAGVM}$, respectively. Oligonucleotides encoding the sequence ^{386}GPPGAAVMGPPGPPGFPGERGQKGDEGPP^{414} from the human collagen IV a5 chain were designed and synthesized with *Xma*I and *Apa*I sticky ends (Thermo Fisher Scientific, MA). Annealed dsDNA was cloned into pColdIII-VCL vector after the DNA sequences encoding triplet number 30 in the CL-domain, generating pColdIII-VCL-X_{29} (protein designated as VCL-X_{29}). Versions of pColdIII-VCL-X_{29} with Gly substitutions were constructed by site-directed mutagenesis at residues 400, 406 and 409 using Q5 Site-Directed Mutagenesis Kit (NEB), yielding plasmids VCL-X_{29}G400E, VCL-X_{29}G406V, and VCL-X_{29}G409D (proteins denoted as G400E, G406V and G409D), respectively. The insertion of Gly codon between the nucleotide sequences of GAA^{391} and ^{392}VM within pColdIII-VCL-X_{29} was performed using Q5 Site-Directed Mutagenesis Kit, creating plasmid VCL-X_{29}G (protein designated as VCL-X_{30}). All enzymes were purchased from NEB (New England Biolabs, MA). The recombinant plasmids were confirmed by DNA sequencing.

Recombinant expression and purification

All constructs were transformed and expressed in *E. coli* BL21 cells under the conditions described previously [25]. Cells were harvested by centrifugation after overnight induction and resuspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4). After cell disruption by sonication, the supernatant fraction was purified on an ÄKTA pure system (GE Healthcare) as described previously [19]. Briefly, after loaded with the supernatant, the pre-equilibrated Ni-NTA column was washed with 5 column volumes of binding buffer, 3 column volumes of binding buffer plus 50 mM imidazole and binding buffer plus 100 mM of imidazole, sequentially. The protein bound to the Ni-NTA resin was eluted by elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4), and the purity was confirmed by SDS-PAGE. The purified proteins were dialyzed against 1 × PBS buffer (10 × PBS, pH 7.4; Fisher Scientific). Protein concentrations were measured by UV-Vis spectrophotometer (Aviv Biomedical Inc., Lakewood, NJ) using an extinction coefficient at 280 nm of $\varepsilon = 9{,}970\ M^{-1}\ cm^{-1}$.

Circular dichroism

CD (Circular dichroism) spectra were obtained on an AVIV Model 420 CD spectrometer (AVIV Biomedical Inc.). Wavelength scans were collected at 0°C from 260 to 190 nm recording points at every 0.5 nm for 4 s using a bandwidth of 1 nm, averaging three scans for each sample. Temperature scans were monitored by measuring MRE (mean residue

ellipticity) at 220 nm from 0 to 70 °C with a 10-s averaging time and 1.5-nm bandwidth. Samples were equilibrated for 2 min at each temperature, and the temperature was increased at an average rate of 0.1°C/min. The melting temperature (T_m) is defined as the temperature at which the fraction-folded $F(T_m)$ is equal to 0.5 as described previously [37].

Differential scanning calorimetry

DSC (Differential Scanning Calorimetry) profiles of recombinant collagens were obtained on a NANO DSC II model 6100 (Calorimetry Sciences Corp, Lindon, UT). Each sample was dialyzed against PBS overnight before measurement. Dialysis buffer was collected and used as a reference for the corresponding sample. Samples were loaded into the cells at 0°C and heated at a rate of 1°C /min till 70°C.

Trypsin and chymotrypsin digestion assay

Purified protein samples (0.1 mg/ml in PBS) were treated with 0.01 mg/ml of trypsin (Sigma-Aldrich) or 5.0 μg/ml of chymotrypsin (Sigma-Aldrich) at 20°C for up to overnight. The reaction was terminated by adding phenylmethylsulfonyl fluoride (Sigma-Aldrich) to a final concentration of 1 mM, and the digests were further analyzed by SDS-PAGE.

Solid-state integrin binding assay

Binding of integrin α2 I-domain to the recombinant bacterial collagen proteins was carried out according to the method reported previously [38,39]. Here, 100 μl of protein substrate (20 μg/ml) was coated onto the high-binding 96-well plate (R&D Systems, Minneapolis, MN) per well for 1 h at room temperature. Wells were washed with 1 × PBS and subsequently blocked with 200 μl of 50 mg/mL BSA in 1 × PBS for 1 h, and washed four times with washing buffer (PBS containing 1 mg/ml BSA and 2 mM $MgCl_2$). Then 100 μl of 20 μg/ml recombinant GST-tagged integrin α2 I-domain in wash buffer (provided by Professor B. Brodsky, Professor S. Hamaia and Professor R. W. Farndale) was added and incubated at room temperature with immobilized collagen for 1 h. After washing for four times, 100 μl of anti-GST HRP antibody (1:10,000 dilution in washing buffer) was applied to wells and incubated for 1 h at room temperature. After neutralization with 3 M HCl, binding was assessed through colorimetric analysis using a TMB substrate kit (Pierce). Absorbance was measured at 450 nm with SpectraMax M2 microplate reader (Molecular Devices Corp., Sunnyvale, CA). VCL-Int protein was used as the positive control for all the binding assays. All assays were performed in triplicate.

Mammalian cell adhesion assay

Human fibrosarcoma HT1080 cells adhesion assays were performed as described [20]. High binding 96-well plates (R&D Systems) were coated with 60 μg/ml protein substrate at room temperature for 1 h. Blocking was performed by coating the wells with 200 μl of 50 mg/ml BSA solution in 1 × PBS at room temperature for 1 h. After washing, 100 μl of HT1080 cells (2×10^4 cells/ml) in DMEM (Invitrogen) were seeded onto wells and incubated for 1 h at 37°C (supplemented with 10% FBS and 2 mM $MgCl_2$ with 5% CO_2 supply). Non-adherent cells were removed by washing with DMEM (no FBS; 2mM $MgCl_2$ added) four times. For cell imaging, 100 μl of DMEM containing 0.5 μl of calcein-AM solution (Sigma-

Aldrich) was added onto each well and incubated at 37°C for 30 min. All wells were washed with to remove excess calcein-AM, then fluorescence images of cells were taken with 490-nm excitation and 515-nm emission filters on a Leica DFC340 FX camera. Quantitative analysis for the cell attachment was performed by ImageJ. All adhesion assays were performed in triplicate.

Molecular models of triple-helical collagen IV peptides

Homotrimeric collagen molecules with the sequence in Figure S1 (VCL-X_{30}) were generated with the Tcl/Tk-based Triple-Helical collagen Building Script (THeBuScr) [40]. Triplets of GPP amino acids were added at both ends of the peptides to avoid potential side effects from the proximity to the N- and C-terminals:

$$(GPP)_3GPPGAAGVMGPPGPPGFPGERGQKGDEGPP'GPP(GPP)_3$$

THeBuScr can generate well-structured triple-helical collagen molecules with well-defined, repeating triplets of $(Gly\text{-}Xaa\text{-}Yaa)_n$ motifs. However, THeBuScr cannot generate sequences with missing or mutated Gly residues. Thus, we manually replaced or removed the targeted amino acid by utilizing the modules in Matriarch [41]. Matriarch is a python-based open-source script, allowing mathematical structural designs of proteins with arbitrary sequences of amino acids. For manipulating the collagen structures, we implemented new functions to import PDB files and convert them into variables that Matriarch can interpret and then generate more realistic atomic coordinates. After importing all atomic coordinates from the collagen peptides generated by THeBuScr, we removed the glycine to obtain the structure of VCL-X_{29} (Figure S1) that has the natural interruption. We estimated the helical angles from the locations of nitrogen atoms in glycine. To conserve the triple-helical feature of these collagen peptides, we rotated the amino acids on the N-terminal end based on the helical angles after removing the glycine. The same process was applied to replace the glycine to other mutation amino acid at each position as illustrated in Figure S1 and Figure S3 to obtain G400E, G406V, and G409D. The helical structures were all conserved from the manipulations.

Stability of collagen IV peptides under tension

Experimentally, we assessed the stability of the triple-helical structure of the collagen by quantifying their calorimetric enthalpy. Computationally, we estimated the stability of collagen molecules from MD simulations by calculating the external work done for unfolding, using steered MD (SMD) to naturally unfold the triple-helical structures. Subsequently, we divided the corresponding stress-strain curves into two regions: under ε = 0.1 and over ε = 0.5. The stiffness of each region was fitted with a linear function (Figure S4C). We assumed that the intersection point of these two linear functions was the critical strain where the triple helix was fully unfolded. The estimated critical strains were around 0.25~0.3 where the hydrogen bonds ruptured, as shown by a previous study [23]. The accurate estimation of collagen stability is challenging. However, the trends of the external work to unfold the structures showed good qualitative agreement with the trends observed in the calorimetric enthalpies.

Stability of collagen type IV peptides under axial rotation

To obtain further insights into the energetics of unfolding the triple helices, we performed additional simulations to directly unfold the triple-helical structures through axial rotations. We prepared the same collagen strands in $20 \times 5 \times 5$ nm^3 explicit water boxes, which were neutralized with Na^+ or Cl^- ions (Figure S5A). Now, we have more space for lateral deformation compared to SMD simulations. We used the same relaxation process and boundary conditions described in the previous section. To ensure that the starting angles were the same, we only relaxed the structure in the x-direction while fully constraining any axial rotations. Then, we rotated the nitrogen atoms by 1° along the x-axis at each step to unfold the structures (Figure 2C). At each step, we recorded the force applied to the nitrogen atoms for 5 ps. During the simulations, we anchored the nitrogen atoms with a restraining force (10 kcal/mol·Å) to prevent collapse due to the rotation. We averaged the force during the final 2.5 ps and estimated the energy required to rotate by 1° based on the displacement. We note that these comparisons were only possible for models with natural interruptions because the helical angles were very different once we introduced the natural interruptions, while the mutations did not change the angles very much. Figure 2C shows the results where the minimum points of the helical angles varied a lot. We found that the rotation angles also correlated qualitatively with the calorimetric enthalpies while the rank of energies to unfold by 600° showed the same trends as their melting points. The structures at their minimum energy states are shown in Figure S5B. We found that the slopes from the plots of energy versus angle after the minimum energy were very similar, which indicated that the stability of the collagen strands corresponded to the helical angles at the equilibrium state. It is challenging to estimate accurate energies for unfolding the structures because the simulation does not allow the three molecules to separate. However, atomistic simulations suggested that the stability of collagen with the mutations differed due to varying equilibrium helical structures.

Molecular models of collagen type IV bound to integrin

Using Chimera [42], the integrin-binding domain of collagen IV (the protein sequence GFPGER) was structurally aligned to the same domain in the crystal structure of collagen I in complex with integrin α2 I-domain (PDB: 1DZI) [43], and then the collagen I peptide was removed. In VMD [44], this complex of collagen IV and integrin was solvated and neutralized with ions, having a minimum gap of 2 nm between periodic images of the complex to avoid spurious effects of self-interactions. Conformations of the complex were extensively sampled with replica exchange molecular dynamics with solute tempering (REST2) [45–47], performed using the CHARMM22 forcefield [48] and NAMD 2.10 [49]. For each system, 16 replicas were simulated in parallel and their temperatures were distributed over the range of 300 to 400 K. Exchanges were attempted every 2 ps and exchange acceptance was based on the Metropolis criterion. This scheme led to an exchange acceptance ratio of more than 30%, implying sufficient efficiency in conformational sampling. Each simulation's timestep was 2 fs and each replica was simulated for 20 ns. The temperature was controlled with the Langevin thermostat and pressure was exerted through the Nosé-Hoover Langevin piston. Electrostatic interactions were calculated through the Particle Mesh Ewald method. After REST2 simulations, clustering using the GROMOS method was performed in GROMACS for the lowest temperature replica in each system to

determine the collagen-integrin complex conformations that occurred the most frequently. Ramachandran plots of the collagen peptides were generated in VMD [44]. The binding energy of collagen in complex with integrin was rapidly sampled with 16 replicas using the method of multiple-walker adaptive biasing force in NAMD 2.12 [26,27] with the bound Mg ion removed so that only the binding between collagen and integrin was determined. The pertinent collective variable for sampling was the distance between the center-of-masses of each collagen triple-helix and the integrin α2 I-domain.

Molecular docking of trypsin to collagen type IV peptides

With Haddock [50], the crystal structure of native bovine pancreatic trypsin (PDB: 1S0Q) was docked to the structures of collagen type IV that occurred the most frequently after clustering while the integrin was removed. Protein-protein contacts were calculated with MDTraj [51] where contacts were defined as heavy atoms that were within a distance of 1.0 nm between trypsin and collagen.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The authors acknowledge support from the US Department of Defense, Office of Naval Research (N00014-16-1-233) and the National Institutes of Health (U01EB014976). Computational simulations were performed on the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1053575, the MIT Engaging Cluster, and Singapore's A*STAR Computational Resource Centre and National Supercomputing Centre.

References

[1]. Hurst AF, Hereditary familial congenital haemorrhagic nephritis, Br. Med. J 1 (1927) 643. doi:10.1136/bmj.1.3456.643-a.

[2]. Kashtan CE, Alport syndrome and thin glomerular basement membrane disease., J. Am. Soc. Nephrol 9 (1998) 1736 LP–1750. http://jasn.asnjournals.org/content/9/9/1736.abstract. [PubMed: 9727383]

[3]. Kashtan CE, Ding J, Garosi G, Heidet L, Massella L, Nakanishi K, Nozu K, Renieri A, Rheault M, Wang F, Gross O, Alport syndrome: a unified classification of genetic disorders of collagen IV α345: a position paper of the Alport Syndrome Classification Working Group, Kidney Int. 93 (2018) 1045–1051. doi:10.1016/j.kint.2017.12.018. [PubMed: 29551517]

[4]. Cosgrove D, Liu S, Collagen IV diseases: A focus on the glomerular basement membrane in Alport syndrome, Matrix Biol. 57–58 (2017) 45–54. doi:10.1016/j.matbio.2016.08.005. [PubMed: 27576055]

[5]. Gubler MC, Inherited diseases of the glomerular basement membrane, Nat. Clin. Pract. Nephrol 4 (2008) 24–37. doi:10.1038/ncpneph0671. [PubMed: 18094725]

[6]. Abrahamson DR, Hudson BG, Stroganova L, Borza D-B, St. John PL, Cellular Origins of Type IV Collagen Networks in Developing Glomeruli, J. Am. Soc. Nephrol 20 (2009) 1471 LP–1479. http://jasn.asnjournals.org/content/20/7/1471.abstract. [PubMed: 19423686]

[7]. Khoshnoodi J, Pedchenko V, Hudson B, Mammalian Collagen IV, Microsc. Res. Tech 71 (2008) 357–370. doi:10.1002/jemt.20564. [PubMed: 18219669]

[8]. Yurchenco PD, Ruben GC, Basement membrane structure in situ: evidence for lateral associations in the type IV collagen network, J. Cell Biol 105 (1987) 2559–2568. doi:10.1083/jcb.105.6.2559. [PubMed: 3693393]

[9]. Srinivasan M, Uzel SGM, Gautieri A, Keten S, Buehler MJ, Alport syndrome mutations in type IV tropocollagen alter molecular structure and nanomechanical properties, J. Struct. Biol 168 (2009) 503–510. doi:10.1016/j.jsb.2009.08.015. [PubMed: 19729067]

[10]. Srinivasan M, Uzel SGM, Gautieri A, Keten S, Buehler MJ, Linking genetics and mechanics in structural protein materials: a case study of an Alport syndrome mutation in tropocollagen, Math. Mech. Solids 15 (2010) 755–770. doi:10.1177/1081286510374550.

[11]. Omachi K, Kamura M, Teramoto K, Kojima H, Yokota T, Kaseda S, Kuwazuru J, Fukuda R, Koyama K, Matsuyama S, Motomura K, Shuto T, Suico MA, Kai H, A Split-Luciferase-Based Trimer Formation Assay as a High-throughput Screening Platform for Therapeutics in Alport Syndrome, Cell Chem. Biol 25 (2018) 634–643.e4. doi:10.1016/j.chembiol.2018.02.003. [PubMed: 29526710]

[12]. Hwang ES, Thiagarajan G, Parmar AS, Brodsky B, Interruptions in the collagen repeating tripeptide pattern can promote supramolecular association, Protein Sci. 19 (2010) 1053–1064. doi:10.1002/pro.383. [PubMed: 20340134]

[13]. Sun X, Chai Y, Wang Q, Liu H, Wang S, Xiao J, A Natural Interruption Displays Higher Global Stability and Local Conformational Flexibility than a Similar Gly Mutation Sequence in Collagen Mimic Peptides, Biochemistry. 54 (2015) 6106–6113. doi:10.1021/acs.biochem.5b00747. [PubMed: 26352622]

[14]. Xiao J, Sun X, Madhan B, Brodsky B, Baum J, NMR Studies Demonstrate a Unique AAB Composition and Chain Register for a Heterotrimeric Type IV Collagen Model Peptide Containing a Natural Interruption Site, J. Biol. Chem. 290 (2015) 24201–24209. doi:10.1074/jbc.M115.654871. [PubMed: 26209635]

[15]. Xiao J, Yang Z, Sun X, Addabbo R, Baum J, Local amino acid sequence patterns dominate the heterogeneous phenotype for the collagen connective tissue disease Osteogenesis Imperfecta resulting from Gly mutations, J. Struct. Biol 192 (2015) 127–137. doi:10.1016/j.jsb.2015.05.002. [PubMed: 25980613]

[16]. Sun X, Liu S, Yu W, Wang S, Xiao J, CD and NMR investigation of collagen peptides mimicking a pathological Gly–Ser mutation and a natural interruption in a similar highly charged sequence context, Protein Sci. 25 (2016) 383–392. doi:10.1002/pro.2828. [PubMed: 26457583]

[17]. Hwang ES, Brodsky B, Folding Delay and Structural Perturbations Caused by Type IV Collagen Natural Interruptions and Nearby Gly Missense Mutations, J. Biol. Chem. 287 (2012) 4368–4375. doi:10.1074/jbc.M111.269084. [PubMed: 22179614]

[18]. Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JAM, Inouye M, Brodsky B, Self-association of streptococcus pyogenes collagen-like constructs into higher order structures, Protein Sci. 18 (2009) 1241–1251. doi:10.1002/pro.134. [PubMed: 19472339]

[19]. Chhum P, Yu H, An B, Doyon BR, Lin Y-S, Brodsky B, Consequences of Glycine Mutations in the Fibronectin-binding Sequence of Collagen, J. Biol. Chem. 291 (2016) 27073–27086. doi:10.1074/jbc.M116.753566. [PubMed: 27799304]

[20]. Yigit S, Yu H, An B, Hamaia S, Farndale RW, Kaplan DL, Lin Y-S, Brodsky B, Mapping the Effect of Gly Mutations in Collagen on α2β1 Integrin Binding., J. Biol. Chem 291 (2016) 19196–207. doi:10.1074/jbc.M116.726182. [PubMed: 27432884]

[21]. Des Parkin J, San Antonio JD, Pedchenko V, Hudson B, Jensen ST, Savige J, Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes, Hum. Mutat 32 (2011) 127–143. doi:10.1002/humu.21401. [PubMed: 21280145]

[22]. Li Y, Brodsky B, Baum J, NMR Shows Hydrophobic Interactions Replace Glycine Packing in the Triple Helix at a Natural Break in the (Gly-X-Y)n Repeat, J. Biol. Chem. 282 (2007) 22699–22706. doi:10.1074/jbc.M702910200. [PubMed: 17550894]

[23]. Gautieri A, Buehler MJ, Redaelli A, Deformation rate controls elasticity and unfolding pathway of single tropocollagen molecules, J. Mech. Behav. Biomed. Mater 2 (2009) 130–137. doi:10.1016/j.jmbbm.2008.03.001. [PubMed: 19627816]

[24]. Seo N, Russell BH, Rivera JJ, Liang X, Xu X, Afshar-Kharghan V, Höök M, An Engineered α1 Integrin-binding Collagenous Sequence, J. Biol. Chem. 285 (2010) 31046–31054. doi:10.1074/jbc.M110.151357. [PubMed: 20675378]

[25]. Qiu Y, Mekkat A, Yu H, Yigit S, Hamaia S, Farndale RW, Kaplan DL, Lin Y-S, Brodsky B, Collagen Gly missense mutations: Effect of residue identity on collagen structure and integrin binding, J. Struct. Biol 203 (2018) 255–262. doi:10.1016/j.jsb.2018.05.003. [PubMed: 29758270]

[26]. Darve E, Rodríguez-Gómez D, Pohorille A, Adaptive biasing force method for scalar and vector free energy calculations, J. Chem. Phys 128 (2008) 144120. doi:10.1063/1.2829861. [PubMed: 18412436]

[27]. Hénin J, Fiorin G, Chipot C, Klein ML, Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables, J. Chem. Theory Comput 6 (2010) 35–47. doi:10.1021/ct9004432. [PubMed: 26614317]

[28]. Thiagarajan G, Li Y, Mohs A, Strafaci C, Popiel M, Baum J, Brodsky B, Common Interruptions in the Repeating Tripeptide Sequence of Non-fibrillar Collagens: Sequence Analysis and Structural Studies on Triple-helix Peptide Models, J. Mol. Biol 376 (2008) 736–748. doi:10.1016/j.jmb.2007.11.075. [PubMed: 18187152]

[29]. Mohs A, Popiel M, Li Y, Baum J, Brodsky B, Conformational Features of a Natural Break in the Type IV Collagen Gly-X-Y Repeat, J. Biol. Chem. 281 (2006) 17197–17202. doi:10.1074/jbc.M601763200. [PubMed: 16613845]

[30]. Bella J, Eaton M, Brodsky B, Berman HM, Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution, Science (80-.). 266 (1994) 75 LP–81. doi:10.1126/science.7695699.

[31]. V Persikov A, Ramshaw JAM, Kirkpatrick A, Brodsky B, Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability, Biochemistry. 44 (2005) 1414–1422. doi:10.1021/bi048216r. [PubMed: 15683226]

[32]. Xiao J, Cheng H, Silva T, Baum J, Brodsky B, Osteogenesis Imperfecta Missense Mutations in Collagen: Structural Consequences of a Glycine to Alanine Replacement at a Highly Charged Site, Biochemistry. 50 (2011) 10771–10780. doi:10.1021/bi201476a. [PubMed: 22054507]

[33]. Brodsky B, Baum J, Modelling collagen diseases, Nature. 453 (2008) 998 10.1038/453998a. [PubMed: 18563144]

[34]. An B, Abbonante V, Yigit S, Balduini A, Kaplan DL, Brodsky B, Definition of the Native and Denatured Type II Collagen Binding Site for Fibronectin Using a Recombinant Collagen System, J. Biol. Chem. 289 (2014) 4941–4951. doi:10.1074/jbc.M113.530808. [PubMed: 24375478]

[35]. White DJ, Puranen S, Johnson MS, Heino J, The collagen receptor subfamily of the integrins, Int. J. Biochem. Cell Biol 36 (2004) 1405–1410. doi:10.1016/j.biocel.2003.08.016. [PubMed: 15147720]

[36]. KERN A, EBLE J, GOLBIK R, KÜHN K, Interaction of type IV collagen with the isolated integrins α1β1 and α2β1, Eur. J. Biochem 215 (1993) 151–159. doi:10.1111/j.1432-1033.1993.tb18017.x. [PubMed: 8344274]

[37]. Bryan MA, Cheng H, Brodsky B, Sequence environment of mutation affects stability and folding in collagen model peptides of osteogenesis imperfecta, Pept. Sci 96 (2011) 4–13. doi:10.1002/bip.21432.

[38]. Knight CG, Morton LF, Peachey AR, Tuckwell DS, Farndale RW, Barnes MJ, The Collagen-binding A-domains of Integrins α1β1 and α2β1Recognize the Same Specific Amino Acid Sequence, GFOGER, in Native (Triple-helical) Collagens, J. Biol. Chem. 275 (2000) 35–40. doi:10.1074/jbc.275.1.35. [PubMed: 10617582]

[39]. Tuckwell D, Calderwood DA, Green LJ, Humphries MJ, Integrin alpha 2 I-domain is a binding site for collagens, J. Cell Sci 108 (1995) 1629 LP–1637. http://jcs.biologists.org/content/108/4/1629.abstract. [PubMed: 7615681]

[40]. Rainey JK, Goh MC, An interactive triple-helical collagen builder, Bioinformatics. 20 (2004) 2458–2459. doi:10.1093/bioinformatics/bth247. [PubMed: 15073022]

[41]. Giesa T, Jagadeesan R, Spivak DI, Buehler MJ, Matriarch: A Python Library for Materials Architecture, ACS Biomater. Sci. Eng 1 (2015) 1009–1015. doi:10.1021/acsbiomaterials.5b00251. [PubMed: 27570830]

[42]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE, UCSF Chimera—A visualization system for exploratory research and analysis, J. Comput. Chem 25 (2004) 1605–1612. doi:10.1002/jcc.20084. [PubMed: 15264254]

[43]. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC, Structural Basis of Collagen Recognition by Integrin α2β1, Cell. 101 (2000) 47–56. doi:10.1016/S0092-8674(00)80622-4. [PubMed: 10778855]

[44]. Humphrey W, Dalke A, Schulten K, VMD: Visual molecular dynamics, J. Mol. Graph 14 (1996) 33–38. doi:10.1016/0263-7855(96)00018-5. [PubMed: 8744570]

[45]. Liu P, Kim B, Friesner RA, Berne BJ, Replica exchange with solute tempering: A method for sampling biological systems in explicit water, Proc. Natl. Acad. Sci. U. S. A 102 (2005) 13749–13754. doi:10.1073/pnas.0506346102. [PubMed: 16172406]

[46]. Wang L, Friesner RA, Berne BJ, Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2), J. Phys. Chem. B 115 (2011) 9431–9438. doi:10.1021/jp204407d. [PubMed: 21714551]

[47]. Jo S, Jiang W, A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations, Comput. Phys. Commun 197 (2015) 304–311. doi:10.1016/j.cpc.2015.08.030.

[48]. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, J. Phys. Chem. B 102 (1998) 3586–3616. doi:10.1021/jp973084f. [PubMed: 24889800]

[49]. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K, Scalable molecular dynamics with NAMD, J. Comput. Chem 26 (2005) 1781–1802. doi:10.1002/jcc.20289. [PubMed: 16222654]

[50]. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin AMJJ, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes, J. Mol. Biol 428 (2016) 720–725. doi:10.1016/j.jmb.2015.09.014. [PubMed: 26410586]

[51]. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang L-P, Lane TJ, Pande VS, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories, Biophys. J 109 (2015) 1528–1532. doi:10.1016/j.bpj.2015.08.015. [PubMed: 26488642]

Constructs	Description of construct	Inserted amino acid sequence within VCL
VCL-Int	Wild type containing integrin binding sequence from COL1A1	GARGERGFPGERGVQGPP
VCL-X_{29}	Wild-type sequence from COL4A5 (386 - 414)	GPPGAA- VMGPPGPPGFPGERGQKGDEGPP
VCL-X_{30}	Constructed based on VCL-X_{29}	GPPGAAGVMGPPGPPGFPGERGQKGDEGPP
G400E	Gly missense mutation within sequence from COL4A5 in Alport Syndrome carriers based on VCL-X_{29}	GPPGAA- VMGPPGPPEFPGERGQKGDEGPP
G406V		GPPGAA- VMGPPGPPGFPGERVQKGDEGPP
G409D		GPPGAA- VMGPPGPPGFPGERGQKDDEGPP

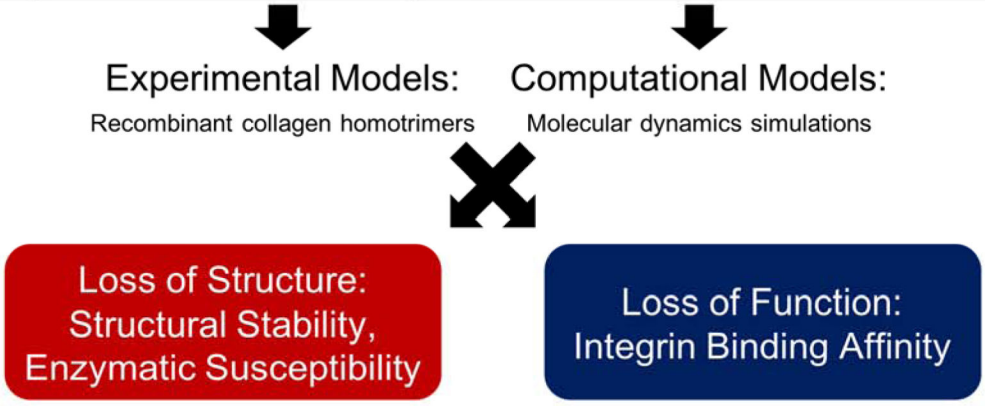


Figure 1.

Amino acid sequences for the reference, wild-type, and mutant collagen homotrimers for creating experimental recombinant proteins and computational molecular dynamics models. Sequences in green denote regions with natural glycine interruptions in collagen type IV. Sequences in blue denote integrin binding regions. Underlines denote mutated residues. The metrics of structural stability, enzymatic resistance, and integrin binding affinity were then investigated.

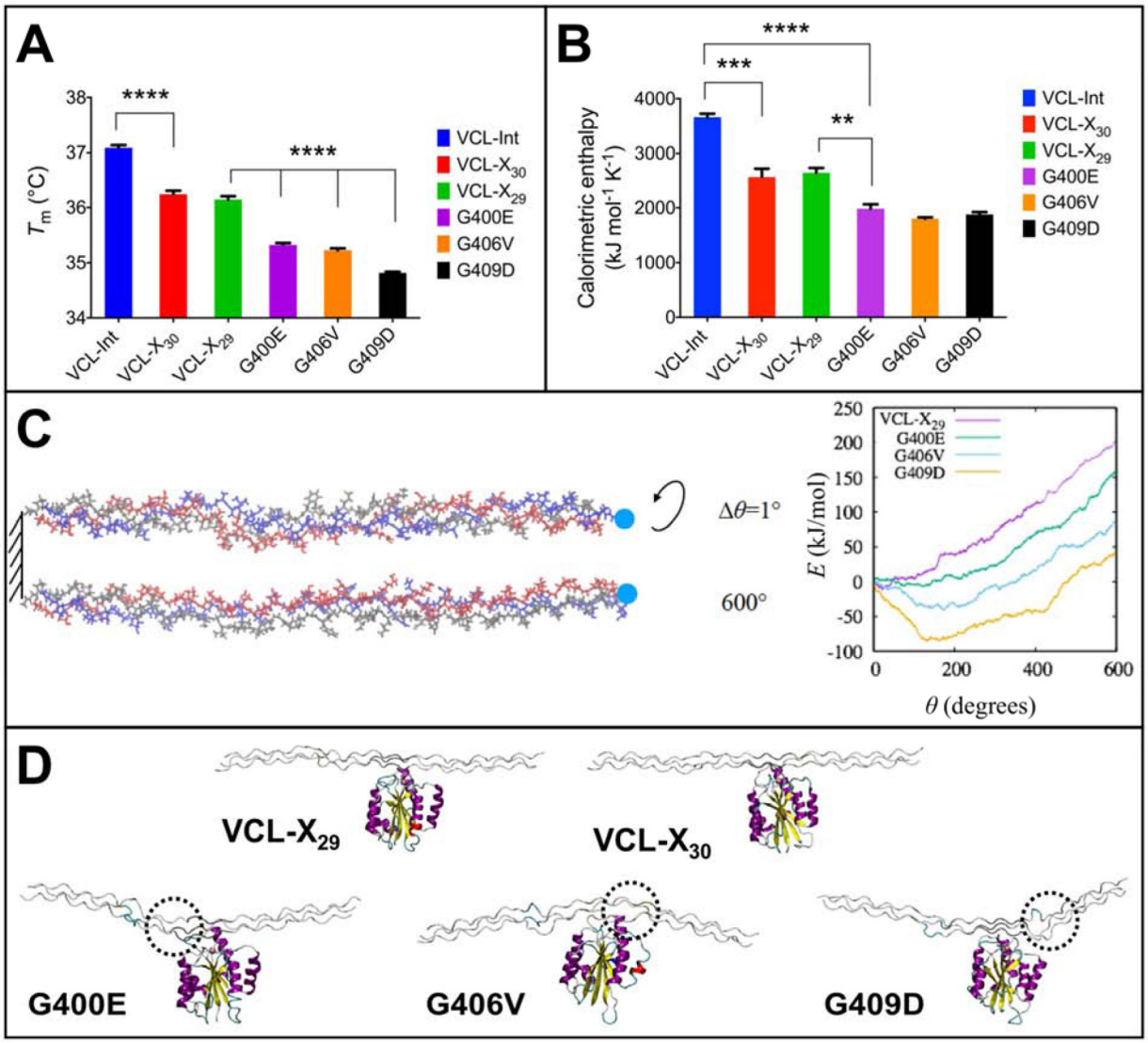
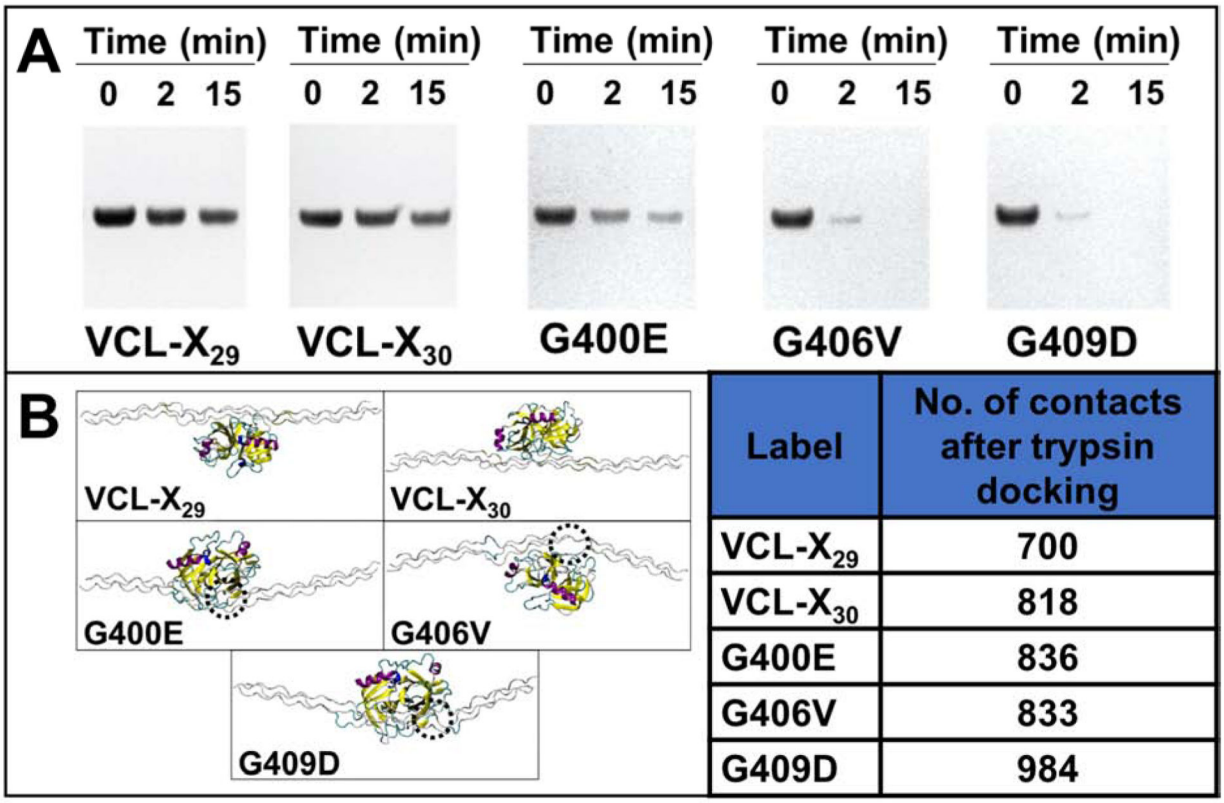


Figure 2.

Mutant collagen type IV proteins were much less stable structurally with significantly reduced (a) melting temperature, (b) calorimetric enthalpy, and (c) work for corresponding peptides needed to unfold. (d) Qualitative agreement was also found with significant structural distortions in the mutated peptides when bound to integrin while the reference and wild-type collagen peptides remained relatively straight. Experiments were carried out in triplicates, with results based on the averages of data points and standard deviations presented as error bars. The significance level was determined by *p* value using paired Student's *t* test between the means of two samples. Statistically significant levels were marked with asterisks: *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001.



Label	No. of contacts after trypsin docking
$VCL\text{-}X_{29}$	700
$VCL\text{-}X_{30}$	818
G400E	836
G406V	833
G409D	984

Figure 3.

(a) Collagen mutants were more susceptible to trypsin digestion, especially for G406V and G409D mutations. (b) From docking of bovine trypsin to the collagen type IV homotrimers, trypsin tended to dock closer to the distorted regions. The number of trypsin-collagen contacts was much higher for mutated collagen peptides, particularly for G409D, in good correlation with the experimental trypsin results.

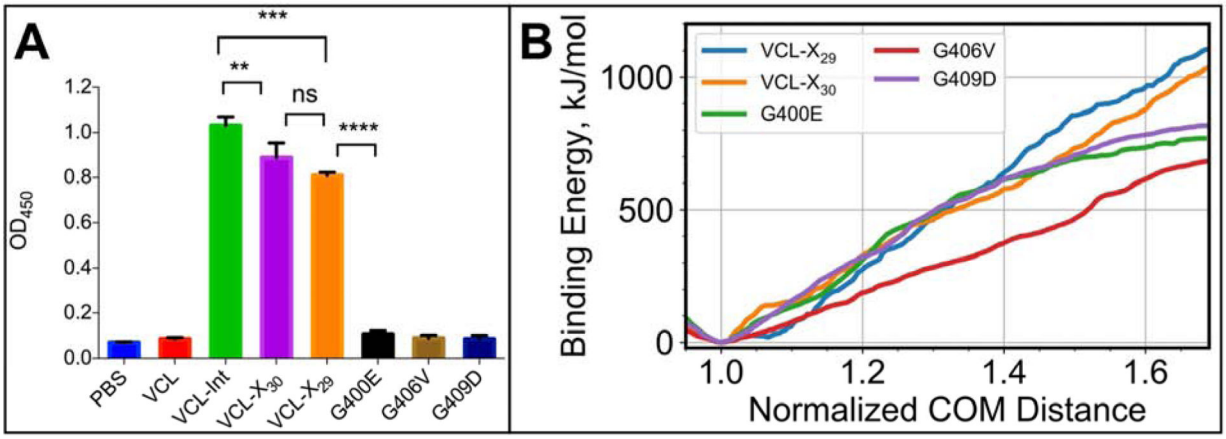


Figure 4.

(a) Solid-phase binding assay showed absences of integrin binding in recombinants with Gly substitutions, while both the VCL-X_{29} and VCL-X_{30} proteins had similar binding affinity. (b) MD sampling with adaptive biasing forces showed reductions in integrin binding energy ranging between 200 to 400 kJ/mol for the mutant peptides compared to the VCL-X_{29} and VCL-X_{30} peptides that also had similar binding energies. Experiments were carried out in triplicates, with results based on the averages of data points and standard deviations presented as error bars. The significance level was determined by *p* value using paired Student's *t* test between the means of two samples. Statistically significant levels were marked with asterisks: *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$.

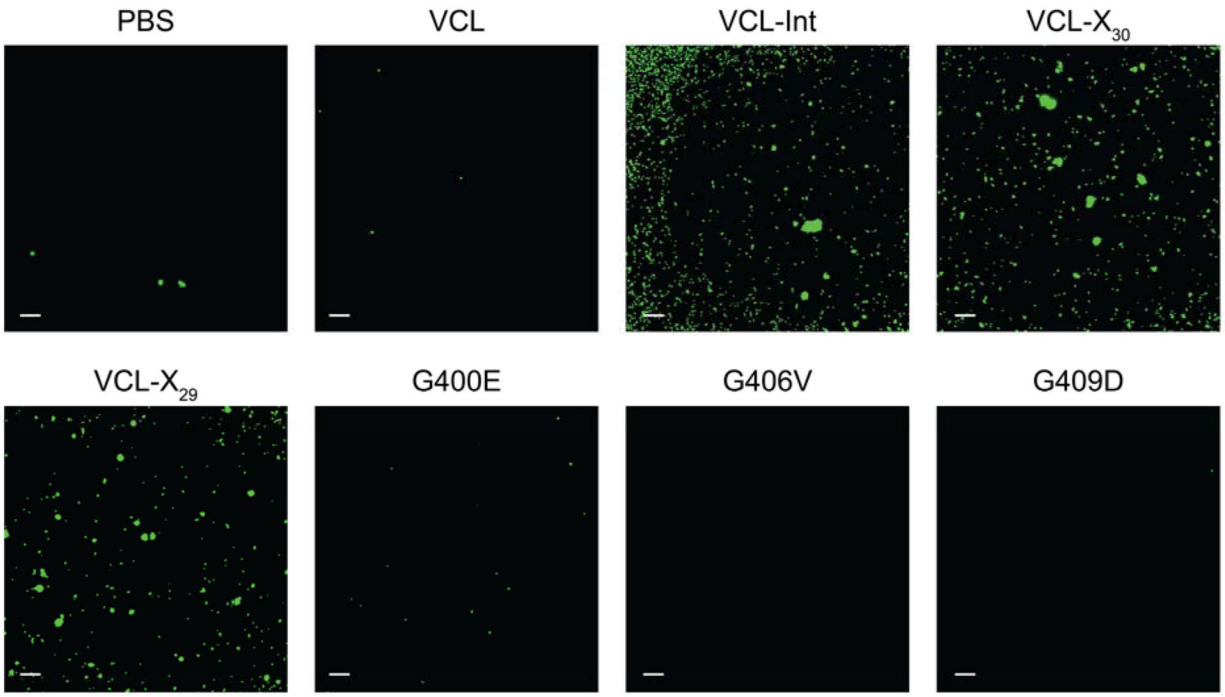


Figure 5.

Mammalian cell adhesion assay showed the *in vitro* binding profile of recombinant proteins with Gly interruptions and Gly substitutions. Live imaging was performed after calcein-AM staining of bound cells. Scale bars denote 100 μm.