The extracellular matrix protein EMILIN1 (elastin microfibril interface located protein 1) is implicated in maintaining blood pressure homeostasis via the N-terminal elastin microfibril interface domain and in trophoblast invasion of the uterine wall via the globular C1q (gC1q) domain. Here, we describe the first NMR-based homology model structure of the human 52-kDa homotrimer of the EMILIN1 gC1q domain. In contrast to all of the gC1q (crystal) structures solved to date, the 10-stranded β-sandwich fold of the gC1q domain is reduced to nine β strands with a consequent increase in the size of the central cavity lumen. An unstructured loop, resulting from an insertion unique to EMILIN1 and EMILIN2 family members and located at the trimer apex upstream of the missing strand, specifically engages the α4β1 integrin. Using both Jurkat T and EA.hy926 endothelial cells as well as site-directed mutagenesis, we demonstrate that the ability of α4β1 integrins to recognize the trimeric EMILIN1 gC1q domain mainly depends on a single glutamic acid residue (Glu933). Static and flow adhesion of T cells and haptotactic migration of endothelial cells on gC1q is fully dependent on this residue. Thus, EMILIN1 gC1q-α4β1 represents a unique ligand/receptor system, with a requirement for a 3-fold arrangement of the interaction site.

EMILIN1 (elastin microfibril interface located protein 1) is a secreted extracellular matrix multidomain glycoprotein (1, 2). It is characterized by a unique arrangement of structural domains, including the elastin microfibril interface domain at the N terminus, an α-helical domain predicted to form a coiled-coil structure in the central part of the molecule, a short collag enous sequence, and a region homologous to the globular domain of C1q (gC1q domain) at the C-terminal end (3, 4). Although the role of the coiled-coil region has not yet been elucidated, it has conclusively been demonstrated that EMILIN1 interacts with pro-tumor growth factor-β (5) through the elastin microfibril interface domain (6). EMILIN1 deficiency causes systemic arterial hypertension, and the expression of EMILIN1 at physiological levels by binding to pro-tumor growth factor-β prevents its maturation by protein convertases (5). Thus, EMILIN1 favorably located at the subendothelium of blood vessels is a new specific antagonist of tumor growth factor-β, and the function of this constituent of elastic tissues is linked to the pathogenesis of hypertension. The C-terminal gC1q domain is involved in the oligomerization of EMILIN1 (7), in cell adhesion and migration via interaction with the α4β1 integrin (8), and in trophoblast invasion (9).

The gC1q signature is found in a variety of proteins, and the essential features of the specific structure-function relationship were recognized with the elucidation of the crystal structure of the homotrimeric gC1q domain of mouse ACRP30 (adipocyte complement-related protein of 30 kDa) (10). It suggested a structural and evolutionary link between the tumor necrosis factor and the gC1q domains and led to the recognition of a gC1q/tumor necrosis factor superfamily. Three other gC1q structures, namely those of human collagen X (11), mouse collagen VIII (12), and human complement C1q (13), have been solved by x-ray crystallography. Each gC1q domain adopts a 10-stranded β-sandwich fold with a jelly roll topology, consisting of five-stranded β-sheets (A, A', H, C, F, B', B, G, D, and E), each made of antiparallel strands. Besides sharing the folding topology, the human complement C1q, ACRP30, collagen X, and collagen α1VIII gC1q domains also have similar oligomerization motifs derived from the noncovalent association of three polypeptide chains to form trimeric globular domains. Homotrimeric domains are formed by ACRP30, collagen X, and collagen α1VIII gC1q domains, whereas the association of the trimeric EMILIN1 gC1q domain reveals a disordered insertion necessary for interaction with the α4β1 integrin.

The Solution Structure of EMILIN1 Globular C1q Domain Reveals a Disordered Insertion Necessary for Interaction with the α4β1 Integrin*§
three different chains gives rise to a heterotrimer in the complement C1q protein.

The EMILIN1 gC1q domain shares similar residues with the gC1q domains of several other extracellular matrix and non-extracellular matrix components. Based on the homotrimeric model of ACRP30 gC1q crystal structure (10), it was supposed that the EMILIN1 gC1q domain could also trimerize via several hydrophobic residues. This was confirmed, first by biochemical (3, 7) and later by preliminary structural evidence (14).

So far, among the C1q superfamily members, EMILIN1 gC1q is the only one capable of interacting with $\alpha 4\beta 1$ integrin, a molecule, predominantly expressed on the surface of hemopoietic cells, that serves as a receptor to allow adhesion to blood vessel wall (15, 16) or extracellular matrix constituents (17). The interaction between $\alpha 4\beta 1$ and EMILIN1 gC1q is particularly efficient because even very low ligand concentrations provide very strong adhesion (8) and migration (9).

In the present study, we report the NMR solution structure of the EMILIN1 gC1q homotrimer domain and demonstrate that the $\alpha 4\beta 1$ binding site is located in a flexible loop at the apex of the trimeric structure and that it is absolutely dependent on the Glu933 residue. The data provide a new understanding of how this ligand/receptor pair interact to mediate $\alpha 4\beta 1$ integrin-dependent haptotactic cell migration as well as cell adhesion, under both static and flow conditions, to subendothelial EMILIN1.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**—The recombinant human EMILIN1 gC1q domain spanning residues from 846 to the end of the human EMILIN1 published sequence (3) was expressed as His$_6$-tagged protein and purified as previously described (7). This recombinant product forms a stable trimeric protein with an acidic isoelectric point (pI, 5.32) having a molecular mass of 51,624 Da. The increase of 1,151 Da compared with the value previously published (14) arises from correcting the primary sequence that had been used for calculating the molecular mass by nucleotide sequence controls and mass spectrometry measurements. A 20 mM phosphate, 100 mM NaCl, 0.02% (w/v) NaN$_3$, 5% (v/v) D$_2$O, pH 7.5 solution was used to dissolve the protein, yielding a clear sample that proved to be stable for several months. A 1.8 mM protomer concentration sample (0.6 mm homotrimer concentration) of protein uniformly labeled in $^{15}$N and $^{13}$C was used to measure the protein linewidth. Proton chemical shifts were referenced indirectly from the hydrogen reference frequency by heteronuclear absolute frequency ratio. Induction Decays (RDC) information measured in weakly aligned media were used as orientational restraints for structure determination (25, 26). The homology model refinement protocol followed a low temperature simulated annealing strategy described by Chou et al. (23), implemented in the program Xplor-NIH (24). The molecular dynamics algorithm of Xplor-NIH, besides containing conditional energy term penalties to enforce experimental constraints, includes additional potential energy terms specifically developed for NMR structure determination that bias the conformational space sampling in favor of regions that are physically accessible, thus improving the performance of the structure calculation. For our purposes, a crucial feature of the Xplor-NIH program was the possibility of including a dipolar energy term to exploit the experimental residual dipolar coupling (RDC) information measured in weakly aligned media to derive orientational restraints for structure determination (25, 26). For the analysis of EMILIN1 gC1q domain in anisotropic phase, a uniformly compressed polyacrylamide gel system was obtained over a $^{13}$C sweep width of 28–60 ppm, with 512 t1 increments of 1024 complex data points and 32–96 scans/t1. Three-dimensional spectra were typically acquired with 64 × 48 × 512 complex data points and 32–96 scans/FID (Free Induction Decay). Proton NOE connectivities were identified from the analysis of a three-dimensional $^{15}$N NOE spectroscopy heteronuclear single quantum coherence spectrum (20) recorded at 600 MHz (1H frequency) with a mixing time of 150 ms and acquisition times of 24.9 ms (t1, 1H), 8.8 ms (t2, 15N), and 57.3 ms (t3, 1H) for a three-dimensional matrix of 220 × 18 × 512 complex data points. Two-dimensional $^{15}$N[1H] NOE measurements (21) were also performed at 500 MHz and 37 °C, with a 5-s relaxation or saturation delay, 160 t1 increments of 1024 complex data points, and 480 scans/t1. The data were processed and analyzed with Felix (Accelrys), and Sparky 3.5. Original sizes ($^{15}$N and $^{13}$C indirect dimensions) were typically increased by linear prediction. Zero filling followed by apodization with shifted sine bell functions was applied prior to Fourier transformation. Proton chemical shifts were referenced to internal dioxane (3.750 ppm), whereas $^{13}$C and $^{15}$N chemical shifts were referenced indirectly from the hydrogen reference frequency by heteronuclear absolute frequency ratio.

**Homology Modeling and Refinement**—The homology model for the EMILIN1 gC1q domain was built after alignment with ClustalW (22), including 13 human proteins of the C1q protein family. The A chain of ACRP30 protein (ACRP30-A) (10) was recognized as the best sequence matching the EMILIN1 gC1q domain (supplemental Table S1), based on the pairwise alignment score and the number of matched amino acids, and used as a template for the modeling. Overall, 20% of sequence identity and 31.3% of sequence similarity were found on comparing the query protein and the template. Proton NOE connectivities were identified from the analysis of a three-dimensional $^{15}$N NOE spectroscopy heteronuclear single quantum coherence spectrum (20) recorded at 600 MHz (1H frequency) with a mixing time of 150 ms and acquisition times of 24.9 ms (t1, 1H), 8.8 ms (t2, 15N), and 57.3 ms (t3, 1H) for a three-dimensional matrix of 220 × 18 × 512 complex data points. Two-dimensional $^{15}$N[1H] NOE measurements (21) were also performed at 500 MHz and 37 °C, with a 5-s relaxation or saturation delay, 160 t1 increments of 1024 complex data points, and 480 scans/t1. The data were processed and analyzed with Felix (Accelrys), and Sparky 3.5. Original sizes ($^{15}$N and $^{13}$C indirect dimensions) were typically increased by linear prediction. Zero filling followed by apodization with shifted sine bell functions was applied prior to Fourier transformation. Proton chemical shifts were referenced to internal dioxane (3.750 ppm), whereas $^{13}$C and $^{15}$N chemical shifts were referenced indirectly from the hydrogen reference frequency by heteronuclear absolute frequency ratio.
chosen and heteronuclear one-bond RDCs were measured for N-H vectors (1D_{NHN}), CO-N vectors (1D_{CN}) and CO-C vectors (1D_{CCO}) through the IPAP (In Phase Anti Phase) approach (27).

Site-directed Mutagenesis—Two sets of alanine mutants were prepared. One comprised a series of mutations in the most conserved residues of EMILIN1 gC1q domain including F852A, F886A, Y944A, L956A, F981A, and L986A. The second set comprised a series of mutants in the unstructured segment unique to EMILIN1 and EMILIN2. The mutations were E930A (L932A/E933A), E933A, (G945A/G948A), F950A, and F950Y. The codon mutation within the DNA sequence of EMILIN1 gC1q were generated by site-directed mutagenesis using the overlapping PCR approach (28). Briefly, in a first PCR the primer carrying the desired mutation was used in combination with 5'- and 3'-flanking primers to generate two overlapping fragments using, as template, the wild type EMILIN1 gC1q coding sequence inserted in the expression vector pQE-30 (Qiagen) between the I and Kpn restriction sites. The overlapping fragments were gel-purified and used as templates in a two-step PCR consisting of 12 elongation cycles in which the overlapping region work as primers and addition of the 5'- and 3'-flanking primers carrying BamHI and KpnI restriction sites, respectively, followed by 25 amplification cycles. The final products were purified by wizard SV columns (Promega), cut with BamHI and KpnI, and ligated in pQE-30 cut with the same enzymes. Ligation products were transformed in MI5 bacterial strain. Four deletion mutants, del(Leu932–Phe950), del(Pro942–Phe950), del(Tyr927–Gly945), and del(Leu932–Gly945) were transformed in MI5 bacterial strain. Four deletion mutants, del(Leu932–Phe950), del(Pro942–Phe950), del(Tyr927–Gly945), and del(Leu932–Gly945) were generated following the same protocol.

Recombinant Protein Purification and SDS-PAGE Analysis—Recombinant proteins were expressed as His_{6}-tagged protein and extracted under native conditions as previously described (7). Briefly, 100 ml of liquid culture grown at 0.6 A_{600} nm was induced with 2 mM isopropyl-1-thio-D-galactopyranoside for 3 h at 37 °C. The culture was then centrifuged at 4,000 g for 20 min, and the cell pellet was resuspended in sonication buffer (50 mM sodium phosphate, pH 8.0, 0.3 M NaCl) at 5 volumes/g of wet weight. The samples were frozen in a dry ice/ethanol bath, thawed in cold H_{2}O, incubated 1 h with 1 mg/ml lysozyme, and sonicated on ice (1-min bursts/1-min cooling/200–300 watts). The cell lysate was centrifuged at 10,000 × g for 20 min; supernatant and pellet were loaded on SDS-PAGE gel to check the respective amounts of soluble and insoluble fractions. The purification of the soluble mutants was performed by affinity chromatography on nickel-nitrilotriacetic acid resin (Qiagen). When resistance to denaturation by heating in the presence of 1% SDS was assayed, the samples were incubated in sample buffer (20 mM Tris-HCl, pH 7.6, 100 mM NaCl, 25 mM EDTA, 1% SDS, 2 mM phenylmethanesulfonyl fluoride, 5 mM N-ethylmaleimide) for 1 h at different temperatures just before gel loading.

Cell Adhesion Assay—The quantitative cell adhesion assay used in this study is based on centrifugation and has been extensively described (29, 30). Human Jurkat T cell leukemia or endothelial EA.hy926 cells that display a number of features characteristic of vascular endothelial cells (31) were labeled with the vital fluorochrome calcein AM (Molecular Probes Inc., Eugene, OR) for 15 min at 37 °C and then aliquoted into the bottom miniplates assemblies. In experiments aimed at examining the effects of blocking antibodies, i.e. 1H8 against EMILIN1 gC1q (8) and rabbit polyclonal against EMILIN1 (7), the various antibodies were added directly to the wells, just before plating the cells. The relative number of cells bound to the substrate (i.e. remaining in the wells of the bottom miniplates) and cells that fail to bind to the substrate (i.e. remaining in the wells of the top miniplates) was estimated by top/bottom fluorescence detection in a computer-interfaced GENios Plus microplate fluorometer (TECAN Italy) (30).

Cell Motility Assay—Haptotactic-like migration of EA.hy926 cells in response to wt Glu933 or mutant E933A substrates was assessed by fluorescence-assisted transmigration invasion and motility assay (30–32). The procedure is based on the use of Transwell-like inserts carrying fluorescence-shielding porous polyethylene terephthalate membranes (polycarbonate-like material with 8 μm pores) and HTS FluoroBlok™ inserts (Becton-Dickinson, Falcon, Milan, Italy). The membranes were coated (20 μg/ml in 50 μl) on the underside in bicarbonate buffer at 4 °C overnight and blocked with 1% bovine serum albumin for 1 h at room temperature. The cells were fluorescently tagged with the lipophilic dye DiI (Molecular Probes) used at a final concentration of 5 μg/ml for 10–15 min at 37 °C. The cells were washed, resuspended, and then added to the upper side of the inserts (2 × 10^5 cells/insert), and migratory behavior was monitored at different time intervals using the computer-interfaced SPECTRAFluor Plus instrument (TECAN Italy).

Perfusion Experiments—Proteins were coated on glass coverslips (24 × 50 mm) that represented the lower surface of a modified Richardson’s parallel flow chamber mounted on an inverted microscope; a silicon rubber gasket determined the flow path height (125 μm) between the glass coverslip and the upper plate. The chamber was assembled and filled with phosphate-buffered saline, pH 7.4. A syringe pump (Harvard apparatus Inc., Boston, MA) and silicone tubing connections were used to aspirate the Jurkat cell suspension through the chamber at a flow rate of 20 s^{-1}. The perfusion system was mounted on an inverted microscope equipped with epifluorescent illumination (Diaphot-TMD; Nikon) and an intensified CCD video camera (C-2400-87, Hamamatsu Photonics). The images were captured at a sampling rate of 25 frames/s during the experiment, and image analysis was performed using recently developed algorithms and programs (33) based on the MatLab Image tool (the MathWorks, Inc.).

RESULTS

NMR Structure of the EMILIN1 gC1q Domain—The recombinant EMILIN1 gC1q domain exists in solution as a stable trimeric protein formed by three identical polypeptide chains of 162 amino acids (150 from the natural domain sequence, 7 from N-terminal fused histidine tag and initial Met residue, and 5 derived from the cloning strategy) with an overall molecular mass of 51,624 Da (3). The 3-fold symmetry, expected from homotrimeric assembly, makes NMR analysis viable, albeit very challenging. We determined the three-dimensional solution structure of the EMILIN1 gC1q domain using a hybrid approach that involved homology modeling and structure...
EMILIN1 gC1q Structure and α4β1 Interaction Site

The secondary structure of EMILIN1 gC1q domain was derived from the consensus of three different NMR approaches, summarized in the scheme. Below the amino acid sequence of the protein (third line), experimental data are reported as follows: amide protons showing slow exchange with the bulk solvent classified into two groups, slow exchange (violet circle) and medium-slow exchange (light blue circle); TALOS backbone torsion angle predictions for residues belonging to β-strands (green B) or tight turns (green T); 13Cα chemical shift deviations from random coil values (red histogram, vertical scale in ppm). These data led to the prediction of nine segments of β-strands represented with yellow arrows. For comparison the β-strands elements found in our best refined model are reported as blue arrows in the top line.

FIGURE 1. NMR-based secondary structure of EMILIN1 gC1q domain. A possible location of secondary structure elements for the EMILIN1 gC1q domain has been derived from the consensus of three different NMR approaches, summarized in the scheme. Below the amino acid sequence of the protein (third line), experimental data are reported as follows: amide protons showing slow exchange with the bulk solvent classified into two groups, slow exchange (violet circle) and medium-slow exchange (light blue circle); TALOS backbone torsion angle predictions for residues belonging to β-strands (green B) or tight turns (green T); 13Cα chemical shift deviations from random coil values (red histogram, vertical scale in ppm). These data led to the prediction of nine segments of β-strands represented with yellow arrows. For comparison the β-strands elements found in our best refined model are reported as blue arrows in the top line.
A, A’, H, C belonging to sheet 1, and the unstructured region spanning residues Leu947–Pro925, whereas strands B, B’, D, E, and G constitute the external second sheet. The hydrophobic core of the EMILIN1 gC1q monomer is formed mainly by the side chains of residues Phe852, Tyr879, Phe886, Tyr894, Leu956, Val962, and Gly983 (Fig. 2C), all belonging to the group of most conserved residues in C1q family (4). Notably, the corresponding side chains of the homologous protein ACRP30 (chain A) have similar orientations, which highlights a conservative overall packing of the cores in the two molecules. Thus, the refined model maintains the same folding topology as the starting homology model, although some changes are observed: (i) loops, which were not restrained during the simulated annealing, show some deviations from the starting conformation; (ii) small variations are found in the length of β-strand elements; and (iii) segment 1953–1956, originally classified as β-strand (strand F) becomes unstructured. This region, however, still remains in contact with the adjacent subunit participating to the trimer interface, as confirmed also by the reduced local mobility assessed by heteronuclear NOE measurements (not shown).

The most relevant changes that occur between the structures before and after refinement are in the strand F region where comparison shows that: (i) in the initial regularized homology model, strand E (Ala920–Gly925, sheet 2) is joined to strand F (Ile953–Leu956, sheet 1) by a long unstructured loop (Tyr927–Gly945); (ii) in the refined model, strand F melts away so that the whole secondary structure is a β-strand (E), a long unstructured loop (Tyr927–Gly945), and an interfacial region with no secondary classification (F). Although shorter in length, the position of strand E remains constant before and after the refinement. Also the location of region F remains invariant, despite the fact that the new φ and ψ angles no longer define a β strand. Region F, however, is still involved in intersubunit interactions, as observed in the structure prior to refinement. In other words, despite the loss of local secondary structure, region F is still expected to be rigid. In contrast, the region Tyr927–Gly945 was classified as flexible and disordered in every analysis performed, according to 15N1H NOE evidence. Therefore, we simply excluded that segment from the structure refinement consistently with the absence of any corresponding experimental restraints. The Tyr927–Gly945 loop is located mostly in the apical part of the trimer in the initial regularized model, prior to simulated annealing refinement. Because, either before and after refinement, we know precisely the location and orientation of the preceding strand E and of the following strand/molten strand F, we infer that the position of the loop ends should remain invariant in the final structure. Region Tyr927–Gly945 appears as a long flexible loop located by the apex of the molecule, possibly available for interactions with ligands/receptors of EMILIN1 gC1q.

The quaternary structure consists of three subunits that are mostly identical, apart from very small variations intrinsically arising during molecular dynamics simulations. The interface contacts between subunits are mainly hydrophobic and located near the base of the trimer, the top being still slightly open (≈150 Å2) to form a cavity ~16 Å deep. These contacts involve the following strand and adjacent segment residues: strand A’ (Arg870–Leu873), strand A (Ala851–Ala855), strand H (Phe981–Gly988), strand C (Arg893–Ala898), and region F (Leu947–Leu956). Among the 30 residues that establish intersubunit interactions, only two are charged (Arg870 and Arg893) and located at opposite ends of an extended hydrophobic surface. The presence of hydrophobic contacts near the base of the trimer has been highlighted for other members of the C1q family and apparently is a common feature of these molecules (collagen X, ACRP30, and collagen VIII). In contrast, the occurrence of hydrophilic interactions at the apex of the trimeric adduct, seen in collagen X, ACRP30, and collagen VIII, cannot be confirmed in our structure, which is relatively loose in that area.

By comparing the association of the monomers in the initially regularized homology model and in the final refined structure, a definite increase of the size of the central cavity encompassed by the three subunits is observed, probably because of topology changes undergone by each monomer during the refinement to optimize the molecular packing. The buried surface upon trimerization calculated with a probe radius of 1.4 Å is 3,128 Å2, a value that indicates a weaker intermonomer interaction with respect to the tight associations that lead to burying at least 5,000 Å2 in ACRP30 (5,320 Å2) and complement C1q protein (5,490 Å2) and up to 7,360 Å2 in collagen X.

Structure-Function Relationship—The available solution structure for the EMILIN1 gC1q domain allows the identification of some key residues in the molecule that might be crucial for the native structure formation and function. In addition, based on the correlation between native structure and function, it is also possible to predict functional roles for specific amino acids. These hypotheses have been tested here by functional assay of recombinant mutants of the EMILIN1 gC1q domain with respect to the wild type protein. Fig. 3 reports some of the results from site-directed mutagenesis experiments (additional data are available in supplemental Table S3).

As expected, introduction of single amino acid mutations (Fig. 3) for any of the hydrophobic core residues of the monomeric domain hampered the formation of the native conformation and resulted in insoluble recombinant proteins confined in bacterial inclusion bodies. Mutations to Ala were introduced at Phe852, Phe886, Gly892, Tyr894, Leu956, Phe981, and Leu986, and the resulting proteins were analyzed by SDS-PAGE to detect the expressed products either in the supernatant or following extraction from the cellular pellet. Compared with wild type EMILIN1 gC1q that was soluble to a large extent, none of the mutated proteins appeared soluble (supplemental Fig. S4). Changes in the side chains of those conserved residues might greatly impair correct folding and hence the solubility of EMILIN1 gC1q trimeric assembly. In addition, the region preceding and including the F strand, despite the loss of its β-structure, is functionally relevant for a correct trimer formation as indicated by the finding that the point mutation L956A greatly affects solubility of the protein, as does deletion of preceding segments. Intriguingly, the mutation F950A did not affect solubility but strongly modified trimer stability in 1% SDS because all of the protein was monomeric already at 0 °C (Fig. 3B).

The Unstructured Loop Tyr927–Gly945 Does Not Influence the Folding of EMILIN1 gC1q and Contains the α4β1 Recognition
EMILIN1 gC1q Structure and α4β1 Interaction Site

Site—Although α4β1 integrin is known to mediate cell adhesion and migration on an EMILIN1 substrate via the gC1q domain (8, 9), no EMILIN1 specific site recognized by this integrin has yet been identified. The finding of a unique structural characteristic, i.e. the presence of a flexible and disordered region in EMILIN1 compared with all the other resolved gC1q structures, prompted us to investigate the possible functional role of the Tyr927–Gly945 segment.

Thus a number of deletion and single point mutants were designed (Fig. 3A), expressed as recombinant fragments in a bacterial system, purified under native conditions, and monitored by SDS-PAGE (Fig. 3B). The two deletion mutants del(Leu932–Phe950) and del(Pro942–Phe950), which lack, respectively, part of the unstructured segment Tyr927–Gly945 and five of the residues joining the unstructured loop to the molten F region, were almost completely retained in bacterial inclusion bodies highlighting the key role of the “molten” F strand for correct molecular folding. In contrast the deletion mutant del(Tyr927–Gly945), lacking the entire unstructured loop, was fully soluble; the slight reduction in thermostability observed is most likely a consequence of the loss of conformational entropy (Fig. 3B). This result is consistent with the NMR analysis that indicated the presence of an unstructured loop spanning the region Tyr927–Gly945, which does not appear to be involved in monomer folding nor trimer formation.

By comparing the sequence of this region with the known α4β1 binding sequence of fibronectin, vascular cell adhesion molecule 1 and thrombospondin 1 and 2 (37–39), we identified a potential EMILIN1 gC1q integrin recognition site corresponding to the amino acids Leu932–Glu943, i.e. the core of EMILIN1 gC1q fragment 931–934 (GLEN) matching an analogous fragment in vascular cell adhesion molecule 1 (KLEK) (Fig. 3D). These
EMILIN1 gC1q Structure and α4β1 Interaction Site

core residues belong to the stretch EXLE, which is not only shared between human EMILIN1 and EMILIN2 gC1q domains but is also highly conserved in the orthologs from several distant species (Fig. 3C). Thus, the L932A/E933A double mutant was designed together with the single mutant involving the acidic residues, i.e. E933A. To exclude the involvement of other acidic residues located nearby, a shorter deletion mutant del(Leu932–Gly945) and two more single mutants, E930A and E939A, were generated and analyzed. Finally, the importance of the residues in the region between the unstructured loop and the molten F strand was assessed by expression of the double mutant G945A/G948A. All of the designed single and double-point mutants proved soluble and had the same thermal stability as the wild type EMILIN1 gC1q, suggesting correct folding. An example of E933A and del(Tyr927–Gly945) solubility compared with wild type EMILIN1 gC1q is shown in Fig. 3B.

The different point-mutated proteins and the shortest deletion mutant del(Leu932–Gly945) were then assayed in standard cell adhesion assays. Jurkat cells that express high levels of the α4β1 integrin attached in a dose-response assay very efficiently to wild type and mutated EMILIN1 gC1q except del(Gly932–Gly945), E933A, L932A/E933A, and wild type EMILIN1 gC1q denatured in 8 M urea (Fig. 4A). The cells did not adhere to the nonfunctional mutants even at high concentration of plated ligand (up to 10 μg/ml). Thus, amino acid substitution mutants revealed that the conserved residue Glu933 is essential for binding of EMILIN1 gC1q to integrin α4β1, whereas single mutations of several residues close to Glu933, do not affect the interaction between the two molecules. When the Glu933 was substituted with an aspartic acid, the cell anchorage was still very efficient, especially at the lowest doses of plated ligand.

FIGURE 2. Solution structure of the homotrimeric EMILIN1 gC1q domain. Structure of the homotrimeric EMILIN1 gC1q domain as obtained by homology model refinement with RDCs. A ribbon representation of the assembly, as side view, is presented in A. The three protomers in the trimer are shown in different colors (red, blue, and green). Each monomer has a nine-stranded folding topology profiled in B as stereoview. β-Strands are labeled according to the C1q/tumor necrosis factor superfamily nomenclature (10). The location of the unstructured segment Tyr927–Gly945 is highlighted in magenta (in A and B). C shows the amino acids forming the monomer hydrophobic core. The drawings were prepared using Open-Source PyMOL (DeLano Scientific LLC, South San Francisco, CA) and MOLMOL (45).
EMILIN1 gC1q Structure and α4β1 Interaction Site

FIGURE 4. Cell adhesion to EMILIN1 gC1q variant forms. A, wt and mutant gC1q polypeptides were plated on miniplate wells at different protein concentration. The reported data are the mean values ± S.D. of three independent experiments. To avoid graphic crowding the data from urea-denatured wild type EMILIN1 gC1q and mutant L932A/E933A are not reported as the corresponding curves overlap the traces of E933A (filled triangles), F950A (empty circles), and dell(Leu932–Gly945) (filled squares). The other mutants are: E933D (open diamonds), wt (open triangles), E930A (filled circles), and E939A (open squares). B, morphology of EA.hy926 human endothelial cells attached to wt Glu933 or mutant E933A under static conditions. C and D, adhesion (C) and haptotaxis (D) of EA.hy926 human endothelial cells to EMILIN1 gC1q variant forms. The reported data are the mean values ± S.D. of three independent experiments.

FIGURE 5. Attachment of Jurkat T cells under flow conditions. The cells were stained with 5 μM DIO in complete medium. The cells were injected into the flow chamber at fixed shear rate of 20 s⁻¹. A, Representative fields at 0 and 4 min. Arrested cells are indicated by arrows. B, attached cells after increasing flow times. The number of arrested cells/mm² was calculated as the mean ± S.D. of 10 different fields, and the global results are presented from two independent assays.

norsorbent assay assays. Neither E933A, nor dell(Leu932–Gly945), nor any other point mutation affected binding (supplemental Table S3).

The Glu933 Residue Is Required for Haptotaxis and for Cell Adhesion under Flow Conditions—As a further demonstration of the relevance of the residue Glu933, EA.hy926 human endothelial cells that express lower levels of α4β1 integrin were used in cell adhesion and haptotactic migration. The results demonstrated that mutation of Glu933 residue fully impaired cell adhesion (Fig. 4, B and C) as well as cell haptotaxis (Fig. 4D). Because static adhesion and haptotaxis experiments showed that cells are highly dependent on Glu933 for full functionality, we next mimicked cell flow in blood vessels to determine whether firm arrest of Jurkat T cells mediated by α4β1 was also related to residue Glu933. Under flow conditions at a shear rate of 20 s⁻¹ almost no Jurkat cells had arrested on E933A-gC1q-coated surfaces up to 4 min of flow. In marked contrast up to 160 cells/mm² were firmly arrested on coated wt gC1q-coated surfaces (Fig. 5).

DISCUSSION

To match structural evidence and functional inference, we have described the structural basis for the interaction between integrin α4β1 and the globular C1q domain of EMILIN1, a protein with important roles in regulating blood pressure homeostasis (5), cell adhesion (8), and cell migration (9). The interaction restraints were dictated by the solution structure of the isolated EMILIN1 gC1q domain trimer, as determined by NMR spectroscopy. The quaternary structure of EMILIN1 gC1q assembly identified, for each monomer, a unique unstructured loop, mostly located at the apex of the homotrimer, that mediates the interaction with α4β1. This loop is not present in any of the already determined gC1q structures. Sequence alignments suggest that a corresponding loop should occur in the analogous domain of the cognate human protein EMILIN2 (6) as well as in the paralogues from other organisms. Unique among α4β1 ligands, static and adhesion under flow of T cells as well as haptotactic migration of endothelial cells was fully dependent on a single acidic residue, Glu933.

Determining the structure of EMILIN1 gC1q trimer, i.e. the monomeric subunit conformation and the homotrimeric arrangement of an adduct with a molecular mass of ~52 kDa, required the application of several different experimental and modeling approaches that led to the structure depicted in Fig. 2, which represents the refined geometry with the lowest devia-
motif, i.e. arrangement because of a reverse turn at residues Ser951 and structure, are fundamental for a correct folding and trimer for-

assembly trimer.

gency limits on the number of strands that are necessary to edge strand melting, as originally referred to, could reflect strin-
mation at strand E of a single subunit within the trimer. The

Leu952. This geometry leads to a substantial increase in the size of the central cavity encompassed by the three subunits (Fig. 6). The cavity lumen is shadowed, with an upper entrance that is outlined by the ideal triangle identified by residues Gly926 (red, 19 Å apart) of each subunit. The drawing was prepared using Open-Source PyMOL (DeLano Scientific LLC).

It has been recognized that a number of highly conserved hydrophobic residues, buried in the core of the gC1q monomer structure, are fundamental for a correct folding and trimer for-

mation. Several single-point mutation experiments confirm, also for EMILIN1 gC1q, the essential role of those residues (in particular, Phe950, Phe956, Gly981, Tyr927–Gly945, and Leu956); the corresponding expressed recombinant proteins precipitated in inclusion bodies. Mutagenesis of residue Phe950, however, results in the successful isolation of a soluble F950A protein. This residue is located in the intersubunit cavity, precisely between the flexible and disordered region Tyr927–Gly945, and the diverging (βI-like) turn preceding the molten F strand remnant, i.e. fragment Ile953–Leu956. Therefore the hydrophobic Phe950 side chain, expected to establish interface interactions between monomers, does not appear essential for correct folding because its replacement does not affect EMILIN1 gC1q domain solubility, a prerequisite for successful assembly. However, residue Phe950 exerts a key role in trimer stability as inferred from the inability to detect the trimeric form in the presence of 1% SDS even at 0 °C (Fig. 3B) and the failure of specific recognition by monoclonal antibody 1H2G8 and a rabbit polyclonal antibody.

The complete loss of the cell adhesive function displayed by the EMILIN1 gC1q mutant devoid of segment Leu932–Gly945 indicates that the integrin binding capacity resides entirely in that region, a statistically disordered loop according to NMR evidence. Solubility and thermostability of this deletion mutant, and of the del(Tyr927–Gly945) mutant, are analogous to full-length EMILIN1 gC1q protein, indicating that region Tyr927–Gly945 is not involved in the folding and assembly of the domain, a conclusion further reinforced by a monoclonal antibody recognition assay. The limited decrease in melting temperature observed with the deletion mutants could reflect destabilization caused by conformational entropy loss, which would confirm further the NMR-based structural picture of a flexible loop. The mutagenesis experiments clearly demonstrate that the interaction between α4β1 and EMILIN1 gC1q is highly specific, because the del(Leu932–Gly945) mutation dramatically impairs cell binding. The conformational characteristics of fragment Tyr927–Gly945 provide a further example of the functional relevance of disordered regions in proteins (40). The structures of several polypeptides with integrin-binding capability have been determined; all of them display exposed aspartic or glutamic acid residues that are critical for integrin recognition and are generally located in mobile loops protrud-
ing from the main core of the whole ligand (41, 42). Although the aspartic acid-based sequences (e.g. RGD, LDV, KGD, RTD, and KQAGD) bind to the majority of integrins (43), other integ-

trins, including all β2 integrins, prefer ligands containing glutamtic acid-based motifs (44). Thus, we focused our attention on the three Glu residues occurring in the unstructured seg-

ment Tyr927–Gly945, Glu930, Glu933, and Glu939. Notably, Glu930 and Glu933 are completely conserved among all known EMILIN1 and EMILIN2 sequences, from fishes to mammals (Fig. 3C). The E933A mutation fully inhibited Jurkat T cell and E.A.hy926 endothelial cells adhesion and haptotactic migration. This loss of activity was not due to a global folding change of the protein as shown by biochemical and immunorecognition evidence. Therefore, Glu933 appears to play a crucial role in α4β1 integrin mediated interaction with the EMILIN1 gC1q domain.

FIGURE 6. Intersubunit central cavity of EMILIN1 gC1q homotrimer (top view). The cavity lumen is shadowed, with an upper entrance that is outlined by the ideal triangle identified by residues Gly926 (red, 19 Å apart) of each subunit. The drawing was prepared using Open-Source PyMOL (DeLano Scientific LLC).
EMILIN1 gC1q Structure and α4β1 Interaction Site

Furthermore, the present study suggests a potential structural basis for mediating firm arrest and attachment to subendothelial EMILIN1 gC1q. The shear flow rates used in our experiments likely mimic physiological inflammatory conditions in microvessels, when microaggregations induced by inflammation blood flow to these rates. T cells trafficking from blood into peripheral tissues occurs through a sequential process that involves tethering, rolling, firm adhesion, and transendothelial migration. The α4β1 integrin has been implicated in all of these phases of trafficking. When inflammation occurs, EMILIN1 gC1q might be exposed at sites of endothelial cells in postcapillary venules, and activated T lymphocytes, expressing high affinity integrins, might be capable of tethering and arrest.

This is the first report of an integrin-binding site located on a homotrimeric assembly that could employ a different mode of integrin engagement.

Because the NMR structure locates residue Glu933 in a flexible loop structure at the apex of the EMILIN1 gC1q domain assembled into a symmetric trimer, once the stoichiometry of three integrin molecules for a single gC1q trimer is ruled out for sterical reasons, the interaction pattern with α4β1 might well entail binding of the three Glu933 carboxylate groups to integrin via metal ion(s). This model for the interaction geometry patterns a testable hypothesis to characterize the functional modulation of the recognition event for this unique integrin-binding site located on a homotrimeric assembly of a functional ligand.

Acknowledgments—We acknowledge Drs. A. Makek and P.-A. Nicolesi for assistance and Dr. C.-J. Edgell for providing EA.hy926 cells.

REFERENCES

12. Kvasnakul, M., Bogen, O., Hohenester, E., and Yayon, A. (2003) Matrix Biol. 22, 145–152
13. Gaboriaud, C., Juanhuix, J., Grua, E., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J. C., and Arlaud, G. J. (2003) J. Biol. Chem. 278, 46974–46982
14. Verdone, G., Colebrooke, S. A., Boyd, I., Viglino, P., Corazza, A., Dolianna, R., Muniguierra, G., Colombatti, A., Esposito, G., and Campbell, I. D. (2004) J. Biol. Chem. 279, 91–92
15. Alon, R., Kassner, P. D., Carr, M. W., Finger, E. B., Hemler, M. E., and Springer, T. A. (1995) J. Cell Biol. 128, 1243–1253
16. Masumoto, A., and Hemler, M. E. (1993) J. Biol. Chem. 268, 228–234
17. Hauzenberger, D., Klominek, J., Bergström, S. E., and Sundqvist, K. G. (1995) Crit. Rev. Immunol. 15, 285–316
18. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 13585–13590
19. Salzmann, M., Wider, G., Pervushin, K., Senn, H., and Wuthrich, K. (1999) J. Am. Chem. Soc. 121, 844–848
20. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) J. Am. Chem. Soc. 111, 1515–1517
21. Grzesiek, S., and Bax, A. (1993) J. Am. Chem. Soc. 115, 593–595
22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
23. Chou, J. J., Li, S., and Bax, A. (2000) J. Biol. Chem. 275, 187–227
24. Schwiers, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) J. Magn. Reson. 160, 66–74
25. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) Nat. Struct. Biol. 4, 732–738
26. Tjandra, N., and Bax, A. (1997) Science 278, 1111–1114
27. Ottiger, M., Delaglio, F., and Bax, A. (1998) J. Magn. Reson. 131, 373–378
28. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. S., and Struhl, K. (1998) in Current Protocols in Molecular Biology, pp. 10.20.1, John Wiley & Sons, Inc., New York
29. Giacomello, E., Neumayer, J., Colombatti, A., and Perris, R. (1999) BioTechniques 26, 758–766
30. Spessotto, P., Giacomello, E., and Perris, R. (2000) Methods Mol. Biol. 139, 321–343
31. Edgell, C. J., McDonald, C. C., and Graham, J. B. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3734–3737
32. Spessotto, P., Gronkowska, A., De Marco, L., and Canu, P. (2006) J. Am. Chem. Soc. 128, 215–222
33. Clements, J. M., Newham, P., Shepherd, M., Gilbert, R., Dudgeon, T. J., Needham, L. A., Edwards, R. M., Berry, L., Brass, A., and Humphries, M. J. (1994) J. Cell Sci. 107, 2127–2135
34. Chou, J. J., Li, S., and Bax, A. (2000) J. Magn. Reson. 131, 373–378
35. Spessotto, P., Cervi, M., Pignol, D., Volpin, D., Bax, A. (1998) J. Magn. Reson. 131, 373–378
36. Machin, M., Santamasso, A., Mazzuccato, M., Cozzi, M. R., Battiston, M., De Marco, L., and Canu, P. (2006) Am. Biochem. Eng. 34, 833–846
37. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12366–12371
38. Wishart, D. S., and Sykes, B. D. (1994) J. Biol. Chem. 280, 3705–3712
39. Cornilescu, G., Delaglio, F., and Bax, A. (1999) J. Biol. Chem. 13, 289–302
40. Clements, J. M., Newham, P., Shepherd, M., Gilbert, R., Dudgeon, T. J., Needham, L. A., Edwards, R. M., Berry, L., Brass, A., and Humphries, M. J. (1994) J. Cell Sci. 107, 2127–2135
41. Kilger, G., Clements, J., and Holzmann, B. (1997) Int. Immunol. 9, 219–226
42. Vonderheide, R. H., Tedder, T. F., Springer, T. A., and Staunton, D. E. (1994) J. Cell Biol. 125, 215–222
43. Lebby, D. J., Aukhil, I., and Erickson, H. P. (2007) J. Magn. Reson. 192, 373–378
44. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12366–12371
45. Vonderheide, R. H., Tedder, T. F., Springer, T. A., and Staunton, D. E. (1994) J. Cell Biol. 125, 215–222
46. Lebby, D. J., Aukhil, I., and Erickson, H. P. (1996) Cell 84, 155–164
47. Casasnovas, J. M., Stehle, T., Liu, J. H., Wang, J. H., and Springer, T. A. (1997) Proc. Natl. Acad. Sci. U. S. A. 95, 2152–2157
48. Yamada, K. M. (1991) J. Biol. Chem. 266, 12809–12812
49. Michishita, M., Viden, V., and Arnaout, M. A. (1993) Cell 72, 857–867
50. Koradi, R., Billeter, M., and Wüthrich, K. (1996) J. Mol. Graph. 14, 51–55