PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Computational modeling of laminin N-terminal domains using sparse distance constraints from disulfide bonds and chemical cross-linking

Stefan Kalkhof,[1] Sebastian Haehn,[2] Mats Paulsson,[2] Neil Smyth,[3] Jens Meiler,[4]* and Andrea Sinz[1]*

[1] Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[2] Center 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, Cologne, Germany

[3] School of Biological Sciences, University of Southampton, Bassett Crescent, East Southampton, United Kingdom

[4] Department of Chemistry and Center for Structural Biology, Vanderbilt University Nashville, Tennessee

ABSTRACT

Basement membranes are thin extracellular protein layers, which separate endothelial and epithelial cells from the underlying connecting tissue. The main noncollagenous 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, that is, 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.

Proteins 2010; 78:3409–3427.

Key words: laminin; N-terminal domains; disulfide bonds; chemical cross-linking; mass spectrometry; computational modeling.

INTRODUCTION

Laminins are the major noncollagenous proteins of basement membranes and are known to form networks due to crucial noncovalent 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 Ref. 2) that embed one or two additional domains with a total length of 40–60 nm each. Toward 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 with three short arms 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 laminin fragments with two or fewer LN domains fail to form networks.[5]

Additional Supporting Information may be found in the online version of this article.

Abbreviations: BS^2G, *bis*(sulfosuccinimidyl)glutarate; BS^3, *bis*(sulfosuccinimidyl)suberate; DHB, 2,5-dihydroxybenzoic acid; DTT, dithiothreitol; ESI-FTICR-MS, electrospray ionization Fourier transform ion cyclotron resonance MS; 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 N-terminal; LOO, leave-one-out; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight MS; NHS, *N*-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.

Grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Grant numbers: SI 867/7-1 and SM 65/1-3; Grant sponsor: Köln Fortune program of the Medical Faculty of the University of Cologne.
Stefan Kalkhof's current address is Department of Proteomics, Helmholtz-Centre for Environmental Research - UFZ, Permoserstrasse 15, D-04318 Leipzig, Germany.
*Correspondence to: Prof. Dr. Jens Meiler, Department of Chemistry and Center for Structural Biology, Vanderbilt University Nashville, TN 37212. E-mail: jens.meiler@vanderbilt.edu or Prof. Dr. Andrea Sinz, Martin Luther University Halle-Wittenberg, Wolfgang-Langenbeck-Strasse 4, D-06120 Halle (Saale), Germany. E-mail: andrea.sinz@pharmazie.uni-halle.de.
Received 23 April 2010; Revised 16 July 2010; Accepted 25 July 2010
Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.22848

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 γ1 and γ3 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 by 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.[7–10] 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 used), (II) the analysis is rapid, (III) it requires very small (10^{-15} to 10^{-18} 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, posttranslational 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 (ca. 5 to 25 Å) allow setup of fine-tuned experimental strategies.[7–11]

However, despite the straightforwardness of the cross-linking approach, the identification of the cross-linked products can be cumbersome because of the complexity of the reaction mixtures. Several strategies have been used to enrich cross-linker-containing species by affinity chromatography or to facilitate the identification of the cross-linked products, for example, 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-MS).[12] The aim of this 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, that is, 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, that is, deuterated, cross-linkers to facilitate the identification of cross-linker-containing species in the mass spectra based on their characteristic isotope patterns.[13–15]

Sparse experimental distance restraints can restrict the conformational space for a protein substantially and, therefore, enable determination of tertiary structure by 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 ~0.1 constraints per amino acid the correct fold was determined *de novo* for T4-lysozyme and αA-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 Supporting Information will facilitate this process.

MATERIALS AND METHODS

Materials

The cross-linking reagents BS^3-D_0/D_4 (*bis*(sulfosuccinimidyl)suberate-D_0/D_4), and BS^2G-D_0/D_4 (*bis*(sulfosuccinimidyl)glutarate-D_0/D_4) were obtained from Pierce Inc. (Rockford, IL). 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 BS^2G and BS^3 were used as 1:1 mixtures of nondeuterated and four-times deuterated derivatives (D_0/D_4). Cross-linking reactions were conducted with 2 μ*M* protein solutions in 20 m*M* HEPES buffer, 100 m*M* NaCl, 5 m*M* $CaCl_2$, and 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 NH_4HCO_3 to a final concentration of 20 m*M*.

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 und 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 μL 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 μL.

Nano-HPLC/MALDI TOF/TOF-MS

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 by an autosampler with a 200-μL sample loop onto a precolumn (PepMap, C18, 300 μm × 5 mm, 3 μm, 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 μm × 150 mm, 3 μm, 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 μL of matrix solution (0.7 μg/μL HCCA in 90% ACN/0.1% TFA, 1 m*M* $NH_4H_2PO_4$) and directly prepared onto a 384 MTP 800-μm 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 2000 laser shots in the range m/z 800–4000 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 Sophisticated Numerical Annotation Procedure (SNAP) algorithm. Afterward, 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 2000 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 2000 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 MS

Proteolytic peptide mixtures were additionally analyzed by nano-HPLC/nano-ESI-MS 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 by the autosampler with a 20-μL sample loop and were desalted and concentrated by washing the precolumn (PepMap, C18, 300 μm × 5 mm, 5 μm, 100 Å, Dionex) with 0.1% FA at a flow rate of 2 μL/min for 10 min, before peptides were eluted onto the separation column (PepMap, C18, 75 μm × 150 mm, 3 μm, 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 solvents 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 software (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 μm, New Objective, Woburn, MA) were used. 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 400–2000) with 256 k data points per spectrum. Ten 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 ICR analyzer and is equipped with a 7-T magnet. Samples were injected by the autosampler with a 100-μL sample loop, desalted, and concentrated on a precolumn (PepMap, C18, 300 μm × 5 mm, 3 μm, 100 Å). Separation of peptides was performed on a C18 column (PepMap, C18, 75 μm × 150 mm, 3 μm, 100 Å) using a 90-min gradient (0%–60% solvent B in 90 min, followed by isocratic elution at 90% B for 3 min; solvents A: 5% ACN, 0.1% FA and 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 LTQ. Dynamic exclusion (exclusion time 20 sec, exclusion window ±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*, *Biotools* and *MS2Assign*. The in-house software tool *Isofind* determines signals with a defined mass difference of 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 (0.25:1–1:4) and similar LC retention times (±1 min) were considered as potential cross-linked products. The General Protein Mass Analysis for Windows (GPMAW, 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 with the predicted fragment masses for a cross-linked product. In addition, the freely accessible program MS2Assign (http://roswell.ca.sandia.gov/~mmyoung/ms2assign.html; part of the Collaboratory for MS3D, http://ms3d.org/home.php) was used for calculation of MS/MS data and comparison with experimental data. The software package Biotools 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] In addition, oxidation (Met), carboxamidomethylation (Cys), amidation (at cross-linker and at C-terminus), deamidation (Gln, Asn), and cyclization (Cys) were taken into account as potential modifications as well as incomplete cleavage (up to eight 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] Fig. 1(A)]. Potential templates (PDB entries: 1CZT, 1D7P, 1KEX, and 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 used 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; 10 iterative rounds). In this experiment, the sequences of netrin I, II, and IV as well as these of LN domains of laminin chains α_1, α_2, α_3B, α_5, β_2, β_3, and 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, and 1XPW were submitted to the 3D-Jury metaserver for a second iteration. Two additional template candidates (1SDD and 1GOF) were

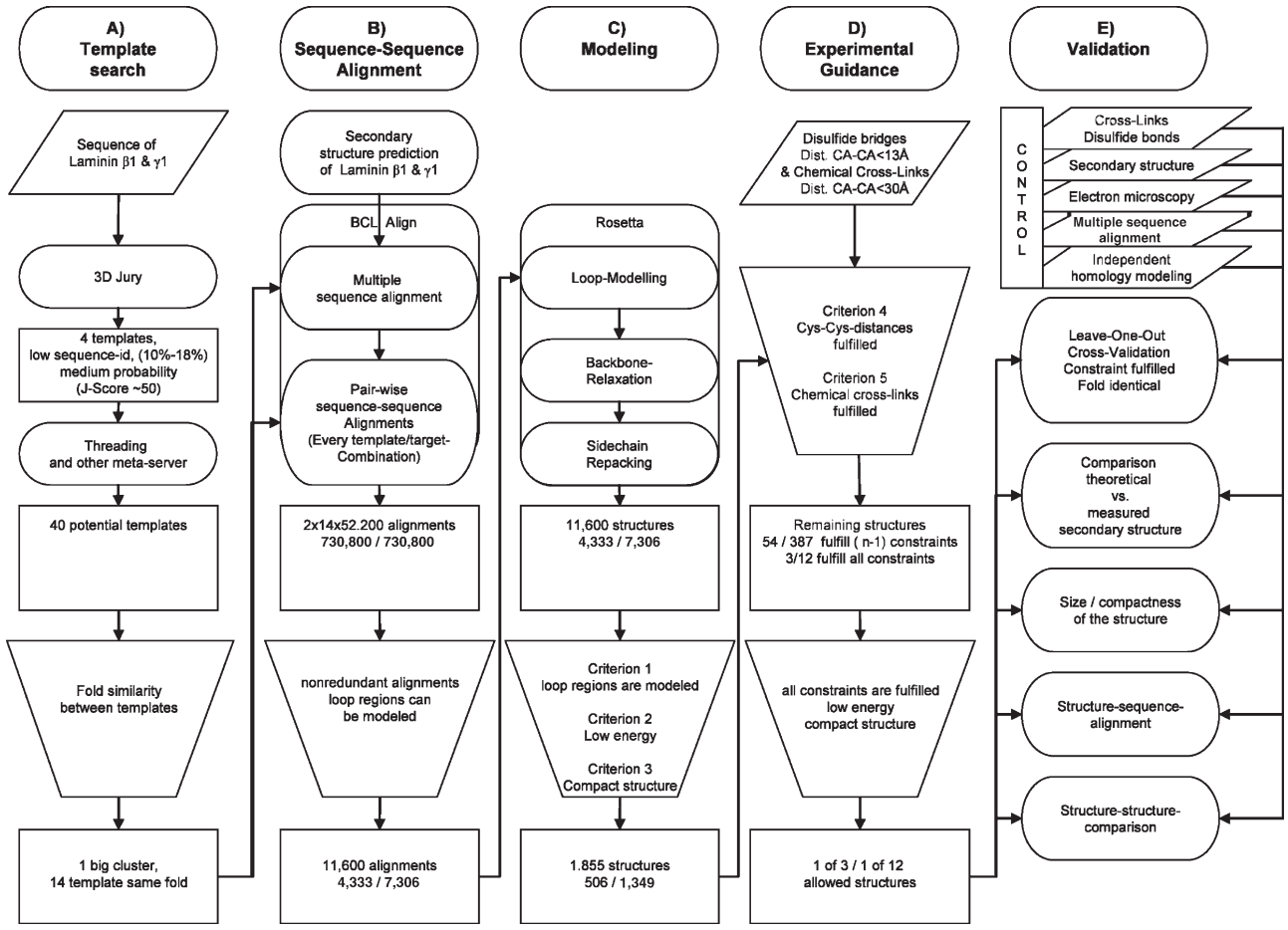

Figure 1

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.

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.tuebingen.mpg.de/hhpred[31]], and Loopp [http://loopp.org/[32]] were applied. Thirty-five putative templates were identified in addition to the previously detected structures yielding a total of 40 candidate templates [Fig. 1(A)].

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. Fourteen 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 IV). These 14 structures were considered as templates in the subsequent comparative modeling efforts [Fig. 1(B)]. Other structures have been dismissed as potential templates as in structure clustering (Fig. 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, similarly 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

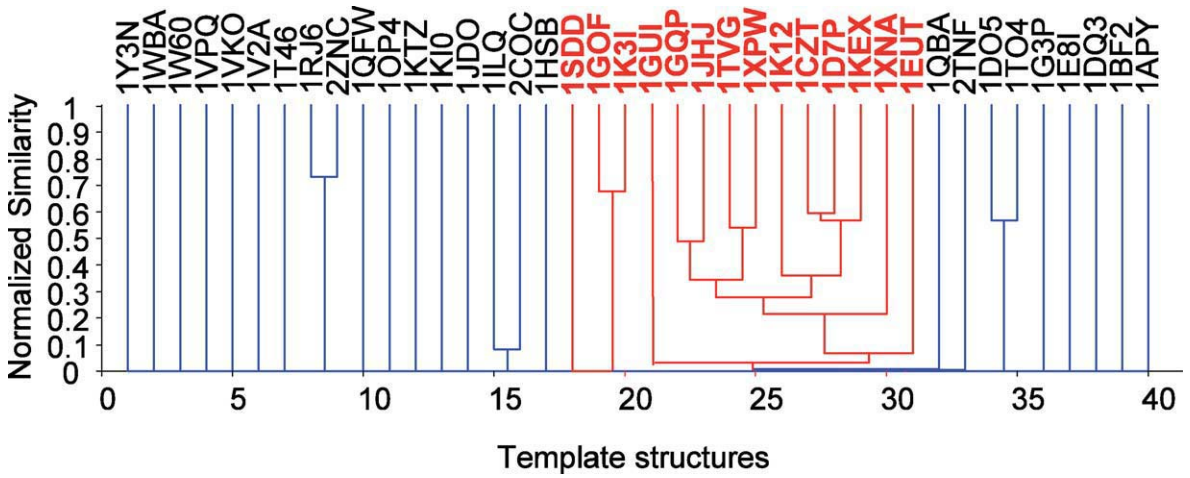

Figure 2

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. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

SAM[37] yielding large content of random coil and β-strands. To sample the space of possible alignments densely, a grid search of alignment parameters was performed[38] [Supporting Information Table I, Fig. 1(B)] 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 ~3 Å. 4,333 (laminin β_1) and 7306 (laminin γ_1) target/template-alignments fulfilling this rule were carried forward [Fig. 1(B)].

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 10 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 [Fig. 1(C)].

Validation and refinement of LN domain structures

Models, in which not all loops could be closed, were removed (Criterion 1). In addition, 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 $Å^2$ for lack of compactness were removed (Criterion 3) [Fig. 1(C)]. Further models were removed that displayed C^A–C^A distances for cross-links larger than 30 Å (Criterion 4) or C^A–C^A distances of disulfide bridges larger than 13 Å (Criterion 5) [Fig. 1(D) and Table V]. A total of seven constraints were available for the laminin β_1 LN domain and five for the laminin γ_1 LN domain. Only 3 models of the laminin β_1 LN domain and 12 models of the γ_1 LN domain fulfilled all experimental constraints [Fig. 1(D)]. 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 this study, <1% of the initial models fulfill all distance constraints (Table V). The models were ranked according to an empirical composite score derived from these four criteria (Table VI). 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 Supporting Information).

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 of seven (for laminin β_1) and four 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 with the models selected

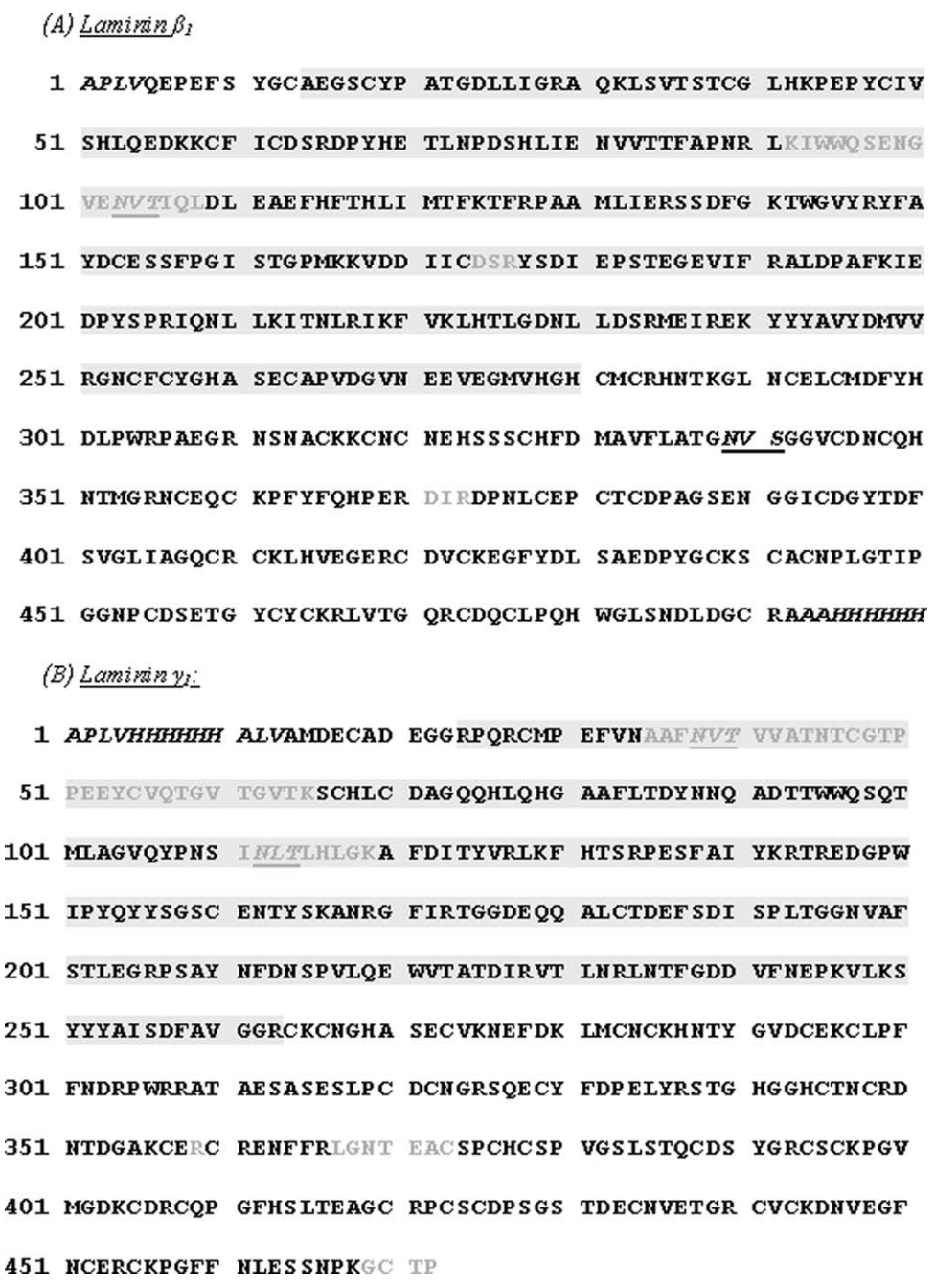

Figure 3

Amino acid sequences of recombinant (A) laminin β1 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 fragments are printed in italics. LN domains are shaded.

based on 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 VII and Supporting Information Table II). The final assessment of the models was based on precision as measured by the composite score (Table VI and Supporting Information Table II).

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 Supporting Information.

Table I
Intramolecular Cross-Linked Products of Isolated Laminin β_1 and γ_1 LN Domains Using BS^3 as Cross-Linker

Protein	$MH+_{exp}$	$MH+_{theo}$	Sequence	Amino acids	Δm (ppm)	MS/MS	D_0/D_4	Digestion
β_1	1293.794	1293.793	218–227	R/I*K*FV*K*LHTLG/D	1	b5–9	1.9	*In-solution*
	1297.821	1297.818			2	y4–5		
β_1	1293.786	1293.793	218–227	R/I*K*FV*K*LHTLG/D	5	b5–9	1.8	*In-gel*
	1297.810	1297.818			6	y3–5		
β_1	1672.824	1672.822	194–206	L.DPAF*K*IEDP*Y*SPR.I	1	y1, y12	1.4	*In-solution*
	1676.846	1676.847			1	b3		
β_1	1672.804	1672.822	194–206	L.DPAF*K*IEDP*Y*SPR.I	11	y1–3, y12,	1.4	*In-gel*
	1676.830	1676.847			10	b2, b4		
β_1	1856.951	1856.943	192–206	R.ALDPAF*K*IEDP*Y*SPR.I	4	y1–3; y12, b2	1.4	*In-solution*
	1860.976	1860.968			4			
β_1	2282.310	2282.334	194–200	L.DPAF*K*IE.D	11	1: y6	1.6	*In-solution*
	2286.329	2286.359	+207–217	+ R.IQNLL*K*ITNLR.I	13	2: y1-y5, b2-b5		
β_1	2466.415	2466.433	192–200	R.ALDPA*K*IE.D	7	2: b1, 3, y1–4	1.5	*In-gel*
	2470.444	2470.458	+207–217	+R.IQNLL*K*ITNLR.I	6			
β_1	2466.455	2466.433	192–200	R.ALDPA*K*IE.D	9	1: b2, 3	1.6	*In-solution*
	2470.479	2470.458	+207–217	+R.IQNLL*K*ITNLR.	9	2: b2		
γ_1	2118.135	2118.149	128–143	R.L*K*FHTSRPESFAIY*K*R.T	7	y1, 2, 5–7, 9	1.8	*In-solution*
	2122.156	2122.174			8	b7, 14, 15		
γ_1	2118.145	2118.149	128–143	R.L*K*FHTSRPESFAIY*K*R.T	2	y1, b15	1.7	*In-gel*
	2122.166	2122.174			4			
γ_1	2651.284	2651.299	189–212	S.DISPLTGGNVAFSTLEGRP*S*A*Y*NF.D	6	y1, 2, 6–12, 14, 18, 21, 23	1.3	*In-gel*
	2655.307	2655.324			6	b2–4, 6, 9–11, 18		
γ_1	2651.286	2651.299	189–212	S.DISPLTGGNVAFSTLEGRP*S*A*Y*NF.D	5	y2, 6, 23, b2,	1.5	*In-solution*
	2655.301	2655.324			9	3, 6		

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.

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 earlier (Table VI). These structures were evaluated based on the structure homology to give the overall best model. To assess the importance of each distance constraint 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, that is, in these cases for which all except one constraint are fulfilled.

RESULTS AND DISCUSSION

The aim of this 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, that is, disulfide bonds, as well as by chemical cross-links obtained from reactions with the amine-reactive cross-linkers BS^2G and BS^3. 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 BS^3-D_0/D_4 and BS^2G-D_0/D_4 bridging distances of about 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 toward primary amines, that is, 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-linker (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 (BS^3-D_0) and 96.021 u (BS^2G-D_0), respectively. Peptides, which are modified by a partially hydrolyzed cross-linker, exhibit mass increases of 156.079 u (BS^3-D_0) and 114.032 u (BS^2G-D_0), respectively. BS^3 and BS^2G were used as 1:1 mixtures of their nondeuterated (D_0) and deuterated (D_4) species to facilitate the identification of cross-linked products by means of their distinct doublet isotope patterns with mass differences of 4.025 u (D_0/D_4) 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

Table II
Intramolecular Cross-Linked Products of Cross-Linking Reaction Mixtures Containing Both Laminin β_1 and γ_1 LN Domains

	$MH+_{exp}$	$MH+_{theo}$	Sequence	Amino acids and cross-linker	Δm (ppm)	MS/MS	D_0/D_4	Digestion
Laminin β_1	1293.787	1293.793	218–227	R.I*K*FV*K*LHTLG.D + BS^3	5	y4–5,	1.3	*In-solution*
	1297.812	1297.818			5	b5–9		
	1293.821	1293.793	218–227	R.I*K*FV*K*LHTLG.D + BS^3	22	y3–5, b5–9	1.8	*In-gel*
	1297.841	1297.818			18			
	1672.823	1672.822	194–206	L.DPAF*K*IEDP*YS*PR.I + BS^3	1	b3, y1–3, 12	1.4	*In-solution*
	1676.847	1676.847			0			
	1672.833	1672.822	194–206	L.DPAF*K*IEDP*YS*PR.I + BS^3	7	b3, y1–3, 9, 12	3.0	*In-gel*
	1676.860	1676.847			8			
	2282.322	2282.312	194–200 +	L.DPAF*K*IE.D+R.IQNLL*K*ITNLR.I	4	1 : b3	1.8	*In-gel*
	2286.342	2286.337	207–217	+ BS^3	2	2 : b2–4, y1–5		
	2466.430	2466.433	192–200 +	R.IQNLL*K*ITNLR.I	1	—	1.6	*In-gel*
	2470.454	2470.458	207–217	+ R.ALDPAF*K*IE.D + BS^3	2			
	1421.842	1421.847	207–217	R.IQNLL*K*I*T*NLR.I + BS^2G	4	y1–5, b1–5	0.8	*In-solution*
	1425.867	1425.869			1			
	1630.800	1630.775	194–206	L.DPAF*K*IEDP*YS*PR.I + BS^2G	15	y1, 2, 12	0.6	*In-gel*
	1634.793	1634.800			4			
	1836.901	1836.865	176–191	R.*YS*DIEP*ST*EGEVIFR.A	20	y1–8	0.7	*In-solution*
	1840.924	1840.890		amidated(C-term); + BS^2G	18			
	1836.921	1836.865	176–191	R.*YS*DIEP*ST*EGEVIFR.A	30	y1–4	0.6	*In-gel*
	1840.944	1840.890		amidated (C-term); + BS^2G	29			
	1935.877	1935.883	152–168	Y.DCESSFPGISTGPM*KK*V.D + BS^2G	3	y2–3, 5–8,	0.7	*In-gel*
	1939.900	1939.908			4	10–13, b4, b9, b6, b4		
	2240.265	2240.265	194–200 +	L.DPAF*K*IE.D +R.IQNLL*K*ITNLR.I + BS^2G	0	b3,	0.7	*In-solution*
	2244.286	2244.290	207–217		2	y1, 4, 17		
	2240.260	2240.265	194–200 +	L.DPAFKIE.D +R.IQNLL*K*ITNLR.I + BS^2G	2	1 : y1–5, b2–5	0.7	*In-gel*
	2244.285	2244.290	207–217		2	2 : b2–4		
	2501.238	2501.235	136–147 +	R.SSDFG*K*TWGVYR.Y +	1	1 : b2, y1, 2,	0.5	*In-solution*
	2505.257	2505.260	192–200	R.ALDPAF*K*IE.D + BS^2G	1	5		
	2501.216	2501.235	136–147 +	R.SSDFG*K*TWGVYR.Y	8	A: b2, 3, 5,	0.6	*In-gel*
	2505.239	2505.260	192–200	+ R.ALDPAF*K*IE.D + BS^2G	8	y1, 4–6 b : b3		
	2552.132	2552.135	152–173	Y.DCESSFPGISTGPM*KK*VDDIIC.D + BS^2G	1	b6, 9, 10, 17, 21	0.5	*In-solution*
	2556.149	2556.160			4	y10, 11, 13, 17		
	2552.021	2552.135	152–173	Y.DCESSFPGISTGPM*KK*VDDIIC.D + BS^2G	45	—	0.6	*In-gel*
	2556.044	2556.160			45			
	2751.278	2751.312	33–55	K.LSVTSTCGLH*K*PEPYCIVSHLQE.D + BS^2G	1	y3–y6, 10–14	0.5	*In-solution*
	2755.315	2755.337			8	b2, 4, 10, 11		
	2751.272	2751.312	33–55	K.LSVTSTCGLH*K*PEPYCIVSHLQE.D + BS^2G	15	b2, 4, 10	0.6	*In-gel*
	2755.312	2755.337			9	y5–6, 10, 12		
	3078.499	3078.502	30–55	R.AQ*K*LSVTSTCGLH*K*PEPYCIVSHLQE.D	1	b15, 17	0.3	*In-gel*
	3082.529	3082.527		+ BS^2G	1	y4, 5, 10, 12		
Laminin γ1	2118.151	2118.149	128–143	R.L*K*FH*TS*RPESFAIY*K*R.T + BS^3	1	b7, 9, 15	1.3	*In-solution*
	2122.176	2122.174			1	y1, 7, 9		
	2118.172	2118.149	128–143	R.L*K*FH*TS*RPESFAIY*K*R.T + BS^3	11	(y1, b15)	0.8	*In-gel*
	2122.170	2122.174			2			
	2076.080	2076.103	128–143	R.L*K*FHTSRPESFAIY*K*R.T + BS^2G	11	b1, 15	0.8	*In-solution*
	2080.125	2080.128			1	y1, 6, 7, 10		
	2852.190	2852.226	147–169	E.DGPWIPYQYYSGSCENTYSKANR.G + BS^2G	13	y1–3, b3–b5,	0.4	*In-solution*
	2856.198	2856.251			19	b8, b16		
	2949.432	2949.431	239–263	G.DDVFNEP*K*VL*K*SYYYAISDFAVGGR.C + BS^2G	0	b18, b20,	0.7	*In-solution*
	2953.455	2953.456			0	y1, 3, 19		

The cross-linker (BS^2G or BS^3) is indicated. All cross-linked products were verified by MS/MS data. The cross-linked amino acids are printed in italics and are underlined.

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).

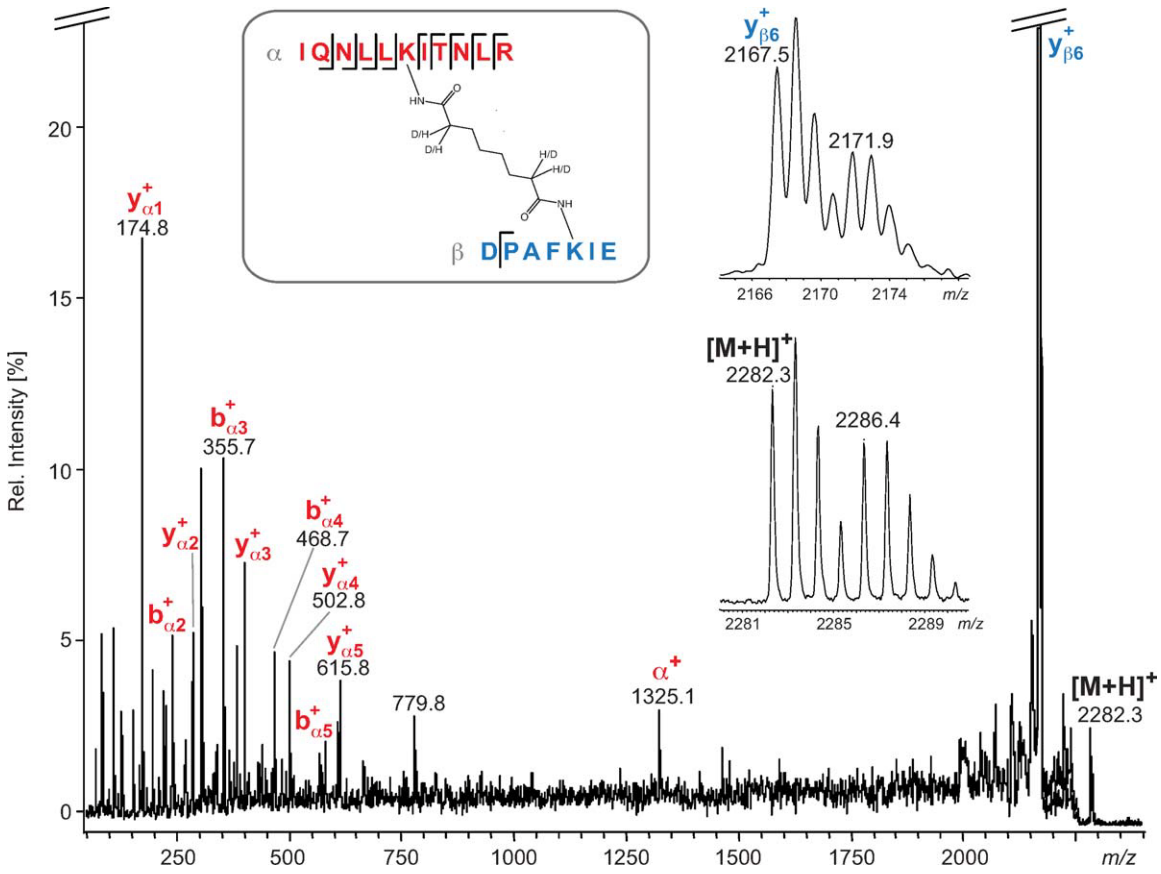

Figure 4

MALDI-TOF/TOF-MS/MS analysis of the BS^3 cross-linking product of laminin β_1 LN domain (amino acids 194–200 connected with amino acids 207–217; $[M + H]^+_{exp}$ 2282.310), lysines at positions 198 and 212 were found to be connected. The insets show the precursor ion and y6 fragment ion, both exhibiting the characteristic D_0/D_4 isotope pattern. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

When isolated laminins were subjected to the cross-linking reaction with BS^3 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 I 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 BS^3 and BS^2G (Table II). 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, 11 cross-linked products for laminin β_1 LN domain and 4 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 MS has been described recently.[12] Briefly, after recording the mass spectra, the putatively disulfide-linked peptides are subjected to laser-induced fragmentation technique (LIFT)-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

Table III
Distance Constraints in Laminin β_1 and γ_1 LN Domains Serving as Basis for Computational Modeling

	Connected amino acids	Connection type
Laminin β1	C13–C18	Disulfide
	K32–K43	BS^2G
	C39–C48	Disulfide
	C59–C62	Disulfide
	K141–K198	BS^2G
	C153–C173	Disulfide
	K166–K167	BS^2G
	K198–Y203/S204	BS^2G, BS^3
	K198–K212	BS^2G, BS^3
	K212–T214	BS^2G
	K219–K222	BS^3
	K222–T225	BS^2G
Laminin γ1	C18–C28	Disulfide
	C47–C55	Disulfide
	C67–C70	Disulfide
	K129–K142	BS^2G, BS^3
	C160–C183	Disulfide
	S208–Y210	BS^3
	K246–K249	BS^3

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 used as a direct structural filter because of the spatial proximity in the primary structure of laminin LN domains.

the recombinant laminin β1 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 III 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 about 30% for laminin β_1 and about 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, and 1XPW) were identified with 3D-Jury scores ranging between 39 and 58 (Table IV). 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 Loopp[28–32] (Table IV). To eliminate false-positives, a pairwise structure–structure comparison of all putative templates was executed. Fourteen of 40 templates displayed the galactose-binding domain-like fold (Table IV). The 26 remaining templates were removed as false-positives as they exhibited similarity with one other template at most (Fig. 2). Sequence identities between the templates and laminin β_1 and γ_1 LN domains ranged between 10% and 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, Blosum45, Pam250, and Blast), secondary structure (programs JUFO,[19,35] Psipred,[36] and SAM[37]), and sequence gaps (gap opening and gap extension) with variable weights. 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 Å), the fold deviated significantly from the starting template, or the model was energetically clearly disfavored, that is, had substantial clashes (Table V).

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 III. The identification of a disulfide bond implies that the C^A atoms of the cysteines involved are within a distance of about 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

Table IV
Results of the Template Search

No.	Protein name	PDB entry	Method	Template score for laminin β1; γ1	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
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*-β-*glucoseaminidase)*	1QBA	Phyre	45%; —	Immunoglobulin-like β-sandwich	48725	Chitobiase, domain 2	3.30.379.10	Glycosyl hydrolase	60
16	Hypothetical protein TM1631	1VPQ	Wurst	9.4; —	TIM β/α-barrel	51350	Hypothetical protein tm1631.	3.20.20.410	Structural genomics, unknown function	61
17	Class I MHC	1HSB	Wurst	9.3; —	Immunoglobulin-like β-sandwich	48725	Immunoglobulins	2.60.40.10	Histocompatibility antigen	62
18	Cu,Zn superoxide dismutase	1TO4	Wurst	9.1; —	Immunoglobulin-like β-sandwich	48725	Immunoglobulin-like	2.60.40.200	Oxidoreductase	63
19	Putative 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	NAD(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 (Dihydrolipoamide 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

(Continued)

Table IV
(Continued)

No.	Protein name	PDB entry	Method	Template score for laminin β1; γ1	SCOP fold name	SCOP fold classification code	CATH fold name	CATH fold classification code	Protein function gene ontology annotation	Reference
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 - Pi-scei; Chain A, domain 1	2.170.16.10	Hydrolase	71
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 β-sandwich	48725	Immunoglobulins	2.60.40.10	Hydrolase	73
29	N-cadherin (neural)	1OP4	Libellula	9.0; —	Immunoglobulin-like β-sandwich	48725	Not yet assigned		Cell adhesion	74
30	Plasminogen	1KI0	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 β-sandwich	48725	Immunoglobulins	2.10.60.10	Immune System	79
35	TGF-β 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 β-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; —	β-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 β-propeller	50938	Galactose-binding domain-like		Hydrolase	85

Fourteen structures were found using the programs 3D-Jury, Superfamily—"Superfam," Phyre, Libellula "Lib," HHPred; the programs Wurst and Loopp yielded only false-positive template structures. The six templates found with 3D-Jury are printed in bold.

Table V
Remaining Structures Based on the Number of Fulfilled Distance Constraints (Disulfide Bonds and Cross-Links, See Table III)

	Modeling parameters Criteria I–III					Number of fulfilled distance constraints Criteria IV and V						
Protein	Sequence–sequence alignments	Nonredundant alignments	Criterion I, modeled loops	Criterion II, similarity to template and energy cut-off	Criterion III, compactness	1	2	3	4	5	6	7
Laminin β1	730,800	5371	4333	506		506	506	501	472	288	54	3
1CZT	52,200	499	296	43	41	43	43	43	43	28	8	0
1D7P	52,200	364	313	84	83	84	84	84	81	52	10	3
1EUT	52,200	317	294	14	7	14	14	14	13	8	1	0
1GOF	52,200	1017	799	48	10	48	44	48	47	27	7	0
1GQP	52,200	241	241	84	45	84	84	84	79	43	0	0
1GUI	52,200	344	344	44	44	44	44	44	44	42	8	0
1JHJ	52,200	300	283	25	6	25	25	25	24	13	2	0
1K12	52,200	261	201	25	24	25	25	25	24	15	3	0
1K3I	52,200	727	591	35	10	35	35	35	34	17	5	0
1KEX	52,200	177	141	16	16	16	16	16	16	11	0	0
1SDD	52,200	614	372	59	3	59	59	55	40	12	2	0
1TVG	52,200	193	184	16	16	16	16	16	16	3	0	0
1XPW	52,200	193	192	12	10	12	12	12	12	12	6	0
Laminin γ1	730,800	9830	7306	1349		1349	1334	1159	387	12		
1CZT	52,200	695	370	150	145	150	150	144	53	3		
1D7P	52,200	616	507	146	142	146	146	132	31	1		
1EUT	52,200	509	432	54	34	54	54	31	14	0		
1GOF	52,200	811	657	57	30	57	56	37	11	0		
1GQP	52,200	354	285	180	170	180	80	142	71	1		
1GUI	52,200	409	319	103	102	103	103	96	26	1		
1JHJ	52,200	797	776	48	18	48	48	47	8	0		
1K12	52,200	572	385	171	171	171	171	168	12	1		
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)[a]		
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		

Compactness (according to Table VI) was used as an additional criterion to validate the obtained structures, which fulfill all cross-linking constraints.
[a]Structure did not fulfill the compactness criterion.

made for the C^A–C^A distances of amino acids that are cross-linked with BS^3: 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, 7 of 12 and 5 of 7 constraints have been applied as restraint for laminin β_1 and γ_1, respectively. Nevertheless, as already described by Alexander *et al.*[20] the information content of the distance constraints depends on the sequence separation [aa]/Euclidian separation [Å] ratio.

The structures of laminin β_1 and γ_1 LN domains calculated by Rosetta were compared with the experimentally obtained distance constraints (Table III and Supporting Information Table III). 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 to account for inaccuracies in the comparative models (Table V). Only approximately 1% of all models fulfilled all distance constraints highlighting the discriminative power of even few experimental distance measurements (Table V).

All models fulfilling all distance constraints exhibited the same topology (more than 100 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 (Table VI and Supporting Information Table IV). In the best scoring models of the LN domains of laminin β_1 and γ_1 all of the distance constraints deviate from the theoretical distance by <5 Å (Fig. 5).

For laminin β_1, three structures (Table V) fulfilled all seven distance constraints and exhibited structural similarity, that is, >110 C^A atoms are superimposed with an

Table VI
Composite Scoring Function that Integrates the Deviation of the Experimental Constraints from the Theoretical Maximum Distance as Well as the Rosetta Energy and Compactness of the Structure
Score = $LN\left[\prod\left(\frac{X_{i,max}-X_i}{X_{i,max}-X_{i,opt}}\right)\right]$ 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 $X_i < X_{opt}$ than $\frac{X_{i,max}-X_i}{X_{i,max}-X_{i,opt}}$ was set to 1, where else in the case $X_i > X_{max}$, the term was set to 0.05.

RMSD below 4 Å. For laminin γ_1, five distance constraints were fulfilled by 12 models (Table V). Nine structures could be superimposed in >120 C^A atoms with an RMSD below 4 Å. For the remaining three models, >50 C^A atoms could be superimposed with an RMSD below 4 Å.

LOO cross-validation and evaluation with structure quality

To assess the quality of laminin β_1 and γ_1, all models fulfilling six of seven or four of five distance constraints, respectively, were pairwise compared with the best solution structures for laminin β_1 and γ_1 LN domains (Fig. 5). A test of the precision of the modeling efforts yielded 17 of 33 (laminin β_1) and 141 of 293 models (laminin γ_1) with a high structural similarity, that is, >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 β_1 and γ_1 LN domains, which were obtained with six of seven (four of five) 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 VII). In conclusion, 78% and 61% of the LOO structures fulfill the left-out distance constraint.

In addition, the programs VADAR and MolProbity were used for a Rosetta-independent measurement of structure quality (Supporting Information Table IV). 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.

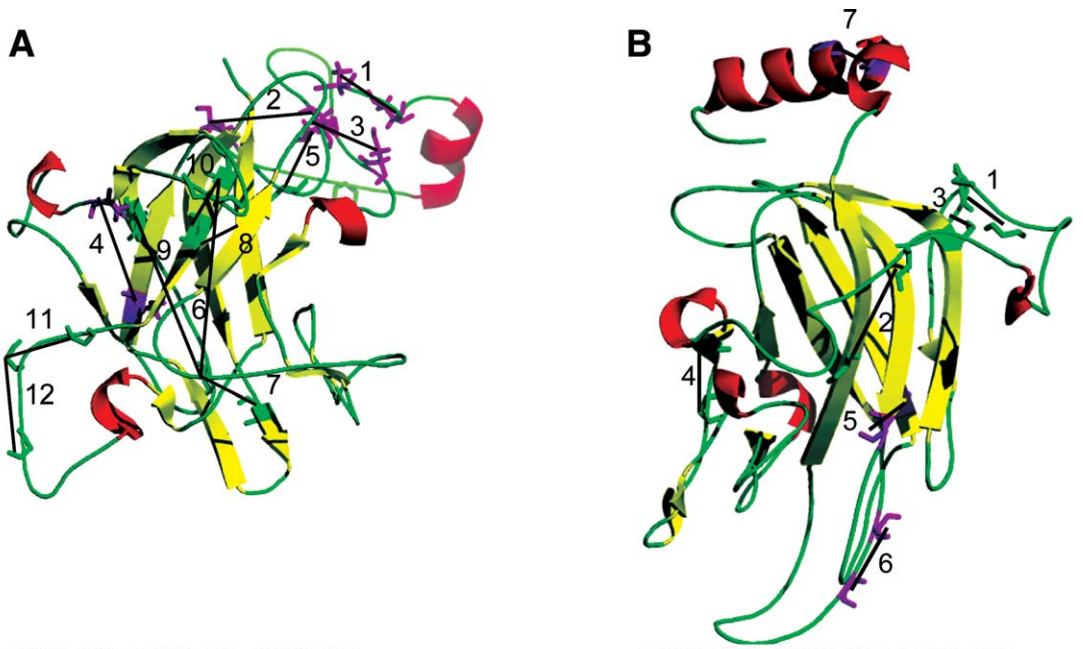

Figure 5
The best scoring Rosetta structures for (A) laminin β_1 and (B) γ_1 LN domains that fulfill all experimental distance constraints. The amino acids involved in the constraints are indicated, and structures are visualized by PyMOL (version 0.99rc6). Cysteines at Positions 13 (laminin β_1) and 18 (laminin γ_1), which are not part of laminin N-terminal domains, are marked with asterisks.

Table VII
Summary of the Filter Function and the *Leave-One-Out* (LOO) Experiments

		Disulfide 1	Disulfide 2	Disulfide 3	Disulfide 4	XL 1	XL 2	XL 3	Mean
Laminin β_1	Filter (total)	484	117	467	47	493	288	435	65%
		95.7%	23.1%	92.3%	9.3%	97.4%	56.9%	86.0%	
	Filter (6/7)	33	27	33	9	33	33	33	84%
		100%	81.8%	100%	27.3%	100%	100%	100%	
	LOO distance	3/3	3/9	3/3	3/27	3/3	3/3	3/3	78%
		100%	33%	100%	11.1%	100%	100%	100%	
	LOO fold	3/3	6/9	3/3	14/27	3/3	3/3	3/3	88%
		100%	66%	100%	52%	100%	100%	100%	
	Information content ($Å^{-1}$)	0.38	0.69	0.23	1.54	0.37	1.90	0.47	
Laminin γ_1	Filter (total)	1226	260	1348	218	1193			63%
		90.8%	19.3%	99.9%	16.1%	88.4%			
	Filter (4/5)	291	161	293	148	291			81%
		99.3%	54.9%	100.0	50.5%	99.3%			
	LOO distance	12/14	12/133	12/12	12/145	12/12			61%
		86%	9%	100%	8%	100%			
	LOO fold	9/14	68/133	9/12	82/145	9/12			64%
		64%	51%	75%	57%	75%			
	Information content ($Å^{-1}$)	0.77	0.62	0.23	1.31	0.43			

"Filter total", "filter (6/7)", and "filter (4/5)" indicate how many models in total and how many models that fulfill six of seven, respectively, four of five 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 >8; information content measure [sequence separation (aa)/Euclidian separation (Å)] of the distance constraints. XL: Cross-link.

Influence of templates

To analyze the influence of the template on the final model best structures were compared for the 14 templates (Table IV). For laminin γ_1, for all 14 templates structures were obtained, which fulfilled at least 4 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 of seven distance constraints. All displayed the identical fold as in the overall best structure (Supporting Information Table II).

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 VII). 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 γ1 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 reported 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, MS, 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, that is, disulfide bonds, and chemical cross-links created by reaction with two amine-reactive cross-linkers, and parameters for model quality, that is, 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 LOO cross-validation combined with a structure–structure alignment as well as by Rosetta-independent structure evaluation programs.

Although none of the techniques used herein is completely new, we argue that the presented protocol is novel as it combines 10 computational methods with a complex experimental setup integrating chemical cross-linking, isotope labeling, and LC/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.

ACKNOWLEDGMENTS

SK acknowledges support from the DFG-funded Graduiertenkolleg 1026 "Conformational Transitions in Macromolecular Interactions" at the Martin Luther University Halle-Wittenberg. 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. Mathias Q. Müller is acknowledged for graphical assistance.

REFERENCES

1. Tunggal P, Smyth N, Paulsson M, Ott M-C. Laminins: structure and genetic regulation. Microsc Res Tech 2000;51:214–227.
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. A simplified laminin nomenclature. Matrix Biol 2005;24:326–332.
3. Odenthal U, Haehn S, Tunggal P, Merkl B, Schomburg D, Frie C, Paulsson M, Smyth N. Molecular analysis of laminin N-terminal domains mediating self-interactions. J Biol Chem 2004;279:44504–44512.
4. Yurchenco PD, Tsilibary EC, Charonis AS, Furthmayr H. Laminin polymerization in vitro. Evidence for a two-step assembly with domain specificity. J Biol Chem 1985;260:7636–7644.
5. McKee KK, Harrison D, Capizzi S, Yurchenco PD. Role of laminin terminal globular domains in basement membrane assembly. J Biol Chem 2007;282:21437–21447.
6. Schneiders FI, Maertens B, Böse K, Li Y, Brunken WJ, Paulsson M, Smyth N, Koch M. Binding of Netrin-4 to laminin short arms regulates basement membrane assembly. J Biol Chem 2007;282:23750–23758.
7. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. Proc Natl Acad Sci USA 2000;97:5802–5806.
8. Back JW, Jong LD, Muijsers AO, Koster CGD. Chemical cross-linking and mass spectrometry for protein structural modeling. J Mol Biol 2003;331:303–313.
9. Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. J Mass Spectrom 2003;38:1225–1237.
10. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom Rev 2006;25:663–682.
11. Hermanson G, editor. Bioconjugate techniques, 2nd ed. Amsterdam, NL: Elsevier; 2008.
12. Kalkhof S, Haehn S, Ihling C, Paulsson M, Smyth N, Sinz A. Determination of disulfide bond patterns in laminin beta1 chain N-terminal domains by nano-high-performance liquid chromatography/matrix-assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2008;22:1933–1940.
13. Kalkhof S, Ihling C, Mechtler K, Sinz A. Chemical cross-linking and high-performance Fourier transform ion cyclotron resonance mass spectrometry for protein interaction analysis: application to a calmodulin/target peptide complex. Anal Chem 2004;77:495–503.
14. Schmidt A, Kalkhof S, Ihling C, Cooper D, Sinz A. Mapping protein interfaces by chemical cross-linking and Fourier transform ion cyclotron resonance mass spectrometry: application to a calmodulin/adenylyl cyclase 8 peptide complex. Eur J Mass Spectrom 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. Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex. J Am Soc Mass Spectrom 2006;17:1100–1113.
16. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. J Mol Biol 1997;268:209–225.
17. Bradley P, Malmström L, Qian B, Schonbrun J, Chivian D, Kim DE, Meiler J, Misura KMS, Baker D. Free modeling with Rosetta in CASP6. Prot Struct Funct Bioinform 2005;61:128–134.
18. Bowers PM, Strauss CEM, Baker D. De novo protein structure determination using sparse NMR data. J Biomol NMR 2000;18:311–318.
19. Meiler J, Baker D. The fumarate sensor DcuS: progress in rapid protein fold elucidation by combining protein structure prediction methods with NMR spectroscopy. J Magn Reson 2005;173:310–316.
20. Alexander N, Al-Mestarihi A, Bortolus M, McHaourab H, Meiler J. De novo high-resolution protein structure determination from sparse spin-labeling EPR data. Structure 2008;16:181–195.
21. Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. Anal Bioanal Chem 2008;392:305–312.
22. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–685.
23. Schulz DM, Kalkhof S, Schmidt A, Ihling C, Stingl C, Mechtler K, Zschörnig O, Sinz A. Annexin A2/P11 interaction: new insights into annexin A2 tetramer structure by chemical crosslinking, high-reso-

lution mass spectrometry, and computational modeling. Prot Struct Funct Bioinform 2007;69:254–269.
24. Mädler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. J Mass Spectrom 2009;44:694–706.
25. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
26. Huang X, Miller W. A time-efficient linear-space local similarity algorithm. Adv Appl Math 1991;12:337–357.
27. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25:3389–3402.
28. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–371.
29. Juan D, Graña O, Pazos F, Fariselli P, Casadio R, Valencia A. A neural network approach to evaluate fold recognition results. Prot Struct Funct Genet 2003;50:600–608.
30. Torda AE, Procter JB, Huber T. Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. Nucl Acids Res 2004;32:W532–W535.
31. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics 2005;21:951–960.
32. Teodorescu O, Galor T, Pillardy J, Elber R. Enriching the sequence substitution matrix by structural information. Prot Struct Funct Bioinform 2004;54:41–48.
33. Ortiz AR, Strauss CEM, Olmea O. MAMMOTH (Matching molecular models obtained from theory): an automated method for model comparison. Prot Sci 2002;11:2606–2621.
34. Dong E, Smith J, Heinze S, Alexander N, Meiler J. BCL::Align–sequence alignment and fold recognition with a custom scoring function online. Gene 2008;422:41–46.
35. Meiler J, Prompers JJ, Peti W, Griesinger C, Brüschweiler R. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. J Am Chem Soc 2001;123:6098–6107.
36. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
37. Karplus K, Barrett C, Cline M, Diekhans M, Grate L, Hughey R. Predicting protein structure using only sequence information. Prot Struct Funct Genet 1999;37:121–125.
38. Chivian D, Baker D. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. Nucl Acids Res 2006;34:e112.
39. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Prot Struct Funct Bioinform 2004;55:656–677.
40. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. Nucl Acids Res 2003;31:3316–3319.
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. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acids Res 2007;35:W375–W383.
42. Schnaible V, Wefing S, Resemann A, Suckau D, Bücker A, Wolf-Kümmeth S, Hoffmann D. Screening for disulfide bonds in proteins by MALDI in-source decay and LIFT-TOF/TOF-MS. Anal Chem 2002;74:4980–4988.
43. Rost B, O'Donoghue S. Sisyphus and prediction of protein structure. Comput Appl Biosci 1997;13:345–356.
44. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA 2000;97:10383–10388.
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. Mutations in lama1 disrupt retinal vascular development and inner limiting membrane formation. J Biol Chem 2010;285:7697–7711.
46. Smyth N, Vatansever HS, Murray P, Meyer M, Frie C, Paulsson M, Edgar DJ. Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation. Cell Biol 1999;144:151–160.
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. Crystal structures of the membrane-binding C2 domain of human coagulation factor V. Nature 1999;402:434–439.
48. Pratt KP, Shen BW, Takeshima K, Davie EW, Fujikawa K, Stoddard BL. Structure of the C2 domain of human factor VIII at 1.5Å resolution. Nature 1999;402:439–442.
49. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure 1995;3:1197–1205.
50. Ito N, Phillips SEV, Stevens C, Ogel ZB, McPherson MJ, Keen JN, Yadav KDS, Knowles PF. Novel thioether bond revealed by a 1.7 [angst] crystal structure of galactose oxidase. Nature 1991;350:87–90.
51. Au SWN, Leng X, Harper JW, Barford D. Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit. J Mol Biol 2002;316:955–968.
52. Boraston AB, Nurizzo D, Notenboom V, Ducros V, Rose DR, Kilburn DG, Davies GJ. Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan-binding modules. J Mol Biol 2002;319:1143–1156.
53. Wendt KS, Vodermaier HC, Jacob U, Gieffers C, Gmachl M, Peters J-M, Huber R, Sondermann P. Crystal structure of the APC10/DOC1 subunit of the human anaphase-promoting complex. Nat Struct Biol 2001;8:784–788.
54. Bianchet MA, Odom EW, Vasta GR, Amzel LM. A novel fucose recognition fold involved in innate immunity. Nat Struct Biol 2002;9:628–634.
55. Firbank SJ, Rogers MS, Wilmot CM, Dooley DM, Halcrow MA, Knowles PF, McPherson MJ, Phillips SEV. Crystal structure of the precursor of galactose oxidase: an unusual self-processing enzyme. Proc Natl Acad Sci USA 2001;98:12932–12937.
56. Lee CC, Kreusch A, McMullan D, Ng K, Spraggon G. Crystal structure of the human neuropilin-1 b1 domain. Structure 2003;11:99–108.
57. Adams TE, Hockin MF, Mann KG, Everse SJ. The crystal structure of activated protein C-inactivated bovine factor Va: implications for cofactor function. Proc Natl Acad Sci USA 2004;101:8918–8923.
58. Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma L-C, Acton TB, Hunt JF, Montelione GT, Baker D, Kennedy MA. Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. Prot Struct Funct Bioinform 2009;75:147–167.
59. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. Nat Struct Mol Biol 1999;6:884–893.
60. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nat Struct Mol Biol 1996;3:638–648.
61. Joint Center For Structural Genomics. Crystal structure of hypothetical protein (TM1631) from Thermotoga maritima at 2.20 Å resolution. PDB entry 1VPQ. http://www.rcsb.org/pdb/explore/explore.do?structureId=1VPQ
62. Guo H-C, Jardetzky TS, Garrettt TPJ, Lane WS, Strominger JL, Wiley DC. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. Nature 1992;360:364–366.
63. Cardoso RMF, Silva CHTP, Ulian de Araujo AP, Tanaka T, Tanaka M, Garratt RC. Structure of the cytosolic Cu,Zn superoxide dismutase from Schistosoma mansoni. Acta Crystallogr D Biol Crystallogr 2004;60:1569–1578.
64. Setayesh FR, DeCorte BL, Horton P, Harris CM, Harris TM, Stone MP. Styrene oxide adducts in an oligodeoxynucleotide containing the human N-ras Codon 12: minor groove structures of the R(12,1)- and S(12,1)-Î±-(N2-Guanyl) stereoisomers determined by 1H nuclear magnetic resonance. Chem Res Toxicol 1998;11:766–777.

65. Momma K, Mishima Y, Hashimoto W, Mikami B, Murata K. Direct evidence for Sphingomonas sp. A1 periplasmic proteins as macromolecule-binding proteins associated with the ABC transporter: molecular insights into alginate transport in the periplasma. Biochemistry 2005;44:5053–5064.
66. Joint Center For Structural Genomics. Crystal structure of inositol-3-phosphate synthase (ce21227) from Caenorhabditis elegans at 2.30 Å resolution. PDB entry 1VKO. http://www.rcsb.org/pdb/explore/explore.do?structureId=1VKO
67. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang B-C, Wilson KP. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. J Biol Chem 2004;279:31655–31663.
68. Skelton NJ, Quan C, Reilly D, Lowman H. Structure of a CXC chemokine-receptor fragment in complex with interleukin-8. Structure 1999;7:157–168.
69. Li H, Tomizawa T, Tochio N, Muto Y, Koshiba S, Inoue M, Kigawa T, Yokoyama S. Solution structure of the C-terminal PH domain of FYVE, RhoGEF and PH domain containing protein 3 (FGD3) from human. PDB entry 2COC. http://www.rcsb.org/pdb/explore/explore.do?structureId=2COC
70. Udomsinprasert R, Pongjaroenkit S, Wongsantichon J, Oakley AJ, Prapanthadara L-A, Wilce MCJ, Ketterman AJ. Identification, characterization and structure of a new Delta class glutathione transferase isoenzyme. Biochem J 2005;388:763–771.
71. Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K. Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. J Mol Biol 2000;300:889–901.
72. López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-I, Kaltner H, Gabius H-J, Romero A. Growth-regulatory human Galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. J Mol Biol 2004;343:957–970.
73. Katsuya Y, Mezaki Y, Kubota M, Matsuura Y. Three-dimensional structure of Pseudomonas isoamylase at 2.2 Å resolution. J Mol Biol 1998;281:885–897.
74. Koch AW, Farooq A, Shan W, Zeng L, Colman DR, Zhou M-M. Structure of the neural (N-) cadherin prodomain reveals a cadherin extracellular domain-like fold without adhesive characteristics. Structure 2004;12:793–805.
75. Abad MC, Arni RK, Grella DK, Castellino FJ, Tulinsky A, Geiger JH. The X-ray crystallographic structure of the angiogenesis inhibitor angiostatin. J Mol Biol 2002;318:1009–1017.
76. Llera AS, Viedma F, Sánchez-Madrid F, Tormo J. Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. J Biol Chem 2001;276:7312–7319.
77. Baeyens KJ, De Bondt HL, Raeymaekers A, Fiers W, De Ranter CJ. The structure of mouse tumour-necrosis factor at 1.4 Å resolution: towards modulation of its selectivity and trimerization. Acta Crystallogr D Biol Crystallogr 1999;55:772–778.
78. Lubkowski J, Hennecke F, Pluckthun A, Wlodawer A. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nat Struct Mol Biol 1998;5:140–147.
79. Tegoni M, Spinelli S, Verhoeyen M, Davis P, Cambillau C. Crystal structure of a ternary complex between human chorionic gonadotropin (hCG) and two Fv fragments specific for the α and β-subunits. J Mol Biol 1999;289:1375–1385.
80. Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS, Hinck AP. Crystal structure of the human TβR2 ectodomain-TGF-β3 complex. 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. Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: molecular basis of isozyme-drug discrimination. Prot Sci 1998;7:556–563.
82. Lamb AL, Wernimont AK, Pufahl RA, O'Halloran TV, Rosenzweig AC. Crystal structure of the second domain of the human copper chaperone for superoxide dismutase. Biochemistry 2000;39:1589–1595.
83. Whittingtons DA, Grubb JH, Waheed A, Shah GN, Sly WS, Christianson DW. Expression, assay, and structure of the extracellular domain of murine carbonic anhydrase XIV. J Biol Chem 2004;279:7223–7228.
84. Dayan SM, Van Donkelaar A, Kortt AA. Crystallization and preliminary crystallographic data of the major albumin from Psophocarpus tetragonolobus (L.) DC. J Biol Chem 1987;262:10287–10289.
85. Watson JN, Newstead S, Dookhun V, Taylor G, Bennet AJ. Contribution of the active site aspartic acid to catalysis in the bacterial neuraminidase from Micromonospora viridifaciens. J. FEBS Lett 2004;577:265–269.