Published in final edited form as:

Proteins. 2010 December ; 78(16): 3409–3427. doi:10.1002/prot.22848.

Computational Modeling of Laminin *N*-Terminal Domains Using Sparse Distance Constraints from Disulfide Bonds and Chemical Cross-Linking

Stefan Kalkhof^{1,5}, Sebastian Haehn², Mats Paulsson², Neil Smyth³, Jens Meiler^{4,*}, and Andrea Sinz^{1,*}

¹Department of Pharmaceutical Chemistry & Bioanalytics, Institute of Pharmacy, Martin Luther University Halle-Wittenberg, Wolfgang-Langenbeck-Strasse 4, D-06120 Halle (Saale), Germany

²Center for Biochemistry, Faculty of Medicine, Center for Molecular Medicine Cologne (CMMC), and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Joseph-Stelzmann-Strasse 52, Cologne D-50931, Germany

³School of Biological Sciences, University of Southampton, Bassett Crescent, East Southampton, SO16 7PX, United Kingdom

⁴Department of Chemistry and Center for Structural Biology, Vanderbilt University Nashville, TN 37212, USA

Abstract

Basement membranes are thin extracellular protein layers, which separate endothelial and epithelial cells from the underlying connecting tissue. The main non-collagenous components of basement membranes are laminins, trimeric glycoproteins, which form polymeric networks by interactions of their *N*-terminal (LN) domains; however, no high-resolution structure of laminin LN domains exists so far. To construct models for laminin β_1 and γ_1 LN domains 14 potentially suited template structures were determined using fold recognition methods. For each target/template-combination comparative models were created with Rosetta. Final models were selected based on their agreement with experimentally obtained distance constraints from natural cross-links, i.e., disulfide bonds as well as chemical cross-links obtained from reactions with two amine-reactive cross-linkers. We predict that laminin β_1 and γ_1 LN domains share the galactose-binding domain-like fold.

Keywords

laminin; *N*-terminal domains; disulfide bonds; chemical cross-linking; mass spectrometry; computational modeling

* Address correspondence to: Prof. Dr. Andrea Sinz, Martin Luther University Halle-Wittenberg, Phone: +49-345-5525170, Fax: +49-345-5527026, andrea.sinz@pharmazie.uni-halle.de, Prof. Dr. Jens Meiler, Vanderbilt University, Phone: +1-615-9365662, Fax: +1-615-9362211, jens.meiler@vanderbilt.edu.

⁵present address: Department of Proteomics, Helmholtz-Centre for Environmental Research – UFZ, Permoserstrasse 15, D-04318 Leipzig, Germany

INTRODUCTION

Laminins are the major non-collagenous proteins of basement membranes and are known to form networks due to crucial non-covalent self-interactions (1). Each member of the laminin protein family consists of three polypeptide chains, with one copy of the α , β and γ polypeptide chains being present. At the N-termini of all three chains globular (LN) domains are followed by a series of laminin-type EGF-like (LE) modules (for nomenclature see (2)) that embed one or two additional domains with a total length of 40–60 nm each. Towards the C-termini the three polypeptide chains intertwine into a long coiled-coil region of 77 nm that ends in five LG domains connected to the α chain. This arrangement gives laminins their typical cross-shaped structure carrying with three short and one long arm. The laminin *N*-terminal (LN) globular domains require the subsequent four LE-domains for efficient expression and proper folding (3). This region of the molecule is involved in Ca^{2+} dependent laminin self-assembly and binds to the corresponding *N*-terminal regions of other laminin chains (4). The LN domains of all three chains are required for efficient polymerization as deletion mutants with two or fewer LN domains fail to form networks (5).

The *N*-terminal portions of laminin chains share sequence homology and domain structure with the netrins, a family of extracellular proteins that were originally identified as neural guidance molecules. Netrin-4 was recently shown to interact with the *N*-terminal portions of laminin $\gamma 1$ and $\gamma 3$ chains in a manner that may regulate basement membrane assembly (6).

No high-resolution three-dimensional structures for laminin LN-domains have been obtained so far. An alternative approach that provides structural insight into proteins is based on chemical cross-linking and subsequent mass spectrometric analysis of the created products. Structural information can be obtained by the insertion of a chemical cross-linker between two functional groups within a protein. The cross-linker has a defined length and is connected via covalent bonds to functional groups of amino acid side chains. The cross-linked amino acids can be identified after enzymatic digestion. This chemical cross-linking approach is also applied to study protein-protein interfaces. The sequence separation of cross-linked amino acids, combined with the cross-linker length, impose a distance constraint on the structure of protein fold or protein complex (7) (8) (9) (10). Analysis of cross-linked peptides by mass spectrometry (MS) uses several advantages associated with MS analysis: (I) The mass of the protein or the protein complex under investigation is theoretically unlimited as the proteolytic peptides of the cross-linked proteins after an enzymatic digest are analyzed (in case a standard “bottom-up” strategy for mass spectrometric protein analysis is employed), (II) the analysis is rapid, (III) it requires very small (10^{14} – 10^{15} mol) amounts of protein, and (IV) as the cross-linking reaction can be executed in a native-like environment protein structure and flexibility are accurately reflected. It is possible to study membrane proteins, post-translational modifications, or splice variants. The broad range of cross-linking reagents with different specificities (primary amines, sulfhydryls, or carboxylic acids) and the wide range of distances (0 Å – >20 Å) allow setup of fine-tuned experimental strategies (7) (8) (9) (10) (11).

However, despite the straightforwardness of the cross-linking approach, the identification of the cross-linked products can be cumbersome due to the complexity of the reaction mixtures.

Several strategies have been employed to enrich cross-linker-containing species by affinity chromatography or to facilitate the identification of the cross-linked products; e.g., by using isotope-labeled cross-linkers or proteins, fluorogenic cross-linkers, or cleavable cross-linkers (10).

We recently reported the identification of the disulfide bond pattern of laminin β_1 LN-domain using *offline* nano-high performance liquid chromatography (nano-HPLC)/matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrometry (12). The aim of the present study was to obtain structural models of laminin β_1 and γ_1 LN-domains based on sparse distance constraints imposed by natural cross-links, i.e., disulfide bonds, as well as by chemical cross-links obtained from reactions with two amine-reactive cross-linkers. For chemical cross-linking of laminin LN domains, we used isotope-labeled, i.e. deuterated, cross-linkers to facilitate the identification of cross-linker-containing species in the mass spectra based on their characteristic isotope patterns (13) (14) (15).

Sparse experimental distance restraints can restrict the conformational space for a protein substantially and therefore enable determination of tertiary structure via computational modeling. For example, the Rosetta *de-novo* protein structure prediction algorithm (16) (17) allows the prediction of protein structures with medium to high resolution (1.5–3.0 Å) with less than one restraint per amino acid (distance and orientation restraints from NMR spectroscopy (18) (19)). Even with only approximately 0.1 constraints per amino acid the correct fold was determined *de novo* for T4-lysozyme and α A-crystallin using EPR distance restraints (20). The present project is challenging in that even fewer restraints are available for modeling. On the other hand, the protein models are not folded *de novo*, but template structures define the fold of the protein. Nevertheless, we underline that the structural models presented here should be treated as working hypotheses that need to be further validated and refined. We are confident that the availability of these models as supplementary information will facilitate this process.

MATERIALS AND METHODS

Materials

The cross-linking reagents BS³-*D₀D₄* bisulfosuccinimidyl)suberate-*D₀D₄*, and BS²G-*D₀D₄* bisulfosuccinimidyl)glutarate-*D₀D₄* were obtained from Pierce Inc. (Rockford, IL, USA). The proteases trypsin, chymotrypsin, LysC, endoproteinase AspN, and GluC (all sequencing grade) were obtained from Roche Diagnostics (Mannheim, Germany). MALDI matrices were obtained from Bruker Daltonik (Bremen, Germany), all other chemicals were purchased from Sigma (Taufkirchen, Germany). Nano-HPLC solvents were spectroscopic grade (Uvasol, VWR, Darmstadt, Germany). Water was purified with a Direct-Q5 water purification system (Millipore, Eschborn, Germany).

Expression and Purification of Laminin β_1 and γ_1 N-terminal Constructs

Laminin β_1 und γ_1 constructs (mouse) comprising one LN domain plus LE1-4 domains were expressed in 293-EBNA human embryonic kidney cells (3). Amino acid sequences

were confirmed by peptide mass fingerprint analysis using trypsin, endoproteinase AspN, chymotrypsin, endoproteinase GluC or a mixture of trypsin and AspN (enzyme:substrate ratio 1:50) as digestion enzymes.

Cross-Linking Reactions

For chemical cross-linking of laminin β_1 and γ_1 N-terminal constructs, the homobifunctional amine-reactive cross-linkers BS²G and BS³ were employed as 1:1 mixtures of non-deuterated and four-times deuterated derivatives (D_0/D_4). Cross-linking reactions were conducted with 2 μ M protein solutions in 20 mM HEPES buffer, 100 mM NaCl, 5 mM CaCl₂, pH 7.4. Freshly prepared stock solutions of the cross-linkers (10 mg/ml in DMSO) were added in 100- and 200-fold molar excess (final concentrations 200 μ M and 400 μ M) to the protein solution. The reactions were conducted at room temperature under gentle shaking of the reaction mixtures and were quenched after 45 and 90 min, respectively, by adding NH₄HCO₃ to a final concentration of 20 mM.

In-Solution Digestion

For *in-solution* digestion, the cross-linking reaction mixtures were denatured, reduced, alkylated, and digested with a mixture of AspN and trypsin (enzyme: substrate ratio 1:50) according to an existing protocol (21).

Gel Electrophoresis and In-Gel Digestion

A part of the cross-linking reaction mixtures was desalted with Microcon YM-10 filters (Millipore, Eschborn, Germany) and separated by one-dimensional SDS-PAGE (5% stacking gel / 5%, 8% or 12% resolving gel) according to Laemmli (22). The bands of monomeric laminin β_1 and γ_1 were excised, reduced, alkylated, and digested at 37°C for 16 hrs with a mixture of AspN and trypsin (enzyme: substrate ratio 1:30) as described previously (21). Peptides were extracted by adding three times 50 μ l of 5% TFA (for MALDI-MS analysis) or 5% (V/V) FA (for ESI-MS analysis); samples were concentrated in a vacuum concentrator to a volume of 5–10 μ l.

Nano-HPLC / MALDI TOF/TOF-Mass Spectrometry

Proteolytic peptide mixtures were analyzed by *offline* coupling of a nano-HPLC system (Ultimate 3000, Dionex, Idstein, Germany) to a MALDI-TOF/TOF mass spectrometer (Ultraflex III, Bruker Daltonik, Bremen, Germany). Samples were injected via an autosampler with a 200- μ l sample loop onto a precolumn (PepMap, C18, 300 μ m $\times$ 5 mm, 3 μ m, 100 Å, Dionex) and desalted by washing the precolumn for 15 min with 0.1% TFA before the peptides were eluted onto the separation column (PepMap, C18, 75 μ m $\times$ 150 mm, 3 μ m, 100 Å, Dionex), which had been equilibrated with 95% solvent A (A: 5% ACN, 0.05% TFA). Peptides were separated with a 30 min-gradient (0–30 min: 5–50% B, 30–31 min: 50–95% B, 31–35 min: 95% B (solvent B: 80% ACN, 0.04% TFA) at a flow rate of 300 nl/min with UV detection at 214 nm and 280 nm. Eluates were fractionated into 15-sec fractions with the fraction collector Proteineer fc (Bruker Daltonik), mixed with 1.1 μ l of matrix solution (0.7 μ g/ μ l HCCA in 90 % ACN/0.1 % TFA, 1 mM NH₄H₂PO₄) and directly prepared onto a 384 MTP 800 μ m AnchorChip target (Bruker Daltonik). Hystar software 3.2

controlled data collection with the nano-HPLC system, UV data acquisition, and fraction collector sampling.

MALDI-TOF-MS analyses were conducted in the positive ionization and reflectron mode by adding 2,000 laser shots in the range m/z 800–4,000 to one mass spectrum. Mass spectra were processed (Savitzky-Golay smoothing and baseline correction) and externally calibrated using Peptide Calibration Standard II (Bruker Daltonik). Monoisotopic mass signals with a signal-to-noise ratio (S/N) > 2 were automatically labeled using the SNAP (*Sophisticated Numerical Annotation Procedure*) algorithm. Afterwards, peak lists of labeled monoisotopic signals were created and signals with S/N > 10 were selected for laser-induced fragmentation. In the MS/MS mode, up to 2,000 laser shots were accumulated for measurement of the intact precursor ion (termination at S/N of 30); the optimum laser energy was determined for each precursor ion by fuzzy logic. Additional 2,000 laser shots were accumulated at 50% higher laser energy for acquisition of fragment ion mass spectra. Spectra were processed, calibrated based on the exact mass of the precursor ion, annotated (SNAP algorithm, S/N > 2) and combined with MS data to one data file. Data acquisition was done automatically by the WarpLC 1.1 software (Bruker Daltonik) coordinating MS data acquisition (FlexControl 1.3) and data processing (FlexAnalysis 3.0) softwares.

Nano-HPLC / Nano-ESI-FTICR Mass Spectrometry

Proteolytic peptide mixtures were additionally analyzed by nano-HPLC/nano-ESI mass spectrometry using the two FTICR (Fourier transform ion cyclotron resonance) mass spectrometers Apex II (Bruker Daltonics, Billerica, MA) and LTQ-FT (ThermoFisher Scientific, Bremen, Germany). The Apex II mass spectrometer was *online* coupled to the nano-HPLC system (Ultimate II with Switchos II and autosampler). Samples were injected via the autosampler with a 20- μ l sample loop, and were desalted and concentrated by washing the precolumn (PepMap, C18, 300 μ m $\times$ 5 mm, 5 μ m, 100 Å, Dionex) with 0.1% FA at a flow rate of 2 μ l/min for 10 min, before peptides were eluted onto the separation column (PepMap, C18, 75 μ m $\times$ 150 mm, 3 μ m, 100 Å, Dionex). The peptides were separated using a 30-min gradient (0–30 min: 5–50% B, 30–31 min: 50–95% B, 31–35 min: 95% B, with solvent A: 5% ACN, 0.1% FA and B: 80% ACN, 0.1% FA) at a flow rate of 200 nl/min monitoring the elution of peptides by their UV absorption at 214 nm und 280 nm. Data acquisition was controlled by the Hystar (version 2.3).

The Apex II FTICR mass spectrometer was equipped with a 7 T supra-conducting magnet and a nano-ESI source (Agilent Technologies, Waldbronn, Germany). For nano-ESI-MS measurements, *fused-silica*-nano-ESI needles (PicoTips, ID 8 μ m, New Objective, Woburn, MA, USA) were employed. The instrument was tuned using the doubly charged signal of the LHRH peptide at m/z 592.2358. Calibration was performed using the LHRH peptide fragments, which were obtained by capillary skimmer dissociation. Data were acquired in broadband mode (m/z range 400–2,000) with 256k data points per spectrum. 10 scans were accumulated to a single spectrum (ca. 12 sec per spectrum), the XMASS software (versions 7.0.2, 7.0.3, 7.0.8, Bruker Daltonik) was used for data acquisition and data.

Nano-HPLC/nano-ESI-MS measurements with the LTQ-FT hybrid mass spectrometer were conducted in *online* coupling with the Ultimate 3000 Nano-HPLC-System (Dionex). The

LTQ-FT combines a linear ion trap (LTQ) and an ion cyclotron resonance (ICR) analyzer and is equipped with a 7 T magnet. Samples were injected via the autosampler with a 100- μ l sample loop, desalted and concentrated on a precolumn (PepMap, C18, 300 μ m $\times$ 5 mm, 3 μ m, 100 Å). Separation of peptides was performed on a C18 column (PepMap, C18, 75 μ m $\times$ 150 mm, 3 μ m, 100 Å) using a 90-min gradient (0 to 60 % solvent B in 90 min, followed by isocratic elution at 90% B for 3 min; solvent A: 5% ACN, 0.1% FA, B: 80% ACN, 0.08% FA) at 300 nl/min. The Chromeleon software (version 2.3, Dionex) was used to control the HPLC system, UV data (214 nm und 280 nm) and MS data acquisition. Data acquisition was performed over 100 min. One duty cycle comprised one high-resolution full scan spectrum (m/z 300–2000, resolution 100,000 at m/z 400) in the ICR cell and 10 fragment ion mass spectra of the most intense signals in the linear ion trap. Dynamic exclusion (exclusion time 20 sec, exclusion window $\pm$ 5 ppm) was used to enhance acquisition of signals with low intensity.

Identification of Cross-Linked Products

Analysis of MS data and identification of cross-linked products was performed with the programs *IsoFind*, *GPMAW*, *Biotoools* und *MS2Assign*. The in-house software tool *Isofind* determines signals with a defined mass difference, i.e., 4.025 u for the identification of reaction products with D_0/D_4 labeled cross-linkers, yielding peak lists with the respective mass differences and peak intensity ratios. Only signals exhibiting the characteristic D_0/D_4 isotope patterns with similar peak intensities (0.25:1 to 1:4) and similar LC retention times ($\pm$ 1 min) were considered as potential cross-linked products. The GPMAW (General Protein Mass Analysis for Windows, version 8.0, Lighthouse Data, Odense, Denmark) (23) software was used for assignment of cross-linked products. Maximum mass deviations between theoretical and experimental masses of 5 ppm (LTQ-FT), 10 ppm (Apex II), and 50 ppm (Ultraflex III) were allowed. MS/MS data were manually compared to the predicted fragment masses for a cross-linked product. Additionally, the freely accessible program MS2Assign (<http://roswell.ca.sandia.gov/~mm-young/ms2assign.html>; part of the Collaboratory for MS3D, <http://ms3d.org/home.php>) was used for calculation of MS/MS data and comparison to experimental data. The software package Biotoools 3.1 (Bruker Daltonik) was used to identify peptides that are modified by a hydrolyzed cross-linker as well as intrapeptidal cross-linked products based on exact mass and MS/MS data. Lysines, serines, tyrosines, and threonines were considered as potential cross-linking sites (21) (24). Additionally, oxidation (Met), carboxamidomethylation (Cys), amidation (at cross-linker and at C-terminus), deamidation (Gln, Asn), cyclization (Cys) were taken into account as potential modifications as well as incomplete cleavage (up to 8 missed cleavage sites) (21).

Fold Recognition

The general workflow of computational modeling is summarized in Figure 1. Laminin β_1 and γ_1 sequences were split into separate domains as defined by the ExPASy Proteomic Server (www.expasy.ch) and modeled independently. The amino acid sequences of murine LN domains of laminin β_1 (ExPASy entry P02469, 31-270) and laminin γ_1 (ExPASy entry P02469, 44-283) were analyzed by the 3D-Jury metaserver (www.bioinfo.pl, (25), Figure 1A). Potential templates (PDB entries: 1CZT, 1D7P, 1KEX, 1XPW) were identified with a 3D-Jury score of 45–55, which roughly correlates with the number of amino acids that can

be superimposed between a template and a target structure with an RMSD <3.5 Å. A score larger 50 has a 90% chance of being a correctly identified fold. Sequence identities between laminin β_1 and laminin γ_1 and the templates were calculated with the software SIM (<http://www.expasy.ch/tools/sim-prot.html>, (26)) and were all smaller than 18%.

The quality of final comparative models is highly dependent on the template that is used for modeling. Given the low sequence similarity of the initial templates two alternative approaches were employed to more comprehensively search the PDB for additional potential template structures. In the first approach, homologous sequences of laminin β_1 and γ_1 LN domains were searched with PSI-BLAST (27) (www.ncbi.nlm.nih.gov/blast/Blast.cgi; E-value threshold 0.005; ten iterative rounds). In this experiment, the sequences of netrin I, II, IV as well as these of LN domains of laminin chains α_1 , α_2 , α_3 B, α_5 , β_2 , β_3 , γ_3 were identified to be related to the β_1 and γ_1 LN domains. These homologues as well as the sequences of the original 3D Jury fits 1CZT, 1D7P, 1KEX, 1XPW were submitted to the 3D-Jury metaserver for a second iteration. Two additional template candidates (1SDD and 1GOF) were identified by that procedure. To test orthogonal fold recognition approaches the threading servers Phyre [www.sbg.bio.ic.ac.uk/phyre/ (28)], Libellula [www.pdg.cnb.uam.es/servers/libellula, (29)], Wurst [www.zbh.uni-hamburg.de/wurst/, (30)], HHpred [<http://toolkit.tue-bingen.mpg.de/hhpred>, (31)], and Loopp [<http://loopp.org/>, (32)] were applied. 35 putative templates were identified in addition to the previously detected structures yielding a total of 40 candidate templates (Figure 1A).

In an unbiased approach the structures of all 40 potential templates were compared pairwise with the program Mammoth (<http://ub.cbm.uam.es/mammoth/> (33)) and clustered based on their structural homologies. 14 structures including 1CZT, 1D7P, 1GOF, 1KEX, 1SDD and 1XPW were grouped in one cluster of which all members share the galactose-binding domain-like fold (SCOP: 49784, CATH: 2.60.120.260, compare Table 4). These 14 structures were considered as templates in the subsequent comparative modeling efforts (Figure 1B). Other structures have been dismissed as potential templates as in structure clustering (Figure 2) only one cluster was observed containing the 14 members with a galactose-binding domain-like fold, whereas further clusters contained only two or less members. These 26 additional structures included immunoglobulin-like β -sandwiches and carbonic anhydrase among others.

Parametric Sequence Alignment

BCL::Align ((34), <http://www.meilerlab.org>) was used to create a pairwise sequence alignment between the sequences of laminin β_1 and γ_1 LN domains and the 14 template proteins. Given the low sequence similarity predicted secondary structure was used as component of the scoring function in sequence alignment. Secondary structure prediction of laminin LN domains was performed with the program JUFO ((19), (35), <http://www.meilerlab.org/>), Psipred ((36); <http://bioinf.cs.ucl.ac.uk/psipred/>) and SAM (37) yielding large content of random coil and β -strands. In order to sample the space of possible alignments densely a grid search of alignment parameters was performed (38) (Supplementary Table 1, Figure 1B) for each of the 28 combinations of a template with one LN domain creating 52,200 pairwise sequence-sequence alignments. After removal of

redundant alignments 5,371 and 9,830 unique alignments were obtained for laminin β_1 and laminin γ_1 , respectively. Moreover, alignments were excluded that prevent construction of all loops assuming that each amino acid can bridge approximately 3 Å. 4,333 (laminin β_1) and 7,306 (laminin γ_1) target/template-alignments fulfilling this rule were carried forward (Figure 1B).

Construction of Comparative Models

For each of the 11,639 alignments the backbone coordinates of the aligned regions were copied from the template structure into the model of the target. Loops were constructed in ten independent runs of Rosetta (39). Always the best of the 10 models with correctly constructed loop regions enter construction of side chain coordinates using the feature “-find_disulf”, which favors the formation of disulfide bonds. A gradient-based relaxation of the protein backbone structure was used to enable changes of the protein backbone conformation. A total of 3,550 and 5,104 models of laminin β_1 and γ_1 LN domains were constructed, respectively, each representing one of the original alignments (Figure 1C).

Validation and Refinement of LN Domain Structures

Models, in which not all loops could be closed, were removed (criterion 1). Additionally, models were filtered by energy (Rosetta total energy – bk_tot – larger than –100, criterion 2), and models with a solvent accessible surface area larger than 20,000 Å² for lack of compactness were removed (criterion 3) (Figure 1C). Further models were removed that displayed C^α-C^α distances for cross-links larger than 30 Å (criterion 4) or C^α-C^α distances of disulfide bridges larger than 13 Å (criterion 5) (Figure 1D & Table 5). A total of seven constraints were available for the laminin β_1 LN domain, five for the laminin γ_1 LN domain. Only three models of the laminin β_1 LN domain and twelve models of the γ_1 LN domain fulfilled all experimental constraints (Figure 1D). This highlights the power of the chemical cross-linking approach enabling an enormous reduction of potential structures, even if only a small number of distance constraints are obtained. In the present study, less than 1% of the initial models fulfill all distance constraints (Table 5). The models were ranked according to an empirical composite score derived from these four criteria (Table 6). The composite score minimizes the deviation between distances observed experimentally and in the model. At the same time, it minimizes Rosetta energy and solvent accessible surface area (see Supplementary Material).

Leave-One-Out Cross-Validation

To test the robustness of the model selection protocol a leave-one-out (LOO) cross-validation was performed by selecting the final models using only six out of seven (for laminin β_1) and four out of five (for laminin γ_1) distance constraints. Agreement of these models with the final constraint was taken as a measure of accuracy. Moreover, these models were compared to the models selected using all experimental data using the structure-structure alignment method (33). A Z-score larger than 8 was used as criterion for high structural homology (Table 7 & Supplementary Table 2). The final assessment of the models was based on precision as measured by the composite score (Table 6 & Supplementary Table 2).

Evaluation of Structure Quality

The structure quality of these models fulfilling all experimental distance constraints was evaluated using the programs VADAR ((40), <http://redpoll.pharmacy.ualberta.ca/vadar/>) and MolProbity ((41), <http://molprobity.biochem.duke.edu/>). The evaluated structures are shown in the Supplementary Material.

Estimation of the Influence of Templates and Distance Constraints

For each of the 14 templates the most likely structure for laminin β_1 and γ_1 was determined based on the composite score discussed above (Table 6). These structures were evaluated based on the structure homology to give the overall best model. To assess the importance of each distance constraints in model selection the fraction of models that fulfill each restraint before filtering was computed. Furthermore, we determined how often a single constraint is exclusively responsible for the rejection of models, i.e., in these cases for which all except one constraint are fulfilled.

RESULTS AND DISCUSSION

The aim of the present study was to obtain structural models of laminin β_1 and γ_1 LN domains based on sparse distance constraints imposed by natural cross-links, i.e., disulfide bonds, as well as by chemical cross-links obtained from reactions with the amine-reactive cross-linkers BS²G and BS³. The amino acid sequences of recombinant laminin β_1 and laminin γ_1 fragments comprised of one LN domain and LE domains 1-4 used in this study are shown in Figure 3.

Intramolecular Cross-Linked Products

The homobifunctional NHS esters BS³-D₀/D₄ and BS²G-D₀/D₄ bridging distances of ca. 7.7 Å and 11.4 Å were used for chemical cross-linking as described in the *Materials and Methods*. A sulfonate group at the NHS moiety improves water solubility of the reagent. NHS esters are highly reactive towards primary amines, i.e., ϵ -amine groups of lysines and the free N-terminus of a protein, but as a side reaction they are also susceptible to hydrolysis. The amino acids that are modified by a partially hydrolyzed cross-linkers (so-called “dead-end” cross-links) do not yield direct distance information, but give valuable insights into the solvent accessibility of a specific amino acid (9). In this work, we did not consider these “dead-end” cross-links for structural modeling. In addition to reacting with amine groups, NHS esters have also been found to react with hydroxyl groups of tyrosine, threonine and serine residues (21) (24). Upon cross-linking, NHS esters create an amide bond with mass increases of 138.068 u (BS³-D₀) and 96.021 u (BS²G-D₀), respectively. Peptides, which are modified by a partially hydrolyzed cross-linker exhibit mass increases of 156.079 u (BS³-D₀) and 114.032 u (BS²G-D₀), respectively. BS³ and BS²G were employed as 1:1 mixtures of their non-deuterated (D₀) and deuterated (D₄) species in order to facilitate the identification of cross-linked products by means of their distinct doublet isotope patterns with mass differences of 4.025 u (D₀/D₄) in the deconvoluted mass spectra (13).

To gain insight into the three-dimensional structures of laminin β_1 and γ_1 LN domains the cross-linking reaction had to be optimized. In contrast to determining protein-protein

interaction sites, intramolecular cross-linking within the laminin LN domain monomers had to be favored. Under the present conditions no formation of laminin homodimers was observed in MALDI-TOF-MS or 1D-SDS-PAGE. Gel bands of laminin LN domain monomers from cross-linking reaction mixtures, in which both laminin β_1 and γ_1 constructs were contained as well as the isolated laminin fragments in solution were digested either *in-gel* or *in-solution* and analyzed by nano-HPLC/MALDI-TOF/TOF-MS(/MS).

When isolated laminins were subjected to the cross-linking reaction with BS³ and digestion was performed *in-solution* and *in-gel*, five and two, intramolecular cross-linked products were obtained for laminin β_1 and γ_1 LN domain, respectively. Table 1 summarizes three unique cross-links for laminin β_1 and two cross-links for laminin γ_1 . In addition, intramolecular cross-linking products were identified from *in-gel* digestions of monomeric bands and *in-solution* digestions of cross-linking reaction mixtures with laminin β_1 and γ_1 LN domains using the cross-linkers BS³ and BS²G (Table 2). The cross-linking products that had already been found when performing *in-solution* digestion of separated laminin β_1 and γ_1 LN domain were confirmed. In total, eleven cross-linked products for laminin β_1 LN domain and four cross-linked products for laminin γ_1 LN domain were observed.

In Figure 4, the identification of a cross-link between lysines at positions 198 and 212 of laminin β_1 is presented exemplarily. MS/MS data unambiguously point to both lysines as reaction sites.

Disulfide Bond Analysis

The analysis of disulfide bond patterns of laminin β_1 LN domain by *offline* nano-HPLC/MALDI-TOF/TOF mass spectrometry has been described recently (12). Briefly, after recording the mass spectra the putatively disulfide-linked peptides are subjected to LIFT (laser-induced fragmentation technique)-TOF/TOF-MS/MS to confirm the disulfide bond. Screening the fragment ion mass spectra of disulfide-linked peptides for characteristic 66-u patterns (34 u + 32 u), arising from symmetric and asymmetric cleavage of disulfide bonds, greatly facilitates their identification (42). In addition, MALDI in-source decay created the reduced 'halves' of the disulfide-linked peptides and subsequent LIFT-TOF/TOF-MS/MS of the reduced peptides confirmed their amino acid sequences. Using different enzymes for proteolytic digestion of the laminin γ_1 chain *N*-terminal protein fragment, a linear bonding pattern of the eight cysteine residues in the LN domain of the laminin γ_1 chain was observed with a strict sequential 1–2, 3–4, 5–6, 7–8 connectivity of the eight cysteines. This disulfide bonding pattern had already been reported for the recombinant laminin β_1 chain fragment and a laminin fragment derived by elastase digestion of mouse tumor laminin-111, confirming that this pattern also occurs in native laminin (12). Table 3 summarizes the cysteines that are connected by disulfide bonds.

Structures of Laminin β_1 and γ_1 LN Domains

The secondary structure of LN domains was predicted using three different programs as described in the *Materials and Methods* resulting in a β -sheet content of ca. 30% for laminin β_1 and ca. 26% for laminin γ_1 . Fold recognition using the 3D-Jury method (25) was performed using the sequences of all murine laminin LN domains and the homologous

sequences of netrin I, II, and IV as input. A total of six template structures (1CZT, 1D7P, 1GOF, 1KEX, 1SDD, 1XPW) were identified with 3D-Jury scores ranging between 39 and 58 (Table 4). All six structures share a common galactose-binding domain-like fold. In a complementary approach, 35 additional templates were determined by sequence-structure alignment threading servers Phyre, Libellula, Wurst, HHPred, and Loop (28), (29), (30), (31), (32) (Table 4). In order to eliminate false positives, a pairwise structure-structure comparison of all putative templates was executed. 14 out of 40 templates displayed the galactose-binding domain-like fold (Table 4). The 26 remaining templates were removed as false positives as they exhibited similarity with one other template at most (Figure 2). Sequence identities between the templates and laminin β_1 and γ_1 LN domains ranged between 10 to 18%.

Sequence Assignment and Modeling of Loops

If sequence identities are lower than 50%, selection of the best possible template and accurate alignment of target and template sequences are crucial for success in comparative modeling (43). To comprehensively test all possible templates and alignments a parametric ensemble alignment approach was used for all 14 template proteins (38). Briefly, for each template laminin β_1 and γ_1 LN domain pair sequence-sequence alignments were created using a scoring function that includes sequence similarity (identity, Blosom45, Pam250, Blast), secondary structure (programs JUFO (19) (35), Psipred (36), SAM (37)), and sequence gaps (gap opening and gap extension) with variable weights. Out of the originally 52,200 alignments redundant alignments were removed as well as alignments that contained gaps too large to be closed by the respective loop. A total of 4,333 and 7,306 sequence-sequence alignments were input to comparative modeling for β_1 and γ_1 LN domains, respectively.

Construction of Comparative Models

The coordinates of the backbone of aligned amino acids were copied from the template structure into an initial model. Missing regions were reconstructed using Rosetta (39). Altogether, 4,333 alignments (for laminin β_1) and 7,306 alignments (for laminin γ_1) resulted in complete comparative models. Side chains were added using Rosetta rotamer libraries (44) assuming that all cysteines are involved in disulfide bonds. Models were removed if not all loop regions could be closed, the compactness of the model did not match native protein domains of the same size, not all cysteines were involved in disulfide bridges (or being at least in close distance to other cysteines $<13\text{\AA}$), the fold deviated significantly from the starting template, or the model was energetically clearly disfavored, i.e. had substantial clashes (Table 5).

Evaluation of the Models Using Experimental Data

The models obtained by using different templates, sequence-sequence alignments, modeling of loops, and relaxation of structures were evaluated using the distance constraints imposed on the laminin LN domains by “natural” cross-links (disulfide bonds) and “chemically introduced” cross-links. The distance constraints used for verifying the models are summarized in Table 3. The identification of a disulfide bond implies that the C^A atoms of the cysteines involved are within a distance of ca. 7 Å. Every C^A - C^A distance of a disulfide

bond, in which the cysteines are separated by at least two amino acids in sequence can be used as structural constraint. In case the cysteines are separated by less than two amino acids, the distance is below 9 Å and the distance constraint is always fulfilled. A similar approximation can be made for the C^A-C^A distances of amino acids that are cross-linked with BS³: If the C^A atoms are at least seven positions apart from each other, these cross-links can be used as useful distance constraints. Thus, seven out of twelve and five out of seven constraints have been applied as restraint for laminin β₁ and γ₁, respectively. Nevertheless, as already described by Alexander *et al.* 41) the information content of the distance constraints depend on the sequence separation [aa] / Euclidian separation [Å] ratio.

The structures of laminin β₁ and γ₁ LN domains calculated by Rosetta were compared with the experimentally obtained distance constraints (Table 3, Supplementary Table 3). All structures were grouped based on the number of fulfilled distance constraints. A distance constraint was considered to be fulfilled if the C^A-C^A distance of a cross-link was below 30 Å and the C^A-C^A distance of a disulfide bridge below 13 Å. Both filter criteria are somewhat relaxed to the distances listed above in order to account for inaccuracies in the comparative models (Table 5). Only approximately 1% of all models fulfilled all distance constraints highlighting the discriminative power of even few experimental distance measurements (Table 5).

All models fulfilling all distance constraints exhibited the same topology (more than 100 out of 240 amino acids could be superimposed with an RMSD < 6 Å), however, shifts in the sequence structure alignment of up to 40 amino acids occur. In a second step, a composite scoring function described in *Materials and Methods* integrating the deviation of the experimental constraints from the theoretical maximum distance as well as the Rosetta energy and compactness of the structure were used to evaluate the quality of the models and create a ranking (Supplementary Table 4). In the best scoring models of the LN domains of laminin β₁ and γ₁ all of the distance constraints deviate from the theoretical distance by less than 5 Å (Figure 5).

For laminin β₁, three structures (Table 5) fulfilled all seven distance constraints and exhibited structural similarity, i.e., more than 110 C^A atoms are superimposed with an RMSD below 4 Å. For laminin γ₁, five distance constraints were fulfilled by 12 models (Table 5). Nine structures could be superimposed in more than 120 C^A atoms with an RMSD below 4 Å. For the remaining three models more than 50 C^A atoms could be superimposed with an RMSD below 4 Å.

Leave-One-Out Cross-Validation and Evaluation with Structure Quality

To assess the quality of laminin β₁ and γ₁, all models fulfilling 6 out of 7 or 4 out of 5 distance constraints, respectively, were pairwise compared to the best solution structures for laminin β₁ and γ₁ LN domains (Figure 5). A test of the precision of the modeling efforts yielded 17 out of 33 (laminin β₁) and 141 out of 293 models (laminin γ₁) with a high structural similarity, i.e. more than 80 C^A atom are superimposed with a RMSD below 4 Å. Accuracy was assessed by testing whether or not the left out distance constrained is fulfilled by the models. On the basis of the LOO cross-validation, the uncertainty of the predicted structures was estimated. For the models of laminin β₁ and γ₁ LN domains, which were

obtained with 6 out of 7 (4 out of 5) distance constraints, 78% (61%) of all models fulfilled the missing distance constraint. If the same models were validated by their structure homology to the best model, 88% (64%) of the models exhibited the correct fold (Table 7). In conclusion, 78% and 61% of the *leave-one-out* structures fulfill the left-out distance constraint.

In addition, the programs VADAR and MolProbity were used for a Rosetta-independent measurement of structure quality (Supplementary Table 4). The quality criteria defined by the two programs are all in a reasonable range for the models that fulfill all distance constraints. As only minor differences in quality scores were observed between the models, these criteria were not used for further ranking the models.

Influence of Templates

To analyze the influence of the template on the final model best structures were compared for the 14 templates (Table 4). For laminin γ_1 , for all 14 templates structures were obtained, which fulfilled at least 4 out of 5 distance constraints and were energetically favored as well as structurally compact. From these, five structures displayed a fold identical to the best overall model. For laminin β_1 , only nine templates yielded models, which fulfilled six out of seven distance constraints. All displayed the identical fold as in the overall best structure (Supplementary Table 2).

Influence of Experimental Distance Constraints

A statistics on the effectiveness of each constraint revealed that all distance constraints reduced conformational space. However, the fraction of models excluded ranged from 0.1% to 80.7% (Table 7). As expected, the effectiveness of an experimental constraint increases with increasing sequence separation and decreasing limit for the Euclidean distance.

Implications of Our Models for Laminin Function

Laminins containing the γ_1 chain are essential for mammalian development, as shown by the fact that mice lacking expression of this chain die at day 5.5 of embryonal development (45). Laminins assemble into a macromolecular network by interactions of their LN domain containing *N*-terminal short arms. This assembly renders them immobile in the basement membrane and allow their C-terminal domains to attach cells onto this structure by interactions with cell surface integrin or dystroglycan receptors. The crucial role of laminin LN domains in laminin network formation has led to intensive attempts to determine their structure by crystallographic techniques. However, it turned out to be extremely difficult to obtain well organized crystals for LN domains from laminins or the homologous netrins and for this reason no LN domain crystal structures have been published so far. The lack of detailed structural information has hampered the study of the crucial LN domain interactions.

The modeled structures that we present herein open new possibilities to identify binding surfaces on LN domains as well as ligand structures. Based on the models site-directed mutagenesis can be performed on surface exposed amino acid residues and the consequences of these mutations can be studied in binding assays for laminin short arm self-interactions

(3). Further chemical cross-linking experiments can be performed on interacting laminin short arms to identify the domains within laminin, to which the LN domains bind. Such studies will provide a structural model of self-interacting laminin molecules, which will significantly add to our understanding of basement membrane structure. As mutations within LN domains that affect their interactions can cause severe retinal phenotypes (46) novel knowledge on pathogenic mechanisms may come about.

CONCLUSIONS

With our approach integrating chemical cross-linking, mass spectrometry, and computational modeling we were able to compute the first experimentally validated low-resolution structures of laminin *N*-terminal domains. After the identification of 14 potentially suited template structures for each laminin β_1 and γ_1 LN domain, several hundreds of models were created with the modeling program Rosetta. The models were filtered and ranked by both the correlation with experimentally derived distance constraints from natural cross-links, i.e., disulfide bonds, and chemical cross-links created by reaction with two amine-reactive cross-linkers and parameter for model quality i.e., energy and solvent accessible surface area.

The questions how well-defined the models are, how many, and which kind of distance constraints are needed for creating a good model, and how experimental data and template structure influence the model were addressed by leave-one-out cross-validation combined with a structure-structure alignment as well as by Rosetta-independent structure evaluation programs.

Although none of the techniques employed herein is completely new, we argue that the presented protocol is novel as it combines ten computational methods with a complex experimental setup integrating chemical cross-linking, isotope labeling, and HPLC/MS analysis. As such, we consider our approach to be a valid alternative for deriving structural models of proteins, which are not amenable to the high-resolution methods for protein 3D structure analysis, such as NMR spectroscopy or X-ray crystallography.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants SI 867/7-1 and SM 65/1-3 from the DFG (Deutsche Forschungsgemeinschaft) and by the Köln Fortune program of the Medical Faculty of the University of Cologne. The authors thank Dr. Tibor Kohajda for programming the software tool *IsoFind* and Dr. Christian Ihling for discussions and valuable advice on MALDI-TOF/TOF-MS and ESI-FTICR-MS (Apex II) measurements. Karl Mechtler and Christoph Stingl, IMP Vienna, are acknowledged for LTQ-FT-MS measurements. SK acknowledges support from the DFG-funded Graduiertenkolleg 1026 “Conformational Transitions in Macromolecular Interactions” at the Martin Luther University Halle-Wittenberg.

ABBREVIATIONS

BS ² G	<i>Bis</i> (sulfosuccinimidyl)glutarate
BS ³	<i>Bis</i> (sulfosuccinimidyl)suberate

DHB	2,5-Dihydroxybenzoic acid
DTT	Dithiothreitol
ESI-FTICR-MS	Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry
FA	Formic acid
HCCA	α -Cyano-4-hydroxy cinnamic acid
HPLC	High-performance liquid chromatography
LHRH	Luteinizing hormone releasing hormone
LIFT	laser-induced fragmentation technique
LN	laminin <i>N</i> -terminal
LOO	Leave-one-out
MALDI-TOF-MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
NHS	<i>N</i> -hydroxysuccinimide
PDB	Protein data bank
RMSD	Root mean square deviation (calculated for the backbone coordinates)
SASA	Solvent accessible surface area
S/N	Signal-to-noise ratio
TFA	Trifluoroacetic acid
TIC	Total ion current

REFERENCES

1. Tunggal P, Smyth N, Paulsson M, Ott M-C. Microscopy Research and Technique. 2000; 51:214–227. [PubMed: 11054872]
2. Aumailley M, Bruckner-Tuderman L, Carter WG, Deutzmann R, Edgar D, Ekblom P, Engel J, Engvall E, Hohenester E, Jones JCR, Kleinman HK, Marinkovich MP, Martin GR, Mayer U, Meneguzzi G, Miner JH, Miyazaki K, Patarroyo M, Paulsson M, Quaranta V, Sanes JR, Sasaki T, Sekiguchi K, Sorokin LM, Talts JF, Tryggvason K, Uitto J, Virtanen I, von der Mark K, Wewer UM, Yamada Y, Yurchenco PD. Matrix Biology. 2005; 24:326–332. [PubMed: 15979864]
3. Odenthal U, Haehn S, Tunggal P, Merkl B, Schomburg D, Frie C, Paulsson M, Smyth N. Journal of Biological Chemistry. 2004; 279:44504–44512. [PubMed: 15310759]
4. Yurchenco PD, Tsilibary EC, Charonis AS, Furthmayr H. Journal of Biological Chemistry. 1985; 260:7636–7644. [PubMed: 3997891]
5. McKee KK, Harrison D, Capizzi S, Yurchenco PD. Journal of Biological Chemistry. 2007; 282:21437–21447. [PubMed: 17517882]

6. Schneiders FI, Maertens B, Bose K, Li Y, Brunken WJ, Paulsson M, Smyth N, Koch M. Journal of Biological Chemistry. 2007; 282:23750–23758. [PubMed: 17588941]
7. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:5802–5806. [PubMed: 10811876]
8. Back JW, Jong Ld, Muijsers AO, Koster CGd. Journal of Molecular Biology. 2003; 331:303–313. [PubMed: 12888339]
9. Sinz A. Journal of Mass Spectrometry. 2003; 38:1225–1237. [PubMed: 14696200]
10. Sinz A. Mass Spectrometry Reviews. 2006; 25:663–682. [PubMed: 16477643]
11. Hermanson, G., editor. Bioconjugate Techniques. 2nd. Amsterdam, NL: Elsevier; 2008.
12. Kalkhof S, Haehn S, Ihling C, Paulsson M, Smyth N, Sinz A. Rapid Communications in Mass Spectrometry. 2008; 22:1933–1940. [PubMed: 18491288]
13. Kalkhof S, Ihling C, Mechtler K, Sinz A. Analytical Chemistry. 2004; 77:495–503.
14. Schmidt A, Kalkhof S, Ihling C, Cooper D, Sinz A. Eur J Mass Spectrom (Chichester, Eng). 2005; 11:525–534.
15. Ihling C, Schmidt A, Kalkhof S, Schulz DM, Stingl C, Mechtler K, Haack M, Beck-Sickinger AG, Cooper DMF, Sinz A. Journal of the American Society for Mass Spectrometry. 2006; 17:1100–1113. [PubMed: 16750914]
16. Simons KT, Kooperberg C, Huang E, Baker D. Journal of Molecular Biology. 1997; 268:209–225. [PubMed: 9149153]
17. Bradley P, Malmström L, Qian B, Schonbrun J, Chivian D, Kim DE, Meiler J, Misura KMS, Baker D. Proteins: Structure, Function, and Bioinformatics. 2005; 61:128–134.
18. Bowers PM, Strauss CEM, Baker D. Journal of Biomolecular NMR. 2000; 18:311–318. [PubMed: 11200525]
19. Meiler J, Baker D. Journal of Magnetic Resonance. 2005; 173:310–316. [PubMed: 15780923]
20. Alexander N, Al-Mestarihi A, Bortolus M, McHaourab H, Meiler J. 2008; 16:181–195.
21. Kalkhof S, Sinz A. Anal Bioanal Chem. 2008; 392:305–312. [PubMed: 18724398]
22. Laemli UK. Nature. 1970; 227:680–685. [PubMed: 5432063]
23. Schulz DM, Kalkhof S, Schmidt A, Ihling C, Stingl C, Mechtler K, Zschörnig O, Sinz A. Proteins: Structure, Function, and Bioinformatics. 2007; 69:254–269.
24. Mädler S, Bich C, Touboul D, Zenobi R. Journal of Mass Spectrometry. 2009; 44:694–706. [PubMed: 19132714]
25. Ginalski K, Elofsson A, Fischer D, Rychlewski L. Bioinformatics. 2003; 19:1015–1018. [PubMed: 12761065]
26. Huang X, Miller W. Adv. Appl. Math. 1991; 12:337–357.
27. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D. Nucl. Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
28. Kelley LA, Sternberg MJE. Nat. Protocols. 2009; 4:363–371. [PubMed: 19247286]
29. Juan D, Graña O, Pazos F, Fariselli P, Casadio R, Valencia A. Proteins: Structure, Function, and Genetics. 2003; 50:600–608.
30. Torda AE, Procter JB, Huber T. Nucl. Acids Res. 2004; 32:W532–W535. [PubMed: 15215443]
31. Soding J. Bioinformatics. 2005; 21:951–960. [PubMed: 15531603]
32. Teodorescu O, Galor T, Pillardy J, Elber R. Proteins: Structure, Function, and Bioinformatics. 2004; 54:41–48.
33. Ortiz AR, Strauss CEM, Olmea O. Protein Science. 2002; 11:2606–2621. [PubMed: 12381844]
34. Dong E, Smith J, Heinze S, Alexander N, Meiler J. Gene. 2008; 422:41–46. [PubMed: 18601985]
35. Meiler J, Prompers JJ, Peti W, Griesinger C, Bruschweiler R. Journal of the American Chemical Society. 2001; 123:6098–6107. [PubMed: 11414844]
36. Jones DT. Journal of Molecular Biology. 1999; 292:195–202. [PubMed: 10493868]
37. Karplus K, Barrett C, Cline M, Diekhans M, Grate L, Hughey R. Proteins: Structure, Function, and Genetics. 1999; 37:121–125.
38. Chivian D, Baker D. Nucl. Acids Res. 2006; 34:e112. [PubMed: 16971460]

39. Rohl CA, Strauss CEM, Chivian D, Baker D. Proteins: Structure, Function, and Bioinformatics. 2004; 55:656–677.
40. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. Nucl. Acids Res. 2003; 31:3316–3319. [PubMed: 12824316]
41. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS, Richardson DC. Nucl. Acids Res. 2007; 35:W375–W383. [PubMed: 17452350]
42. Schnaible V, Wefing S, Resemann A, Suckau D, Buecker A, Wolf-Kuemmeth S, Hoffmann D. Analytical Chemistry. 2002; 74:4980–4988. [PubMed: 12380820]
43. Rost B, O'Donoghue S. Comput. Appl. Biosci. 1997; 13:345–356. [PubMed: 9283749]
44. Kuhlman B, Baker D. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:10383–10388. [PubMed: 10984534]
45. Edwards MM, Mammadova-Bach E, Alpy F, Klein A, Hicks WL, Roux M, Simon-Assmann P, Smith RS, Orend G, Wu J, Peachey NS, Naggert JK, Lefevbre O, Nishina PM. J. Biol. Chem. 2010; 285:7697–7711. [PubMed: 20048158]
46. Smyth NR, Vatansever HS, Murray P, Meyer M, Frie C, Paulsson M, Edgar D. J. Cell Biol. 1999; 144:151–160. [PubMed: 9885251]
47. Macedo-Ribeiro S, Bode W, Huber R, Quinn-Allen MA, Kim SW, Ortel TL, Bourenkov GP, Bartunik HD, Stubbs MT, Kane WH, Fuentes-Prior P. Nature. 1999; 402:434–439. [PubMed: 10586886]
48. Pratt KP, Shen BW, Takeshima K, Davie EW, Fujikawa K, Stoddard BL. Nature. 1999; 402:439–442. [PubMed: 10586887]
49. Gaskell A, Crennell S, Taylor G. Structure. 1995; 3:1197–1205. [PubMed: 8591030]
50. Ito N, Phillips SEV, Stevens C, Ogel ZB, McPherson MJ, Keen JN, Yadav KDS, Knowles PF. Nature. 1991; 350:87–90. [PubMed: 2002850]
51. Au SWN, Leng X, Harper JW, Barford D. Journal of Molecular Biology. 2002; 316:955–968. [PubMed: 11884135]
52. Boraston AB, Nurizzo D, Notenboom V, Ducros V, Rose DR, Kilburn DG, Davies GJ. Journal of Molecular Biology. 2002; 319:1143–1156. [PubMed: 12079353]
53. Wendt KS, Vodermaier HC, Jacob U, Gieffers C, Gmachl M, Peters J-M, Huber R, Sondermann P. Nature Structural Biology. 2001; 8:784–788. [PubMed: 11524682]
54. Bianchet MA, Odom EW, Vasta GR, Amzel LM. Nature Structural Biology. 2002; 9:628–634. [PubMed: 12091873]
55. Firbank SJ, Rogers MS, Wilmot CM, Dooley DM, Halcrow MA, Knowles PF, McPherson MJ, Phillips SEV. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:12932–12937. [PubMed: 11698678]
56. Lee CC, Kreusch A, McMullan D, Ng K, Spraggon G. Structure. 2003; 11:99–108. [PubMed: 12517344]
57. Adams TE, Hockin MF, Mann KG, Everse SJ. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:8918–8923. [PubMed: 15184653]
58. Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma L-C, Acton TB, Hunt JF, Montelione GT, Baker D, Kennedy MA. Proteins: Structure, Function, and Bioinformatics. 2009; 75:147–167.
59. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Nat Struct Mol Biol. 1999; 6:884–893.
60. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. Nat Struct Mol Biol. 1996; 3:638–648.
61. JointCenterForStructuralGenomics. PDB entry 1VPQ To be published.
62. Guo H-C, Jardetzky TS, Garrett TPJ, Lane WS, Strominger JL, Wiley DC. Nature. 1992; 360:364–366. [PubMed: 1448153]
63. Cardoso RMF, Silva CHTP, Ulian de Araujo AP, Tanaka T, Tanaka M, Garratt RC. Acta Crystallographica Section D. 2004; 60:1569–1578.
64. Setayesh FR, DeCorte BL, Horton P, Harris CM, Harris TM, Stone MP. Chemical Research in Toxicology. 1998; 11:766–777. [PubMed: 9671539]

65. Momma K, Mishima Y, Hashimoto W, Mikami B, Murata K. *Biochemistry*. 2005; 44:5053–5064. [PubMed: 15794643]
66. JointCenterForStructuralGenomics. PDB entry 1VKO To be published.
67. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang B-C, Wilson KP. *Journal of Biological Chemistry*. 2004; 279:31655–31663. [PubMed: 15123710]
68. Skelton NJ, Quan C, Reilly D, Lowman H. *Structure*. 1999; 7:157–168. [PubMed: 10368283]
69. Li H, Tomizawa T, Tochio N, Muto Y, Koshiba S, Inoue M, Kigawa T, Yokoyama S. PDB entry 2COC To be published.
70. Udomsinprasert R, Pongjaroenkit S, Wongsantichon J, Oakley AJ, Prapanthadara L-a, Wilce MCJ, Ketterman AJ. *Biochem. J*. 2005; 388:763–771. [PubMed: 15717864]
71. Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K. *Journal of Molecular Biology*. 2000; 300:889–901. [PubMed: 10891276]
72. López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-i, Kaltner H, Gabius H-J, Romero A. *Journal of Molecular Biology*. 2004; 343:957–970. [PubMed: 15476813]
73. Katsuya Y, Mezaki Y, Kubota M, Matsuura Y. *Journal of Molecular Biology*. 1998; 281:885–897. [PubMed: 9719642]
74. Koch AW, Farooq A, Shan W, Zeng L, Colman DR, Zhou M-M. *Structure*. 2004; 12:793–805. [PubMed: 15130472]
75. Abad MC, Arni RK, Grella DK, Castellino FJ, Tulinsky A, Geiger JH. *Journal of Molecular Biology*. 2002; 318:1009–1017. [PubMed: 12054798]
76. Llera AS, Viedma F, Sánchez-Madrid F, Tormo J. *Journal of Biological Chemistry*. 2001; 276:7312–7319. [PubMed: 11036086]
77. Baeyens KJ, De Bondt HL, Raeymaekers A, Fiers W, De Ranter CJ. *Acta Crystallographica Section D*. 1999; 55:772–778.
78. Lubkowski J, Hennecke F, Pluckthun A, Wlodawer A. *Nat Struct Mol Biol*. 1998; 5:140–147.
79. Tegoni M, Spinelli S, Verhoeyen M, Davis P, Cambillau C. *Journal of Molecular Biology*. 1999; 289:1375–1385. [PubMed: 10373373]
80. Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS, Hinck AP. *Nat Struct Mol Biol*. 2002; 9:203–208.
81. Stams T, Chen Y, Boriack-Sjodin PA, Hurt JD, Liao J, May JA, Dean T, Laipis P, Silverman DN, Christianson DW. *Protein Sci*. 1998; 7:556–563. [PubMed: 9541386]
82. Lamb AL, Wernimont AK, Pufahl RA, O'Halloran TV, Rosenzweig AC. *Biochemistry*. 2000; 39:1589–1595. [PubMed: 10677207]
83. Whittingtons DA, Grubb JH, Waheed A, Shah GN, Sly WS, Christianson DW. *Journal of Biological Chemistry*. 2004; 279:7223–7228. [PubMed: 14660577]
84. Dayan SM, Van Donkelaar A, Kortt AA. *Journal of Biological Chemistry*. 1987; 262:10287–10289. [PubMed: 3611061]
85. Watson JN, Newstead S, Dookhun V, Taylor G, Bennet AJ. *FEBS Letters*. 2004; 577:265–269. [PubMed: 15527797]
86. Zhang Y. *BMC Bioinformatics*. 2008; 9:40. [PubMed: 18215316]
87. Zhang Y. *Proteins: Structure, Function, and Bioinformatics*. 2009; 77:100–113.
88. Soding J, Biegert A, Lupas AN. *Nucl. Acids Res*. 2005; 33:W244–W248. [PubMed: 15980461]
89. Wu S, Zhang Y. *Nucl. Acids Res*. 2007; 35:3375–3382. [PubMed: 17478507]
90. Fiser A, Do RKG, Scaronali A. *Protein Science*. 2000; 9:1753–1773. [PubMed: 11045621]

Figure 1.

General workflow of the computational modeling procedure, which includes as main steps (A) detection and identification of templates, (B) sequence-sequence-alignment between target and template sequences, (C) modeling of the LN domains of laminin β_1 and γ_1 , (D) experimentally guided scoring and selection of the best models, and (E) validation. Rectangles: intermediate results; trapeze: verifications and controls; ovals: calculation and predictions steps; parallelograms: input data.

Figure 2.

Dendrogram of the selected and discarded template structures. The clustering was performed based on the structural homology (normalized Z-Score calculated with MAMMOTH). The templates exhibiting a galactose-binding domain-like fold are highlighted in red.

(A) *Laminin β_1*

1 **APLVQ**EPF**S** YGCAEGSCYP ATGDL**LIGRA** QKLSVTSTCG LHKPEPYCIV
 51 **SHLQ**EDKKCF ICDSRPYHE TLNPDSHLIE NVVTT**FAPNR** LKIWWQSENG
 101 **VENVTI**QLDL EA**EFHF**THLI MTFK**TRPAA** MLIERS**SDFG** KTWGVYRYFA
 151 YDCESS**FPGI** STGPMKKVDD IICDSRYSDI EPSTEGEVIF RALDP**AFKIE**
 201 DPYSPRIQNL LKITNLRIKF VKLHTLG**DNL** LDSRMEIREK YYYAVYDMVV
 251 RGNCFCY**GHA** SECAPVDGVN EEVEGMV**HGH** CMC**RHNTKGL** NCELCMDFYH
 301 DLPWRPAEGR NSNACKKCNC NEHSSSCHFD MAVFLAT**GNV** SGGVCDNCQH
 351 NTMGRNCEQC KPFYFQHPER DIRDPNLCEP CTCDPAGSEN GGICDGYTDF
 401 SVGLIAGQCR CKLHVEGERC DVCKEGFYDL SAEDPYGCKS CACNPLGTIP
 451 GGNPCDSETG YCYCKRLVTG QRC**DQCLPQH** WGLSNDLDGC RAA**AAAAHHHHHH**

(B) *Laminin γ_1*

1 **APLVH**HHHHH ALVAMDECAD EGGR**PQRCMP** EFVNAA**FNVT** VVATNTCGTP
 51 PEEYCVQ**TGV** TGVTKSCHLC DAGQ**QHLQH**G AAFLTDYNNQ ADTTWWQSQT
 101 MLAGVQYPNS **INL**TLHLGKA FDITYVRLKF HTSRPESFAI YKRTREDGPW
 151 IPYQYYSGSC ENTYSKANRG FIRTGGDE**QQ** ALCTDEFSDI SPLTGGN**VAF**
 201 STLEGRPSAY NFDNSPVLQE WVTATDIRVT LNRLNTFGDD VFNEPKVLKS
 251 YYYAISDFAV GGRCKCNGHA SECVKNEFDK LMCNCKHNTY GVDCEKCLPF
 301 FNDRPWRRAT AESASESLPC DCNGRSQECY FDP**ELYRSTG** HGGHCTNCRD
 351 NTDGAKCERC RENFFRLGNT EACSPCHCSP VGSLSTQ**CDS** YGRCSCKPGV
 401 MGDKCDRCQP GFHSLTEAGC RPCSCDPSGS TDECNVETGR CVCKDNVEGF
 451 NCERCKPGFF NLESSNP**KGC** TP

Figure 3.

Amino acid sequences of recombinant (A) laminin β_1 and (B) laminin γ_1 fragments comprised of one LN domain and LE domains 1-4. Amino acids that were detected during peptide mass fingerprint and peptide fragment fingerprint analyses are shown in bold. Putative glycosylation sites are shown in italics and underlined. The tags that were added to the laminin *N*-terminal constructs are printed in italics. LN domains are shaded.

Figure 4.

MALDI-TOF/TOF-MS/MS analysis of the cross-linking product with BS³ of laminin β_1 LN domain (amino acids 194–200 connected with amino acids 207–217; $[M+H]^+_{\text{exp}}$ 2282.322), lysines at positions 198 and 212 were found to be connected (marked with asterisks). Upper panel: the identified fragment ions are shown in bold; lower panel: tandem MS (MS/MS); the insets show the y6 fragment ion and the precursor ion, both exhibiting the characteristic D_0/D_4 isotope pattern.

1 C13*-C18; 2 C39-C48; 3 C59-C62; 4 C152-C173; 5 K32-K43; 6 K141-K198; 7 K198-K212; 8 K166-K167; 9 K198-Y203/S204 ; 10 K212-T214; 11 K219-K222; 12 K222-T225

B

1 C18*-C28; 2 C47-C55; 3 C67-C70; 4 C160-C183; 5 K129-K142; 6 S208-Y210; 7 K246-K249

Figure 5.

The lowest scoring Rosetta structures for (A) laminin β_1 and (B) γ_1 LN domains that fulfill all experimental distance constraints. The amino acids involved in the constraints are indicated, structures are visualized by PyMOL version 0.99rc6). Cysteines at positions 13 (laminin β_1) and 18 (laminin γ_1), which are not part of laminin N-terminal domains, are marked with asterisks.

Table 1

Intramolecular cross-linked products of isolated laminin β_1 and γ_1 LN domains using BS³ as cross-linker; all cross-linked products were verified by MS/MS data. The cross-linked amino acids are printed in *italics* and are underlined. The cross-linked product presented in Figure 4 is highlighted.

Protein	MH ⁺ _{exp}	MH ⁺ _{theo}	Sequence	Amino acids	m [ppm]	MS/MS	D ₀ /D ₄	Digestion
β_1	1293.794 1297.821	1293.793 1297.818	218–227	R/I <i>K</i> FV <i>K</i> LH <i>T</i> L <i>G</i> /D	1 2	b5–9 y4–5	1.9	<i>In-solution</i>
β_1	1293.786 1297.810	1293.793 1297.818	218–227	R/I <i>K</i> FV <i>K</i> LH <i>T</i> L <i>G</i> /D	5 6	b5–9 y3–5	1.8	<i>In-gel</i>
β_1	1672.824 1676.846	1672.822 1676.847	194–206	L.DPAF <i>K</i> IEDP <i>Y</i> SPR.I	1 1	y1,y12 b3	1.4	<i>In-solution</i>
β_1	1672.804 1676.830	1672.822 1676.847	194–206	L.DPAF <i>K</i> IEDP <i>Y</i> SPR.I	11 10	y1–3,y12; b2,b4	1.4	<i>In-gel</i>
β_1	1856.951 1860.976	1856.943 1860.968	192–206	R.ALDPAF <i>K</i> IEDP <i>Y</i> SPR.I	4 4	y1–3;y12, b2	1.4	<i>In-solution</i>
β_1	2282.310 2286.329	2282.334 2286.359	194–200 + 207–217	L.DPAF <i>K</i> IE.D + R.IQNLL <i>K</i> ITNLR.I	11 13	1: b3,y6 2: y1,2;b2,3	1.6	<i>In-solution</i>
β_1	2466.415 2470.444	2466.433 2470.458	192–200 +207–217	R.ALDPAF <i>K</i> IE.D + R.IQNLL <i>K</i> ITNLR.I	7 6	2: b1,3;y1–4	1.5	<i>In-gel</i>
β_1	2466.455 2470.479	2466.433 2470.458	192–200 +207–217	R.ALDPAF <i>K</i> IE.D + R.IQNLL <i>K</i> ITNLR.	9 9	1:b2,3 2:b2	1.6	<i>In-solution</i>
γ_1	2118.135 2122.156	2118.149 2122.174	128–143	R.L <i>K</i> FHTSRPESFAIY <i>K</i> R.T	7 8	y1,2,5–7,9 b7,14,15	1.8	<i>In-solution</i>
γ_1	2118.145 2122.166	2118.149 2122.174	128–143	R.L <i>K</i> FHTSRPESFAIY <i>K</i> R.T	2 4	y1, b15	1.7	<i>In-gel</i>
γ_1	2651.284 2655.307	2651.299 2655.324	189 – 212	S.DISPLTGGNVAFTSTLEGRP <i>SA</i> <i>Y</i> NF.D	6 6	y1,2,6– 12,14,18,21,23 b2–4,6,9–11,18	1.3	<i>In-gel</i>
γ_1	2651.286 2655.301	2651.299 2655.324	189 – 212	S.DISPLTGGNVAFTSTLEGRP <i>SA</i> <i>Y</i> NF.D	5 9	y2,6,23; b2,3,6	1.5	<i>In-solution</i>

Table 2

Intramolecular cross-linked products of cross-linking reaction mixtures containing both laminin β_1 and γ_1 LN domains; the cross-linker (BS²G or BS³) is indicated. All cross-linked products were verified by MS/MS data. The cross-linked amino acids are printed in italics and are underlined.

Laminin β_1		Sequence	Amino acids and cross-linker	m [ppm]	MS/MS	D ₀ /D ₄	Digestion
MH ⁺ _{exp}	MH ⁺ _{theo}						
1293.787 1297.812	1293.793 1297.818	218–227	R. <u>I</u> <u>K</u> FV <u>K</u> LHTLG.D + BS ³	5 5	y4–5; b5–9	1.3	<i>In-solution</i>
1293.821 1297.841	1293.793 1297.818	218–227	R. <u>I</u> <u>K</u> FV <u>K</u> LHTLG.D + BS ³	22 18	y3–5; b5–9	1.8	<i>In-gel</i>
1672.823 1676.847	1672.822 1676.847	194–206	L.DPAF <u>K</u> IEDP <u>Y</u> <u>S</u> PR.I + BS ³	1 0	b3, y1–3,12	1.4	<i>In-solution</i>
1672.833 1676.860	1672.822 1676.847	194–206	L.DPAF <u>K</u> IEDP <u>Y</u> <u>S</u> PR.I + BS ³	7 8	b3; y1–3,9,12	3.0	<i>In-gel</i>
2282.322 2286.342	2282.312 2286.337	194–200+ 207–217	L.DPAF <u>K</u> IE.D + R.IQNLL <u>K</u> ITNLR.I + BS ³	4 2	1 : b3 2 : b2–4; y1–5,	1.8	<i>In-gel</i>
2466.430 2470.454	2466.433 2470.458	192–200+ 207–217	R.IQNLL <u>K</u> ITNLR.I + R.ALDPAF <u>K</u> IE.D + BS ³	1 2	-	1.6	<i>In-gel</i>
1421.842 1425.867	1421.847 1425.869	207–217	R.IQNLL <u>K</u> ITNLR.I + BS ² G	4 1	y1–5, b1–5	0.8	<i>In-solution</i>
1630.800 1634.793	1630.775 1634.800	194–206	L.DPAF <u>K</u> IEDP <u>Y</u> <u>S</u> PR.I + BS ² G	15 4	y1,2,12	0.6	<i>In-gel</i>
1836.901 1840.924	1836.865 1840.890	176–191	R. <u>Y</u> SDIEP <u>S</u> TEGEVIFR.A amidated(C-term); + BS ² G	20 18	y1–8	0.7	<i>In-solution</i>
1836.921 1840.944	1836.865 1840.890	176–191	R. <u>Y</u> SDIEP <u>S</u> TEGEVIFR.A amidated (C-term); + BS ² G	30 29	y1–4	0.6	<i>In-gel</i>
1935.877 1939.900	1935.883 1939.908	152–168	Y.DCESSFPGIS TGP <u>M</u> <u>K</u> <u>K</u> V.D + BS ² G	3 4	y2–3,5–8,10–13, b4, b9, b6, b4	0.7	<i>In-gel</i>
2240.265 2244.286	2240.265 2244.290	194–200+ 207–217	L.DPAF <u>K</u> IE.D + R.IQNLL <u>K</u> ITNLR.I + BS ² G	0 2	b3, y1,4,17	0.7	<i>In-solution</i>
2240.260 2244.285	2240.265 2244.290	194–200+ 207–217	L.DPAF <u>K</u> IE.D + R.IQNLL <u>K</u> ITNLR.I + BS ² G	2 2	1 : y1–5 ; b2–5 2 : b2–4	0.7	<i>In-gel</i>
2501.238 2505.257	2501.235 2505.260	136–147+ 192–200	R.SSDFG <u>K</u> TWGVYR.Y + R.ALDPAF <u>K</u> IE.D + BS ² G	1 1	1 : b2, y1,2,5	0.5	<i>In-solution</i>
2501.216 2505.239	2501.235 2505.260	136–147+ 192–200	R.SSDFG <u>K</u> TWGVYR.Y + R.ALDPAF <u>K</u> IE.D + BS ² G	8 8	A : b2,3,5; y1,4–6 b : b3	0.6	<i>In-gel</i>

Laminin β_1						
MH ⁺ _{exp}	MH ⁺ _{theo}	Sequence	Amino acids and cross-linker	m [ppm]	MS/MS	D ₀ /D ₄
2552.132 2556.149	2552.135 2556.160	152–173	Y.DCESSFPGISTGPMKKVDDIIC.D + BS ² G	1 4	b6,9,10,17,21 y10,11,13,17	0.5
2552.021 2556.044	2552.135 2556.160	152–173	Y.DCESSFPGISTGPMKKVDDIIC.D + BS ² G	45 45	-	0.6
2751.278 2755.315	2751.312 2755.337	33–55	K.LSVTSTCGLHKPEPYCIVSHLQE.D + BS ² G	12 8	y3-y6,10–14 b2,4,10,11	0.5
2751.272 2755.312	2751.312 2755.337	33–55	K.LSVTSTCGLHKPEPYCIVSHLQE.D + BS ² G	15 9	b2,4,10 y5-6,10,12	0.6
3078.499 3082.529	3078.502 3082.527	30–55	R.AQKL ⁺ SVTSTCGLHKKPEPYCIVSHLQE.D + BS ² G	1 1	b15,17 y4,5,10,12	0.3

Laminin γ_1						
MH ⁺ _{exp}	MH ⁺ _{theo}	Sequence	Amino acids and cross-linker	m [ppm]	D ₀ /D ₄	MS/MS
2118.151 2122.176	2118.149 2122.174	128–143	R.LKKFH7SRPESFAIYKKR.T + BS ³	1 1	1.3	b7,9,15 y1,7,9
2118.172 2122.170	2118.149 2122.174	128–143	R.LKKFH7SRPESFAIYKKR.T + BS ³	11 2	0.8	(y1,b15)
2076.080 2080.125	2076.103 2080.128	128–143	R.LKKFH7SRPESFAIYKKR.T + BS ² G	11 1	0.8	b1,15 y1,6,7,10
2852.190 2856.198	2852.226 2856.251	147–169	E.DGPWIPYQYYSGSCENTYSKANR.G + BS ² G	13 19	0.4	y1-3; b3– b5, b8, b16
2949.432 2953.455	2949.431 2953.456	239–263	G.DDVFNEPKVLKSYYYAISDFAVGGRC + BS ² G	0 0	0.7	b18,b20, y1,3,19

Table 3

Distance constraints in laminin β_1 and γ_1 LN domains serving as basis for computational modeling (filtering criteria for C^A-C^A distances: disulfide bonds: 13 Å (favored distance 8 Å), cross-links: 30 Å (favored distance 24 Å); shaded areas: these constraints could not be employed as a direct structural filter because of the spatial proximity in the primary structure of laminin LN domains.

Laminin β_1		Laminin γ_1	
Connected amino acids	Connection Type	Connected amino acids	Connection Type
C13-C18	Disulfide	C18-C28	Disulfide
K32-K43	BS ² G	C47 -C55	Disulfide
C39-C48	Disulfide	C67 -C70	Disulfide
C59-C62	Disulfide	K129-K142	BS ² G, BS ³
K141-K198	BS ² G	C160-C183	Disulfide
C153-C173	Disulfide	S208-Y210	BS ³
K166-K167	BS ² G	K246-K249	BS ³
K198-Y203/S204	BS ² G, BS ³		
K198-K212	BS ² G, BS ³		
K212-T214	BS ² G		
K219-K222	BS ³		
K222-T225	BS ² G		

Table 4

Results of the template search; 14 structures were found using the programs 3D Jury, Superfamily – “Superfam”, Phyre, Libellula “Lib”, HHPred; the programs Wurst and Loop yielded only false positive template structures. The six templates found with 3D-Jury are printed in bold.

No	Protein Name	PDB entry	Method	Template Score for Laminin β 1: γ 1	SCOP Fold name	SCOP Fold classification code	CATH Fold name	CATH Fold classification code	Protein Function Gene Ontology Annotation	Reference
1	factor V	1CZT	3DJury HHPred*	52; 58 95; 87	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Blood Clotting	(47)
2	factor VIII	1D7P	3D Jury HHPred	42; 57 88; 82	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Blood Clotting	(48)
3	Sialidase	1EUT	Libellula, Phyre*	25%; 80%	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Hydrolase	(49)
4	Galactose oxidase	1GOF	3DJury	49; 50	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Oxidoreductase(Oxygen(A))	(50)
5	APC10/DOC1 subunit of the anaphase-promoting complex	1GQP	Phyre HHPred*	-; 30% -; 74	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Cell Cycle	(51)
6	Carbohydrate binding module from laminarinase 16A	1GUI	Libellula		Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Carbohydrate Binding Module	(52)
7	APC10/DOC1 subunit of the anaphase-promoting complex	1JHJ	Phyre HHPred*	-; 25% -; 48	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Cell Cycle	(53)
8	Fucose binding lectin	1K12	HHPred*	94; 95	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Sugar Binding Protein	(54)
9	Galactose oxidase	1K3I	Phyre HHPred	30%; 80% 96; 95	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Oxidoreductase	(55)
10	B1 domain of neuropilin-1	1KEX	3D-Jury HHPred	37; 53 96; 93	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Protein Binding	(56)
11	Coagulation factor V	1SDD	3DJury HHPred Phyre	49; 39 67; 41 94%; 94%	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Blood Clotting	(57)
12	Placental protein 25, pp25	1TVG	HHPred Phyre	54; 53 98%; 97%	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Cell Cycle	(58)

No	Protein Name	PDB entry	Method	Template Score for Laminin β ; γ ₁	SCOP Fold name	SCOP Fold classification code	CATH Fold name	CATH Fold classification code	Protein Function Gene Ontology Annotation	Reference
13	xrcc1	1XNA	Libellula		Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	DNA Binding Protein	(59)
14	Placental protein 25, pp25	1XPW	3DJury	47; 50	Galactose-binding domain-like	49784	Galactose-binding domain-like	2.60.120.260	Structural Genomics, Unknown Function	(58)
15	chitobiase (N-acetyl-beta-glucosaminidase)	1QBA	Phyre	45%; -	Immunoglobulin-like beta-sandwich	[48725]	Chitobiase, domain 2	3.30.379.10	Glycosyl Hydrolase	(60)
16	Hypothetical protein TM1631	1VPQ	Wurst	9.4; -	TIM beta/alpha-barrel	[51350]	Hypothetical protein tm1631.	3.20.20.410	Structural Genomics, Unknown Function	(61)
17	Class I MHC	1HSB	Wurst	9.3; -	Immunoglobulin-like beta-sandwich	[48725]	Immunoglobulins	2.60.40.10	Histocompatibility Antigen	(62)
18	Cu,Zn superoxide dismutase	1TO4	Wurst	9.1; -	Immunoglobulin-like beta-sandwich	[48725]	Immunoglobulin-like	2.60.40.200	Oxidoreductase	(63)
19	Puative regulator protein YcfX	1AP1	Wurst	9.1; -	Ribonuclease H-like motif	[53066]		Not yet assigned	Transferase	(64)
20	Alginate-binding periplasmic protein AlgQ1	1Y3N	Wurst	9.1; -	Periplasmic binding protein-like II	[53849]	Periplasmic binding protein-like II	3.40.190.10	Sugar Binding Protein	(65)
21	Myo-inositol 1-phosphate synthase	1VKO	Wurst	-; 11.0	AD(P)-binding Rossmann-fold domains	[51734]		Not yet assigned	Isomerase	(66)
22	c-KIT receptor	1T46	Wurst	-; 10.0	Protein kinase-like (PK-like)	[56111]	Phosphorylase Kinase; domain 1	3.30.200.20	Transferase Activator	(67)
23	Interleukin-8	1ILQ	Wurst	-; 10.0	IL8-like	[54116]	OB fold (Dihydroliipoamide Acetyltransferase, E2P)	2.40.50.40	Cytokine	(68)
24	FYVE, RhoGEF and PH domain containing protein 3, FGD3	2COC	Wurst	-; 10.0	PH domain-like barrel	[50728]		Not yet assigned	Signaling Protein	(69)
25	Class delta GST	1V2A	Wurst	-; 10.0	GST C-terminal domain-like	[47615]	Glutaredoxin	3.40.30.10	Transferase	(70)
26	PI-Pfui intein	1DQ3	Superfam		Hedgehog/intein (Hint) domain	[51293]	Endonuclease - Ptscei; Chain A, domain 1	2.170.16.10	Hydrolase	(71)

No	Protein Name	PDB entry	Method	Template Score for Laminin β_1 ; γ_1	SCOP Fold name	SCOP Fold classification code	CATH Fold name	CATH Fold classification code	Protein Function Gene Ontology Annotation	Reference
27	Galectin-1	1W6O	Superfam		Concanavalin A-like lectins/glucanases	[49898]	Jelly Rolls	2.60.120.200	Lectin	(72)
28	Isoamylase	1BF2	Libellula	2.6; -	Immunoglobulin-like beta-sandwich	[48725]	Immunoglobulins	2.60.40.10	Hydrolase	(73)
29	N-cadherin (neural)	1OP4	Libellula	9.0; -	Immunoglobulin-like beta-sandwich	[48725]	Not yet assigned		Cell Adhesion	(74)
30	Plasminogen	1K10	Libellula	4.3; -	Kringle-like	[57439]	Plasminogen Kringle 4	2.40.20.10	Hydrolase	(75)
31	CD69	1E8I	Libellula	-; 8.6	C-type lectin-like	[56435]	Mannose-Binding Protein A, subunit A	2.60.120.40	Hematopoietic Cell Receptor	(76)
32	Tumor necrosis factor (TNF)	2TNF	Libellula	-; 10.5	TNF-like	[49841]	Jelly Rolls	2.30.27.10	Cytokine	(77)
33	minor coat protein g3p	1G3P	Libellula	-; 11.7	N-terminal domains of the minor coat protein g3p	[50175]	Phage FD Coat Protein, Membrane penetration domain	2.60.40.10	Viral Protein	(78)
34	Immunoglobulin heavy chain	1QFW	Libellula	-; 14.4	Immunoglobulin-like beta-sandwich	[48725]	Immunoglobulins	2.10.60.10	Immune System	(79)
35	TGF-beta type II receptor	1KTZ	Libellula	-; 12.2	Snake toxin-like	[57301]	CD59	3.10.200.10	Cytokine/Cytokine Receptor	(80)
36	Carbonic anhydrase	2ZNC	Loopp	1.1; -	Carbonic anhydrase	[51068]	Carbonic Anhydrase II	2.60.40.200	Lyase	(81)
37	Copper chaperone for superoxide dismutase	1DO5	Loopp	0.9; -	Immunoglobulin-like beta-sandwich	[48725]	Immunoglobulin-like	3.10.200.10	Chaperone	(82)
38	Carbonic anhydrase	1RJ6	Loopp	0.9; -	Carbonic anhydrase	[51068]	Carbonic Anhydrase II	2.80.10.50	Lyase	(83)
39	Winged bean albumin 1	1WBA	Loopp	0.9; -	beta-Trefoil	[50352]	Trefoil (Acidic Fibroblast Growth Factor, subunit A)	2.60.120.260	Seed Storage Protein	(84)
40	Micromonospora sialidase	1W8O	HHPred	94; 97	6-bladed beta-propeller	[50938]	Galactose-binding domain-like		Hydrolase	(85)

Table 5

Remaining structures based on the number of fulfilled distance constraints (disulfide bonds and cross-links, see Table 3). Compactness (according to Table 6) was used as an additional criterion to validate the obtained structures, which fulfill all cross-linking constraints.

Protein	Modeling Parameters criterion I-III						Number of fulfilled distance constraints criterion IV and V					
	sequence-sequence alignments	Non-redundant alignments	Criterion I Modeled loops	Criterion II Similarity to Template and energy cut-off	Criterion III Compactness	1	2	3	4	5	6	7
Ges. 81	730,800	5,371	4,333	506		506	506	501	472	288	54	3
ICZT	52,200	499	296	43	41	43	43	43	43	28	8	0
ID7P	52,200	364	313	84	83	84	84	84	81	52	10	3
IEUT	52,200	317	294	14	7	14	14	14	13	8	1	0
IGOF	52,200	1017	799	48	10	48	44	48	47	27	7	0
IGQP	52,200	241	241	84	45	84	84	84	79	43	0	0
IGUI	52,200	344	344	44	44	44	44	44	44	42	8	0
IJHJ	52,200	300	283	25	6	25	25	25	24	13	2	0
IK12	52,200	261	201	25	24	25	25	25	24	15	3	0
IK3I	52,200	727	591	35	10	35	35	35	34	17	5	0
IKEX	52,200	177	141	16	16	16	16	16	16	11	0	0
ISDD	52,200	614	372	59	3	59	59	55	40	12	2	0
ITVG	52,200	193	184	16	16	16	16	16	16	3	0	0
IXPW	52,200	193	192	12	10	12	12	12	12	12	6	0
Ges. 71	730,800	9,830	7,306	1,349		1,349	1,334	1,159	387	12		
ICZT	52,200	695	370	150	145	150	150	144	53	3		
ID7P	52,200	616	507	146	142	146	146	132	31	1		
IEUT	52,200	509	432	54	34	54	54	31	14	0		
IGOF	52,200	811	657	57	30	57	56	37	11	0		
IGQP	52,200	354	285	180	170	180	80	142	71	1		
IGUI	52,200	409	319	103	102	103	103	96	26	1		
IJHJ	52,200	797	776	48	18	48	48	47	8	0		
IK12	52,200	572	385	171	171	171	171	168	12	1		

Protein	Modeling Parameters criterion I-III					Number of fulfilled distance constraints criterion IV and V						
	sequence-sequence alignments	Non-redundant alignments	Criterion I Modelled loops	Criterion II Similarity to Template and energy cut-off	Criterion III Compactness	1	2	3	4	5	6	7
1K3I	52,200	704	438	52	27	52	52	39	10	0		
1KEX	52,200	180	118	57	56	57	57	56	38	5		
1SDD	52,200	2896	1947	208	57	208	208	147	42	(1)*		
1TVG	52,200	485	374	30	29	30	30	28	20	0		
1XNA	52,200	348	263	8	8	8	8	8	5	0		
1XPW	52,200	454	435	30	22	30	30	29	17	0		

* Structure did not fulfill the compactness criterion.

Table 6

Composite scoring function which integrates the deviation of the experimental constraints from the theoretical maximum distance as well as the Rosetta energy and compactness of the structure

Score = $LN \frac{X_{i,max}}{X_{i,max}} \frac{X_i}{X_{i,opt}}$		
With		
	Optimal X_{opt}	Maximal allowed X_{max}
Chemical Cross-Link	<19 Å	<24 Å
Disulfide	<8 Å	<13 Å
Energy (bk_tot)	<-250	<-180
Compactness (SASA)	<15.000	<20.000

If $X_i < X_{opt}$ then $\frac{X_{i,max}}{X_{i,max}} \frac{X_i}{X_{i,opt}}$ was set to 1, where else in the case $X_i > X_{max}$ the term was set to 0.05.

Table 7

Summary of the filter function and the *leave-one-out* (LOO) experiments. “filter total”, “filter (6/7)”, and “filter (4/5)” indicate how many models in total and how many models that fulfill 6 out of 7, respectively, 4 out of 5 distance constraints pass the filter; LOO distance and LOO fold indicate the results of LOO in respect to distance fulfillment and structural homology with a Z score of larger than 8; information content measure (sequence separation [aa] / Euclidian separation [$\text{\AA}$]) of the distance constraints. XL: Cross-link.

Laminin β_1	Disulfide 1	Disulfide 2	Disulfide 3	Disulfide 4	XL 1	XL 2	XL 3	Mean
Filter (total)	484 95.7%	117 23.1%	467 92.3%	47 9.3%	493 97.4%	288 56.9%	435 86.0%	65%
Filter (6/7)	33 100%	27 81.8%	33 100%	9 27.3%	33 100%	33 100%	33 100%	84%
LOO distance	3/3 100%	3/9 33%	3/3 100%	3/27 11.1%	3/3 100%	3/3 100%	3/3 100%	78%
LOO fold	3/3 100%	6/9 66%	3/3 100%	14/27 52%	3/3 100%	3/3 100%	3/3 100%	88%
Information content [$\text{\AA}^{-1}$]	0.38	0.69	0.23	1.54	0.37	1.90	0.47	

Laminin γ_1	Disulfide 1	Disulfide 2	Disulfide 3	Disulfide 4	XL 1	Mean
Filter (total)	1226 90.8%	260 19.3%	1348 99.9%	218 16.1%	1193 88.4%	63%
Filter (4/5)	291 99.3%	161 54.9%	293 100.0	148 50.5%	291 99.3%	81%
LOO distance	12/14 86%	12/133 9%	12/12 100%	12/145 8%	12/12 100%	61%
LOO fold	9/14 64%	68/133 51%	9/12 75%	82/145 57%	9/12 75%	64%
Information content [$\text{\AA}^{-1}$]	0.77	0.62	0.23	1.31	0.43	