J Biomol NMR (2009) 43:79–96
DOI 10.1007/s10858-008-9290-y

ARTICLE

NMR-based homology model for the solution structure of the C-terminal globular domain of EMILIN1

Giuliana Verdone · Alessandra Corazza · Simon A. Colebrooke · Daniel Cicero · Tommaso Eliseo · Jonathan Boyd · Roberto Doliana · Federico Fogolari · Paolo Viglino · Alfonso Colombatti · Iain D. Campbell · Gennaro Esposito

Received: 26 June 2008 / Accepted: 3 November 2008 / Published online: 21 November 2008

Abstract EMILIN1 is a glycoprotein of elastic tissues that has been recently linked to the pathogenesis of hypertension. The protein is formed by different independently folded structural domains whose role has been partially elucidated. In this paper the solution structure, inferred from NMR-based homology modelling of the C-terminal trimeric globular C1q domain (gC1q) of EMILIN1, is reported. The high molecular weight and the homotrimeric structure of the protein required the combined use of highly deuterated ^{15}N, ^{13}C-labelled samples and TROSY experiments. Starting from a homology model, the protein structure was refined using heteronuclear residual dipolar couplings, chemical shift patterns, NOEs and H-exchange data. Analysis of the gC1q domain structure of EMILIN1 shows that each protomer of the trimer adopts a nine-stranded β sandwich folding topology which is related to the conformation observed for other proteins of the family. Distinguishing features, however, include a missing edge-strand and an unstructured 19-residue loop. Although the current data do not allow this loop to be precisely defined, the available evidence is consistent with a flexible segment that protrudes from each subunit of the globular trimeric assembly and plays a key role in intermolecular interactions between the EMILIN1 gC1q homotrimer and its integrin receptor $\alpha4\beta1$.

Keywords Elastic fibres · EMILIN1 · Globular C1q domain · Homotrimeric proteins · Large system NMR · Partially deuterated proteins

Abbreviations

ACRP30	Adipocyte complement-related protein of 30 kDa
CSA	Chemical shift anisotropy
DD	Dipole–dipole
EMILIN1	Elastin microfibril interface located protein 1
EMILIN2	Elastin microfibril interface located protein 2
gC1q	Globular C1q domain
IPAP	In-phase anti-phase
NOE	Nuclear overhauser effect
RDC	Residual dipolar couplings
TGF	Tumor growth factor
TNF	Tumor necrosis factor

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9290-y) contains supplementary material, which is available to authorized users.

G. Verdone · A. Corazza · F. Fogolari · P. Viglino · A. Colombatti · G. Esposito (✉)
Dipartimento di Scienze e Tecnologie Biomediche – MATI Centre of Excellence, Università di Udine, P. le Kolbe, 4-33100 Udine, Italy
e-mail: gesposito@mail.dstb.uniud.it

Present Address:
G. Verdone
Istituto Biochimico Italiano 'G. Lorenzini', 04011 Aprilia, LT, Italy

A. Corazza · F. Fogolari · P. Viglino · G. Esposito
INBB, Viale Medaglie d'Oro, 305, Rome, Italy

S. A. Colebrooke · J. Boyd · I. D. Campbell
Department of Biochemistry, University of Oxford, Oxford, UK

D. Cicero · T. Eliseo
Dipartimento di Chimica, Università di Tor Vergata, Rome, Italy

R. Doliana · A. Colombatti
Divisione di Oncologia Sperimentale 2, Centro di Riferimento Oncologico di Aviano, Aviano, PN, Italy

Introduction

EMILIN (Elastic Microfibril Interface Located ProteIN) is an extra-cellular matrix glycoprotein originally isolated from chicken aorta in 1983 (Bressan et al. 1983). Abundant in blood vessels, the protein was also found in connective tissues of different organs, always in association with elastic fibres containing elastin (Colombatti et al. 2000). Cloning the cDNA of chicken EMILIN led to the isolation of the human and mouse gene and allowed the identification of three homologous proteins that constitute the protein family of EMILINs. In addition to the first isolated EMILIN, later renamed EMILIN1, the family includes EMILIN2 (Doliana et al. 2001), Multimerin or EMILIN-3 (Hayward et al. 1995) and Endoglyx-1 or Multimerin 2 (Christian et al. 2001). The exact function of EMILINs remained largely unknown until recently when it was found that the secreted protein EMILIN1 is a negative regulator, through the N-terminal domain, of TGF-β signalling, a pathway that has a major role in blood pressure homeostasis (Zacchigna et al. 2006). EMILIN1-deficient mice, in fact, have increased levels of active TGF-β in the vasculature which results in smaller blood vessels, increased peripheral resilience and hypertension (Danussi et al. 2008). Moreover, for the C-terminal domain of EMILIN1 a functional involvement in cell migration has been observed (Spessotto et al. 2006). Human EMILIN1 protein, as well as its homologues, has a modular organization (Doliana et al. 2001). Due to the presence of the globular C1q domain (gC1q), EMILINs belong to the large C1q/TNF superfamily of proteins. A hallmark of gC1q domains is the promotion of trimeric oligomerization of the molecules in which they occur (Kishore et al. 2004).

X-ray crystallography has been used so far to derive structures of the isolated C-terminal gC1q domains of four proteins of the family, namely the mouse ACRP30 (Shapiro and Scherer 1998), human collagen X (Bogin et al. 2002), human collagen VIII (Kvansakul et al. 2003) and human complement C1q (Gaboriaud et al. 2003). These proteins show similar quaternary structures obtained by the non-covalent association of three identical polypeptide chains (homotrimer), or three different chains (heterotrimer) in the case of human complement C1q. The analogy among the members of the family also extends to the tertiary structure of each protomer that adopts a β sandwich topology formed by two-five-stranded β sheets of anti-parallel β-strands.

As expected, the isolated human EMILIN1 gC1q domain also forms homotrimers, with a strong prevalence of β secondary structure elements (Mongiat et al. 2000; Verdone et al. 2004). The EMILIN1 gC1q domain has been identified as a fundamental component of the whole molecule, both from a structural and a functional viewpoint, for two main reasons: first, it is the domain that is able to initiate and drive the trimerization of the whole protein; second, the domain has recognized cell-adhesion and cell-migration stimulating properties that are mediated by interactions with $\alpha 4\beta 1$ integrin, a trans-membrane receptor predominantly expressed on the surface of hemopoietic cells (Spessotto et al. 2003; Spessotto et al. 2006).

To outline the specific structural details that determine those functions, we undertook the NMR characterisation of the molecule. Following the resonance assignment and preliminary secondary structure identification (Verdone et al. 2004), we now present the three-dimensional structure of the EMILIN1 gC1q trimer, obtained by NMR-restrained homology modelling, and a description of the methodology adopted. The functional inferences that were obtained from structure modelling were successfully tested in mutagenesis experiments (Verdone et al. 2008). The EMILIN1 gC1q model reported in this study is the first NMR-based determination of a gC1q domain; at $\sim$52 kDa this is technically challenging for NMR. Due to the substantial lack of NOE-based distance restraints for structure calculation, the definition of the molecule tertiary and quaternary structure was obtained by a combination of homology modelling and structure refinement with heteronuclear residual dipolar coupling restraints measured in anisotropic gel phase.

Materials and methods

Sample preparation

The recombinant gC1q domain of EMILIN1 was expressed as $(His)_6$-tagged protein and purified as previously described (Mongiat et al. 2000). It is a polypeptide chain of 162 amino acids, 150 from the natural domain sequence, seven from a N-terminally fused histidine tag and initial Met residue and five derived from the cloning strategy, i.e. MetArgGlySer$(His)_6$GlySer, at the N-terminal side of the 150-residue EMILIN1 gC1q domain. The recombinant product forms a stable trimeric protein aggregate with an acidic isoelectric point (pI: 5.32) and a molecular weight of 51,624 Da. The increase of 1,151 Da compared to the value formerly published (Verdone et al. 2004) arises from correcting the primary sequence that had been used for calculating the molecular weight. The correct sequence was obtained by nucleotide sequence and mass spectrometry on the product without isotopic enrichment. After an initial screening for the best solubility conditions, a 20 mM phosphate, 100 mM NaCl, 0.02% (w/v) NaN_3, 5% (v/v) D_2O, pH = 7.5 solution was used to dissolve the protein, yielding a clear sample which proved to be stable for

several months. A 1.8 mM protomer concentration sample (0.6 mM homotrimer concentration) of U-^{13}C, U-^{15}N, 80% ^{2}H labelled protein was used for the majority of the NMR experiments. Two more samples with different isotopic enrichment, 1 mM in protein concentration, were employed in the same buffer using U-^{15}N, 70% ^{2}H labelled protein, and U-^{13}C, U-^{15}N labelled protein. All labelled protein samples were purchased as lyophilized powders from Asla Biotech (http://www.asla-biotech.com).

Multidimensional NMR spectroscopy

NMR experiments were performed at 310 K on the Oxford laboratory home-built/GE-Omega spectrometers operating at field strengths of 11.7, 14.1 and 17.6 T, i.e. 500, 600 or 750 MHz proton resonance frequency, respectively, and on the Udine laboratory Bruker Avance 500 NMR system. Various ^{1}H,^{15}N-HSQC experiments (Palmer et al. 1991; Kay et al. 1992) with different spectral widths were recorded in order to establish the range of amide resonances (see "Supplementary material"). Backbone sequential assignment was obtained through a *suite* of triple resonance TROSY experiments (Salzmann et al. 1998, 1999) as detailed in "Supplementary material". Three dimensional spectra were typically acquired with 64 × 48 × 512 complex data points and 32–96 scans/t1.

Multi-dimensional datasets were recorded in a phase-sensitive manner with quadrature detection by the States/TPPI (States et al. 1982; Marion and Wüthrich 1983) and echo-antiecho (Davis et al. 1992) method in the indirectly detected dimensions. The ^{1}H carrier frequency was set to the water resonance and broadband decoupling of the heteronuclei during acquisition was achieved using GARP or WALTZ-16 sequences at decoupling field strengths corresponding to $\gamma_H B_1/2\pi$ of 833 Hz (^{15}N) and 2,500 Hz (^{13}C) (Shaka et al. 1983, 1985). ^{2}H decoupling, using a bandwidth of 385 Hz centred at 115.132091 MHz (^{2}H frequency) and a WALTZ modulation scheme, was also applied during the ^{13}C evolution period of all the experiments recorded on the 750 MHz spectrometer (^{1}H frequency). All the experiments employed a gradient enhanced scheme (Kay et al. 1992), a constant time period for the ^{15}N evolution time (21.8 ms) and an e-snob selective pulse (Boyd and Soffe 1989) to align water magnetization along the $+z$ axis before acquisition. Gaussian shaped pulses (G3) were normally used to obtain selective excitation of ^{13}C′ and 13C$^{\alpha}$ nuclei. ^{1}H, ^{13}C and ^{15}N chemical shifts for the EMILIN1 gC1q domain have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the BMRB accession number 5882.

NOE connectivities were identified from the analysis of a 3D ^{15}N-NOESY-HSQC spectrum (Marion et al. 1989a, b) recorded at 600 MHz (^{1}H frequency) with a mixing time of 150 ms and acquisition times of 24.9 ms (t_1, ^{1}H), 8.8 ms (t_2, ^{15}N) and 57.3 ms (t_3, ^{1}H). Spectral widths of 8,928.57 Hz for both proton dimensions and 2,040.81 Hz for the ^{15}N dimension were used.

A limited number of H$^{\alpha}$ resonances were assigned using 3D-HCCH-COSY (Bax et al. 1990a) and 3D-HCCH-TOCSY (Bax et al. 1990b; Kay et al. 1993). The TOCSY experiment employed a heteronuclear isotropic mixing performed with a 9.4 kHz DIPSI-3 field (Shaka et al. 1988; Kay et al. 1993) over a mixing time of 5.4 ms resulting in transfer through one bond.

^{15}N{^{1}H} NOE measurements (Grzesiek and Bax 1993) at 11.4 T were also performed, with relaxation or saturation delay of 3 and 5 s, 160 t1 increments of 1,024 complex data points and 480 scans/t1.

Data were processed with Felix version 2.3 and 2 k (Accelrys), and analysed with Xeasy (Bartels et al. 1995) and Sparky (T. D. Goddard and D. G. Kneller, University of California). Original sizes were normally increased using mirror-image linear prediction (^{15}N indirect dimension) and backward–forward linear prediction (^{13}C indirect dimension). Zero filling to double the obtained sizes followed by apodization with shifted sine bell functions was applied prior to Fourier transformation. Proton chemical shifts were referenced to dioxane, whose resonance was set to 3.750 ppm while ^{13}C and ^{15}N chemical shifts were referenced indirectly to dioxane, using the absolute frequency ratio (Wishart et al. 1995).

Amide hydrogen exchange

The observation of slow exchange amide hydrogens was performed on the U-^{15}N, 70% ^{2}H labelled C1q sample, freeze-dried from aqueous buffer solution and reconstituted in D_2O. After an initial delay of 30 min, consecutive ^{1}H,^{15}N-HSQC experiments (Palmer et al. 1991; Kay et al. 1992) were recorded to monitor the gradual disappearance of peaks due to the replacement of protons by deuterons. In total, 24 ^{1}H,^{15}N-HSQC spectra were recorded over 68 h.

Residual dipolar coupling measurements

The experiments for RDC determinations were collected using a low-density (3–4% w/v) polyacrylamide gel medium radially compressed to align the protein (Sass et al. 2000; Chou et al. 2001). Heteronuclear one-bond RDCs were obtained as a contribution to existing scalar couplings, comparing peak splitting of spectra recorded in isotropic and anisotropic solutions of the protein. A set of 123 residual dipolar couplings for NH vectors ($^{1}D_{NH}$) was measured using an IPAP approach (Ottiger et al. 1998). The same gel sample was used to acquire a 3D-HNCO

(Grzesiek and Bax 1992) and TROSY ^{1}H-^{15}N HSQC or 2D HNCO (Wang et al. 1998; Yang et al. 1998) spectra for measuring residual dipolar couplings for CO–C$^{\alpha}$ vectors ($^{1}D_{C'C\alpha}$) and CO–N vectors ($^{1}D_{C'N}$), respectively. Using the recorded spectra, 101 $^{1}D_{C'N}$ and 123 $^{1}D_{C'C\alpha}$ couplings values could be measured. After confidence limit assessment (Bax et al. 2001) and exclusion of the unstructured segment data, only a restricted number of RDC values (105 $^{1}D_{NH}$, 70 $^{1}D_{C'N}$, 59 $^{1}D_{C'C\alpha}$) were retained for subsequent restraining purposes. Further details are given in "Supplementary material".

Homology model

The primary sequence of Emilin gC1q domain was aligned with the primary sequence of 11 human proteins of the C1q protein family (Multimerin, Collagen X α1, Collagen VIII α1, Collagen VIII α2, Cerebellin 1, C1q-A, C1q-B, C1q-C, ACRP30-A, ACRP30-B, ACRP30-C) using the multiple sequence alignment program ClustalW (Thompson et al. 1994). The aligned sequences are reported in "Supplementary material". Among the proteins of the alignment the A-chain of ACRP30 protein (ACRP30-A) (Shapiro and Scherer 1998) was recognized as the best sequence matching the EMILIN1 gC1q domain, taking into account both the pairwise alignment score and the number of matched amino acids. 20% of sequence identity and 31.3% of sequence similarity were found on comparing the query protein and ACRP30-A polypeptide chain, after introducing in the query sequence 6 gaps, 2 three- and four-residue insertions and 1 twelve-residue insertion. The alignment with ACRP30-A polypeptide chain was used to generate a homology model for one monomer of the EMILIN1 gC1q domain with the program GeneMine, available at the time this work was started at URL http://www.mag.com. Residues in insertions that did not have a match resulted in loop and C-terminal regions with standard backbone and side-chain geometries. The structure of the C1q domain trimer was obtained by superposition of three identical monomers onto the template's three subunits (chain A, chain B and chain C), followed by regularization and minimization of the whole assembly. This regularization-minimization step, executed with the program Xplor, involved a short low-temperature dynamics run, followed by energy minimization performed imposing a non-crystallographic symmetry (NCS) term with a force constant of 5 kcal mol^{-1} Å^{-2}.

Refinement protocol

Restrained molecular dynamics calculations were performed using the software Xplor NIH that applies a two-stage simulated annealing protocol with ramped potential term constants (Chou et al. 2000). In the first stage the system was cooled from 200 to 20 K, with a temperature step of 10 K and 6.7 ps of Verlet dynamics at each temperature, using a time step of 3 fs. Experimental data introduced at this stage included: backbone dihedral torsion angles, three classes of RDCs ($^{1}D_{NH}$, $^{1}D_{C'C\alpha}$ and $^{1}D_{C'N}$), short and medium-range backbone NOEs. In order to ensure global compactness of the protein, a pseudo-potential for the radius of gyration of the trimer calculated by Xplor from the number of residues in the polypeptide chains was also applied. This gyration radius (22.4 Å) is consistent with the hydrodynamic radius (32.6–31.1 Å) inferred from the experimental ^{15}N{^{1}H} NOE data (for a compact sphere the gyration radius is $0.6^{1/2}$ times the geometric radius). The majority of the force constants for the different energy terms (van der Waals, bond lengths, bond and torsion angles, specific Ramachandran potential, improper torsion angles) were ramped during the annealing, as generally performed in NMR structure calculation procedures. In particular the constants were set as: k (van der Waals) = 0.002–4.0 kcal mol^{-1} Å^{-4}; k (improper dihedrals) = 0.1–1.0 kcal mol^{-1} deg^{-2}; k (bond angles) = 0.4–1.0 kcal mol^{-1} deg^{-2}; k (Ramachandran) = 0.2–2.0; k (NOEs) = 2–20 kcal mol^{-1} Å^{-2}, according to the original recipe (Chou et al. 2000). For the RDC restraints, the originally proposed k ramping from 5×10^{-4} to 0.5 kcal mol^{-1} Hz^{-2} was changed into 0.002–0.4 kcal mol^{-1} Hz^{-2} for $^{1}D_{NH}$ couplings, and similarly, with a multiplication factor of 0.4 and 0.1 for $^{1}D_{C'C\alpha}$ and $^{1}D_{C'N}$ couplings, respectively. The size of the force constants for the dipolar term was empirically optimised in order to have an RMSD between observed and calculated values comparable to the measurement error, taking into account the normalization factors to $^{1}D_{NH}$ couplings used for $^{1}D_{C'C\alpha}$ and $^{1}D_{C'N}$ couplings. The simulated annealing refinement was restrained also with an explicit hydrogen bond term, with a k (H bond) ramped in the range 20–2 kcal mol^{-1} Å^{-2}. No such restraint was introduced for the slow-exchanging amides of the molten F strand region due to the uncertainty about the acceptor nuclei. No ramping was applied to the gyration radius force constant (50 kcal mol^{-1} Å^{-2}). A combination of TALOS predictions and homology model-derived dihedral angle restraints was employed only in regions of the molecule where secondary structure elements are located without constraining loop segments. When TALOS offered a valid prediction that was clearly different from φ and ψ angles in the homology model, we used TALOS information. For all the other residues, dihedral angles derived from the homology model were introduced. During the first stage of molecular dynamics, the predominant experimental term was the dihedral restraint term, which maintained the molecular

backbone quite rigid by means of a high force constant that was kept unramped ($k_{dihedral}$ = 300 kcal mol^{-1} rad^{-2}). The best structures obtained after this step still had a considerably high RDC-restraint energy term which was lowered afterwards. During the second stage, the lowest RDC restraint energy structure obtained from step 1 was subjected to a second simulated annealing run, performed at lower temperature. Throughout this stage the dihedral force constant was gradually lowered (from 300 to 50 kcal mol^{-1} rad^{-2}), while the dipolar restraint force constants and the other previously ramped k values were kept at the maximum value reached in stage one, and larger changes were allowed in the structure in order to satisfy the dipolar restraints. The second step of the simulated annealing was performed without ramping any k value that remained fixed at the final value reached at the end of the first step. The structures were further refined in order to improve the quality of backbone geometry. For this purpose 2,000 steps of restrained energy minimization were performed using the program NAMD (Kale et al. 1999). The forcefield used is CHARMM v. 27b (MacKerell et al. 1998) with the CMAP correction (MacKerell et al. 2004). Since the program does not readily incorporate RDC derived restraints, the *z* co-ordinates of the backbone amide N and H atoms were restrained at their starting values. Only NOE derived restraints and chemical-shift-derived dihedral angle restraints (with unequivocal secondary structure definition) were imposed. All structures were drawn using MOLMOL software (Koradi et al. 1996).

Results

NMR study feasibility

The isolated EMILIN1 gC1q domain forms a trimeric non-covalent complex in solution (Doliana et al. 1999). The NMR spectra obtained from the EMILIN1 gC1q domain are generally of fairly good quality, for an assembly of 51,624 Da although it was necessary to use samples extensively isotope labelled as well as experiments specifically developed for the analysis of large molecules at high magnetic fields, (Pervushin et al. 1997). The TROSY ^{1}H–^{15}N HSQC spectrum of EMILIN1 gC1q domain recorded at 17.6 T (Fig. 1) shows a significant improvement compared to the analogous spectrum recorded using a classical ^{1}H–^{15}N HSQC pulse sequence (shown in "Supplementary material"). The number of resonances for N–H groups observed in the spectrum shown in Fig. 1 is consistent with the 136 non-proline residues of a single EMILIN1 gC1q subunit. The polypeptide chain of the recombinant EMILIN1 gC1q protein contains 162 amino acids, of which 14 are proline residues, and 12 belong to an artificial N-terminal tail that originates from the expression and purification protocol and is expected to be unstructured and spectroscopically highly degenerate. Poor peak dispersion in the central part of the TROSY ^{1}H–^{15}N HSQC spectrum suggests the presence of some unstructured segments in the protein. The number of observed resonances is, however, consistent with a symmetric homo-oligomeric structure, with the corresponding nuclei in different subunits experiencing the same chemical environment, in agreement with previously reported data (Doliana et al. 1999).

Main-chain and C$^{\beta}$ resonance assignment

The sequential assignment of the backbone resonances (^{1}H^{N}, ^{15}N, ^{13}C$^{\alpha}$ and ^{13}C′) for the gC1q domain of EMILIN1 was obtained using heteronuclear three-dimensional triple-resonance experiments acquired with a uniformly ^{13}C, ^{15}N, 80% ^{2}H labelled sample (see "Materials and methods" and "Supplementary material"). All the triple-resonance experiments that were performed were obtained with TROSY-type pulse schemes. ^{13}C$^{\alpha}$ nuclei are efficiently relaxed by dipole–dipole (DD) coupling with both ^{1}H$^{\alpha}$ and remote protons but deuteration significantly decreases the transverse ^{13}C$^{\alpha}$ relaxation rate and extends the accessible molecular size range (Kay and Gardner 1997), although the transverse ^{15}N relaxation during coherence transfer steps is scarcely affected by deuteration (Salzmann et al. 1998). Figure 2 illustrates the data quality obtained in our experiments.

One difficulty encountered with the resonance assignment of gC1q arose from the identification of residues belonging to the segment Tyr927-Gly945, which contains four proline residues. Detection and recognition of resonances for this proved difficult due to a relatively poor signal-to-noise ratio of the nitrogen correlation spectra and was only accomplished when the majority of other peaks had already been classified. Line broadening due to slow conformational averaging of this region of EMILIN1 gC1q could be a reason for the lack of detectable amide resonances. Noticeably, this segment has very low sequence homology with other proteins of the family because it includes part of a 12-residue insertion (from Ser940 to Ser951) that is unique to EMILIN1 and EMILIN2 gC1q domains. Two amino acids in the segment 927–945, i.e. Asn934 and Ser940, still have unassigned HN groups and nine amino acids have typical ^{1}H and ^{15}N random coil chemical shifts (Wishart et al. 1991), i.e. they can be identified in the central crowded portion of the TROSY ^{1}H–^{15}N HSQC spectrum. Using additional experiments (details in "Supplementary material"), ^{13}C$^{\alpha}$ and ^{13}C$^{\beta}$ resonances were unambiguously identified for all

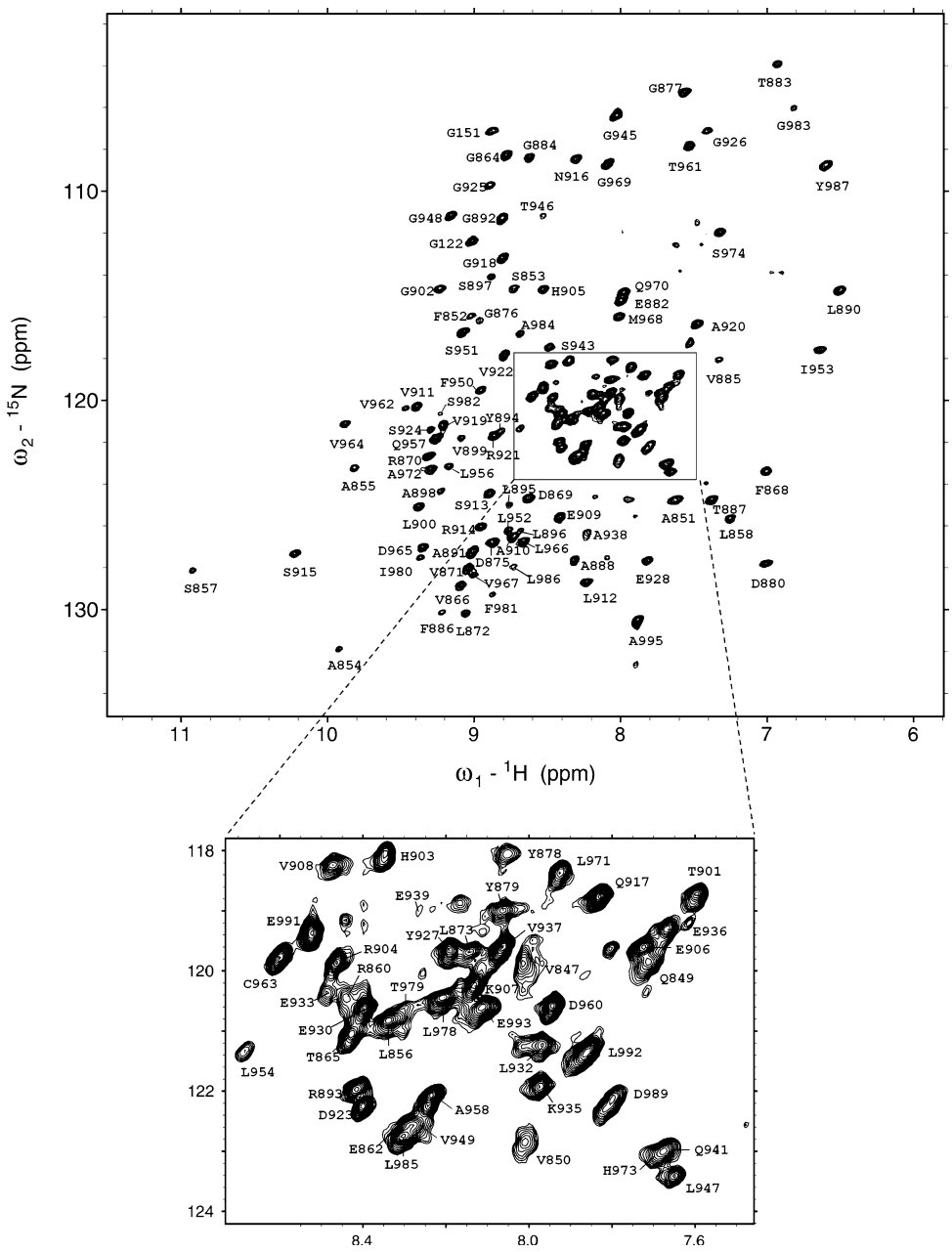

Fig. 1 $^{1}H,^{15}N$-TROSY-HSQC spectrum recorded at 17.6 T (^{1}H frequency 750 MHz) with a 1.8 mM (monomer concentration) U-^{13}C, U-^{15}N, 80% ^{2}H EMILIN1 gC1q sample. Assignments of the central crowded region of the spectrum are reported in the insert

of the 150 amino acids in the protein excluding the 12-residue segment containing the histidine-tag. Percentages of assignment of 98%, 97% were obtained for $^{13}C'$ and $^{1}H^{N}/^{15}N$ nuclei, respectively, and deposited (http://www.bmrb.wisc.edu, accession number 5882). No resonance assignment could be proposed for the backbone amide groups of residues Asn874, Asn934, Ser940 and His994 probably due to fast hydrogen exchange with the solvent. Unless otherwise indicated, the reported data only refer to the 150-residue sequence of the authentic EMILIN1 gC1q domain. The residue numbers follow from the numbering of the homologous C1q/TNF proteins (Shapiro and Scherer 1998).

Secondary structure analysis

According to our data the secondary structure of the EMILIN1 gC1q domain is composed of nine β-strands and one unstructured region including residues 927–945. These structural elements were determined based on a consensus of various different NMR approaches, described below and summarized in Fig. 3.

Secondary chemical shifts

The chemical shift information obtained for the $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ nuclei of the C1q protein was used to identify

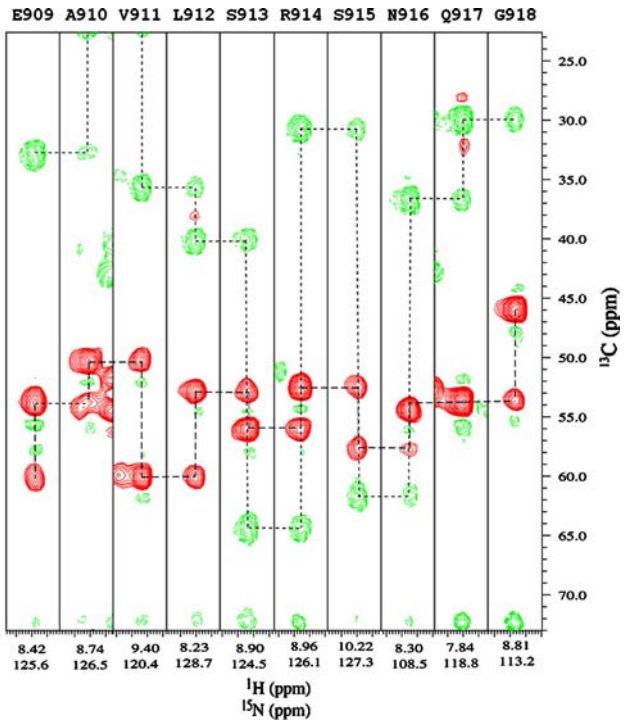

Fig. 2 Strip plot of the 3D [^{13}C, ^{15}N, ^{1}H]-TROSY-HNCACB spectrum obtained at 17.6 T (^{1}H frequency 750 MHz) on the gC1q domain of EMILIN1. For each amide group four peaks are observed: two positive (*red*) $^{13}C^{\alpha}$ resonances from the intraresidue and the preceding residue connectivity and two negative (*green*) $^{13}C^{\beta}$ resonances from the same residue pair. Distinctive $^{13}C^{\beta}$ chemical shift are observed for Ala910, Ser913 and Ser915. The sequential connectivity pattern can be followed using both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances

qualitatively secondary structure elements in the molecule according to the Chemical Shift Index method (Wishart et al. 1994). The C^{α} chemical shift deviations from the random coil values for the C1q domain show a clear predominance of negative values, corresponding to β secondary structure elements. Assuming that when a pattern of four or more consecutive negative $\Delta\delta$ values is found the secondary structure is likely to be β-strand, 11 potential β-strand regions can be predicted in the protein. In detail, these expected strands span the segments Pro846-Leu856, Thr883-Ala891, Arg893-Gly902, Glu906-Ser915, Val919-Asp923, Gly925-Pro929, Thr946-Gln957, Thr961-Leu966, Met968-Ser974, Glu976-Leu985 and Tyr987-Glu993.

Heteronuclear NOE

^{15}N{^{1}H} NOE measurements were run at 11.4 T and 310 K, using two different relaxation/saturation delays, 3 and 5 s. As seen from the histogram in Fig. 4, the results indicate that, besides the N-terminal and C-terminal regions, the highest mobility is observed at segment 927–945. An additional high mobility site is observed at Lys907. The N-terminus appears more rigid than the C-terminus simply because the first reported residue is preceded by a 13-residue fragment including the His-tag. At 310 K, the expected correlation time for a globular protein trimer with an overall molecular weight of 52 kDa ranges between 22.3 and 25.6 ns. This means that at 11.4 T we should find an I/I_0 ratio (the parameter reported in Fig. 4) of 0.90.

Backbone torsion angles

The presence and location of β-strand secondary structure elements in the EMILIN1 gC1q domain has been also established by estimation of the backbone torsion angles, φ and ψ, using TALOS (Cornilescu et al. 1999). The program utilizes a hybrid approach, namely chemical shift and sequence homology, to predict the most likely backbone angles for a given protein residue. By using the experimental $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ and $^{15}N^{H}$ chemical shifts TALOS predicted backbone torsion angles for each assigned amino acid residue. Among these predictions, 82 were ranked as reliable, according to standard TALOS criteria (55% of the total domain residues). While eight angle pairs identify three tight turn conformations, 74 predictions exhibit φ and ψ values very close to the theoretical values of β-strand elements ($-139°/135°$ for antiparallel and -119°/113° for parallel β-strand). The average number of reliable predictions (accuracy of 90%) is usually equivalent to 67% of the residues in a protein. In our case, the lack of the $^{1}H^{\alpha}$ resonance chemical shifts is likely to reduce the algorithm performance to below average. The assignment of $^{1}H^{\alpha}$ resonances, however, was restricted by the highly deuterated samples used to obtain suitably resolved heteronuclear correlation spectra. Although only a partial chemical shift data set was available the TALOS results agree quite well with previous evidence of β secondary structure for the gC1q domain of EMILIN1. Some differences are found mainly at the boundaries of the β strands and for the classification of segments Gly925-Pro929, Ser951-Gln957, Met968-Ser974, Tyr987-Glu993 (Fig. 3).

Amide hydrogen exchange

Information about the secondary structure elements of a polypeptide chain can be gained via the identification of slowly exchanging amide protons. By monitoring over time a series of ^{1}H,^{15}N-HSQC spectra recorded in D_2O (Palmer et al. 1991; Kay et al. 1992) the gradual intensity loss of the exchangeable amide proton peaks allowed the qualitative classification of amide protons to three different categories. Seventy-nine amides exchanged so rapidly that they were not detected in the first ^{1}H,^{15}N-HSQC spectrum

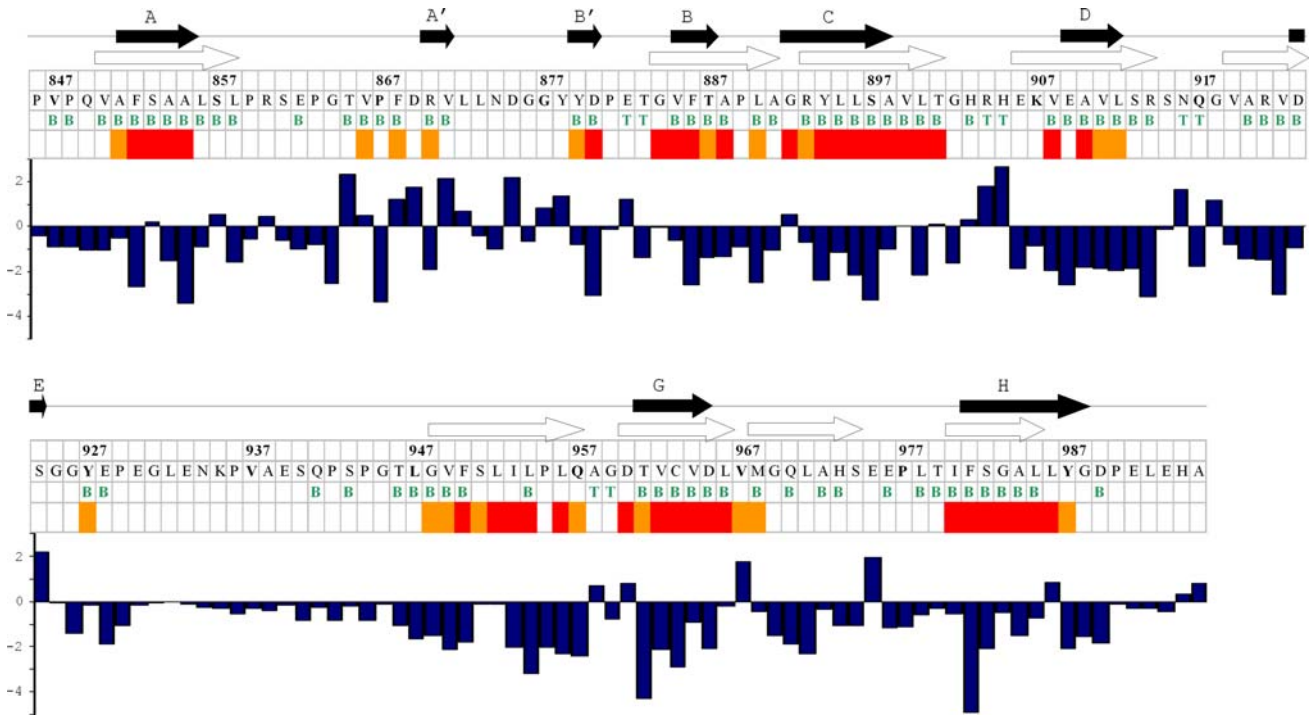

Fig. 3 Secondary structure determination of EMILIN1 gC1q from NMR restraints. Below the aminoacid sequence and the corresponding numbering, the TALOS backbone torsion angle predictions are reported (**B** = β strand, **T** = tight turn). The amide proton exchange rate classification is reported (row 4) with *red* and *yellow* symbols for slow and medium-slow exchange, respectively. At the bottom, the $^{13}C^{\alpha}$ chemical shift deviations from random coil values are plotted as *histogram*. The nine β strand segments that are expected based on the reported NMR evidence are marked by *empty arrows*. The β strands that are found in the best defined model, after the introduction of the RDC restraints, are indicated in the top line by *filled arrows*

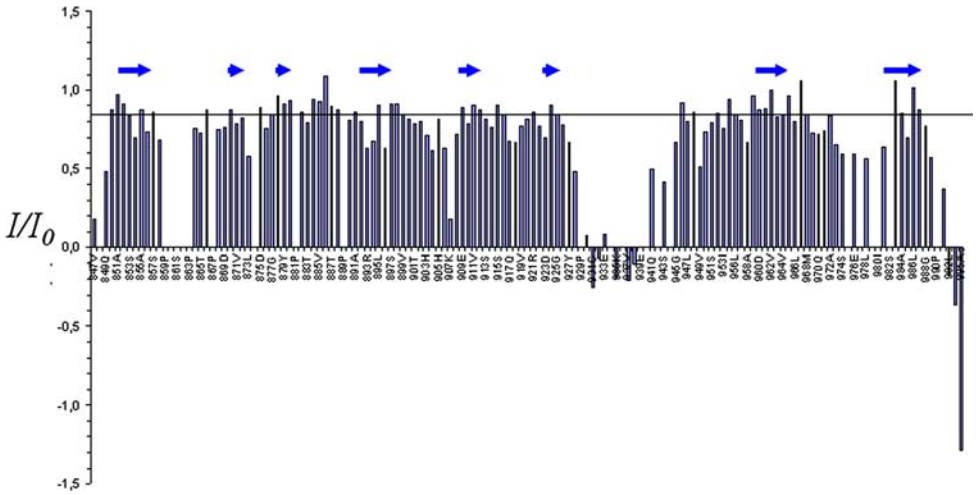

Fig. 4 EMILIN1 gC1q $^{15}N\{^{1}H\}$ NOE data measured at 11.4 T (^{1}H frequency 500 MHz) with 5 s relaxation or saturation delay. The ratio of the ^{1}H-saturated over the unperturbed ^{1}H-^{15} N correlation, I/I_0, was evaluated from the cross-peak intensities. The horizontal line indicates the average I/I_0 level (0.85) calculated by discarding any values below 0.7. *Arrows* indicate the β structure location in the best defined model. *Empty spaces* refer to unresolved, unassigned or proline residues

acquired immediately after transferring the protein to D_2O (i.e. exchange within 30 min after sample preparation). Nineteen further peaks could be detected in that first spectrum but not in the subsequent one and were classified as medium-slow exchanging amide protons. Thirty-eight amide protons gave intense peaks in the very last ^{1}H,^{15}N-HSQC spectrum acquired 68 h after dissolving the protein in D_2O. This latter group constitutes the exhange-protected amide protons in the EMILIN1 gC1q domain. By mapping the distribution of medium-slow and slow exchanging amide protons onto the primary sequence of the protein, regions which are likely to be involved in secondary structure elements in the core of the protein and/or forming hydrogen bonds can be recognized. The segments already recognized to contain β-strands and now identified as also having slowly exchanging amides are Phe852-Ala855, Gly884-Phe886, Tyr894-Thr901, Leu952-Leu954, Val962-Leu966, Ile980-Leu986. Noticeably these segments include

15 out of 19 residues which are conserved in the sequences of the C1q protein family, and their location in the protein core might be crucial for maintaining the protein fold.

Proton NOE correlations

Some ^{1}H NOE connectivities could be identified using 3D ^{15}N-NOESY-HSQC spectra recorded in aqueous solution (Marion et al. 1989a, b). The amount of information expected from this experiment was limited by the extensive deuterium labelling of the sample (80%). Overall only 54 inter-residue NOE connectivities could be identified for restraints. The majority of them represent sequential and short range NOEs between backbone amide protons and very few connectivities could be attributed to characteristic glycine $^{1}H^{\alpha}$ and alanine $^{1}H^{\beta}$ protons. These NOE correlations confirmed the secondary structure elements and loops suggested by previous data (chemical shift deviations analysis, TALOS predictions, rate of amide exchange) and also indicated potential pairing of the β-strands. However, the pattern of observed long-range NOEs was not sufficient to characterize completely the set of paired parallel or anti-parallel β-strands connected by hydrogen bonds.

Non-labile hydrogen observation

In order to extend the assignment database to $^{1}H^{\alpha}$ and possibly to side-chain aliphatic protons, a number of experiments were performed on the available 80% deuterated sample of EMILIN1 C1q. In particular, HNHA, ^{1}H,^{15}N-TOCSY-HSQC, HCCH-TOCSY, HCCH-COSY, ^{1}H,^{1}H-TOCSY and ^{1}H,^{1}H-COSY experiments were recorded but, due the high percentage of deuteration, none of them identified many proton resonances. The same experiments repeated on a protonated ^{13}C, ^{15}N labelled sample did not perform much better because of dipolar broadening. On this latter sample the only valuable experiment was a 3D-HCCH-COSY, a pulse sequence relying on the large one-bond ^{1}H–^{13}C and ^{13}C–^{13}C couplings for the magnetization transfer. The 3D spectrum was still very poor in signals, but careful analysis of the few peaks available allowed the assignment of 16 spin-systems, attributed to the following residues: Ala854, Ala855, Asp880, Ala888, Leu890, Asn934, Lys935, Pro936, Val937, Ala938, Ser943, Val949, Ala958, Glu991, Leu992 and Ala995. Since the listed residues provide the only patterns that are observed in the 3D-HCCH-COSY spectrum, they must have peculiar features. Some residues belong to portions of the polypeptide chain with high mobility, according to ^{15}N{^{1}H} NOE data (Fig. 4), i.e. the C-terminal segment including residues Glu991-Ala995, Val949, Ala958 and the unstructured region previously identified in fragment Tyr927-Gly945. This supports the hypothesis of the existence of a flexible loop (927–945) and a five-residue C-terminal segment that are separate from the main protein fold. It is worth noting that the segment Tyr927-Gly945 and the C-terminal pentapeptide showed typical random-coil chemical shifts and no TALOS prediction output nor slow-exchanging amide protons were available or detected for those amino acids. The remaining residues (854, 855, 880, 888 and 890) that do not exhibit an increased local amide mobility (Fig. 4) were mostly located in β strands, but their ^{1}H–^{13}C connectivities could prove less affected by dipolar broadening because of specific side-chain mobility. Attempts to identify dipolar connectivities involving the available protons from the 16 newly detected amino-acid spin systems were unsuccessful as expected, for flexible moieties.

Residual dipolar coupling estimation

Attempts to obtain an initial global low-resolution fold for the EMILIN1 gC1q domain using NOE-based restrained molecular dynamics were unsuccessful due to the limited amount of data available. A valuable alternative way for structure determination employs the measurement of RDCs, residual dipolar couplings, in weakly aligned media to derive orientational restraints for structure calculation (Tjandra and Bax 1997; Tjandra et al. 1997; Tolman et al. 1995; Brunner 2001). Experimental RDCs for the gC1q domain of EMILIN1 were collected using a low-density polyacrylamide gel medium radially compressed to align the protein, prepared as described in Materials and methods. A gel system was preferred to other orienting media because it can be used over a wide range of conditions (polymer density, temperature, pH and ionic strength). The main difficulty encountered with the measurement of RDCs concerned prolonged and uninterrupted acquisition of high resolution spectra on a delicate sample. The heteronuclear one-bond RDCs for N–H vectors ($^{1}D_{NH}$), CO–N vectors ($^{1}D_{C'N}$) and CO–C$^{\alpha}$ vectors ($^{1}D_{C'C\alpha}$) were obtained using, respectively, ^{1}H–^{15}N HSQC with IPAP method (Ottiger et al. 1998), TROSY ^{1}H–^{15}N HSQC (Wang et al. 1998) and 3D HNCO experiments (Grzesiek and Bax 1992). The complete list of RDC values obtained is available in "Supplementary material". The quality of the results can be appreciated in Fig. 5 that shows sections of the superimposed pairs of spectra recorded on the isotropic (left) and oriented (right) C1q samples. Substantial deviations in the observed splittings from the isotropic $^{1}J_{NH}$ couplings are present in the anisotropic-phase spectra, indicating a significant ordering of the protein. These experimental RDC values were used here to refine a structure obtained by homology

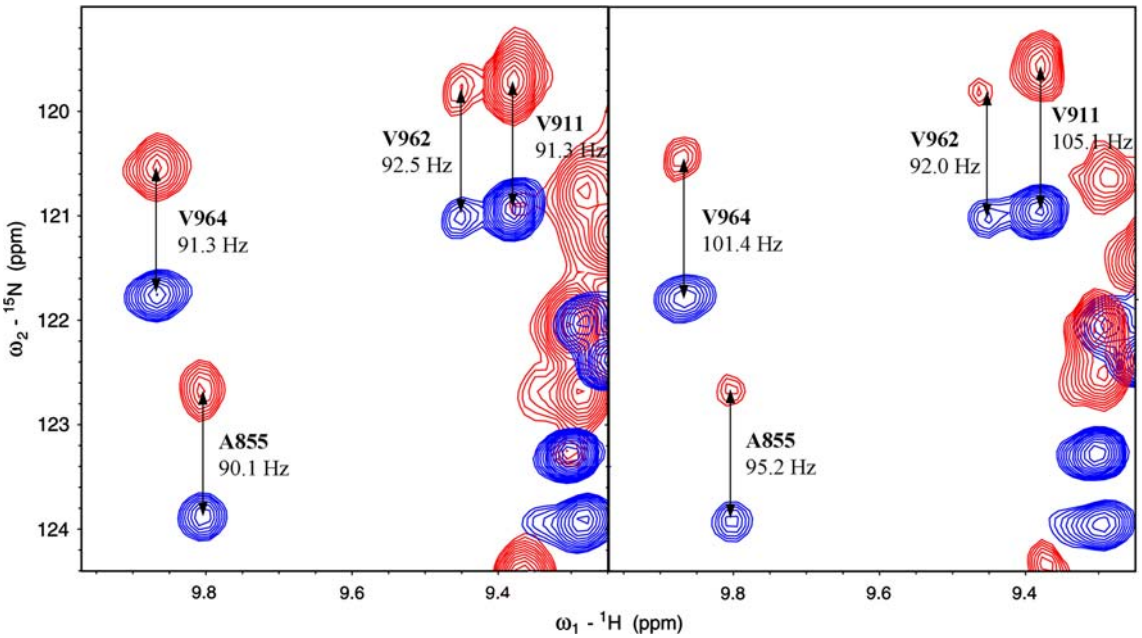

Fig. 5 Sections from IPAP ^{1}H,^{15}N-HSQC spectra acquired on isotropic (*left*) and anisotropic (*right*) solutions at 17.6 T (^{1}H frequency 750 MHz). By comparing $^{1}J_{NH}$ couplings measured in the two different conditions, residual dipolar couplings for ^{15}N–^{1}H nuclei ($^{1}D_{NH}$) were calculated using the equation $^{1}J_{NH}$ (anisot.) = $^{1}J_{NH}$ (isot.) + $^{1}D_{NH}$

modelling according to a protocol described by Chou et al. (2000).

Homology model building

The homology model for the trimeric EMILIN1 gC1q assembly was obtained by using the ACRP30 chain A crystal structure (Shapiro and Scherer 1998) as template protein. This choice was dictated by the best matching obtained upon alignment with the resolved gC1q domain sequences. The quaternary structure of EMILIN1 gC1q homology model is formed by three identical subunits arranged with a three-fold symmetry axis. Each monomer in the trimer shows a β-sandwich folding topology containing ten segments of β-strands and one long unstructured region spanning residues Gly931-Ser951. This region includes an insertion that is unique to EMILIN1 (Doliana et al. 1999) and EMILIN2 (Doliana et al. 2001) members of the C1q protein family and hence could not be modelled using any available structure. Attempts to model this segment by sequence similarity with any other protein in the PDB were fruitless and this region was always classified either as 'disordered' or 'unstructured' by various secondary structure prediction servers such as Psipred, Sam, Jufo (data not shown). Due to accuracy limits of the starting model and to overlaps in the trimer, the resulting structure had rather poor quality at this stage, with only 53.8% residues in the core Ramachandran region, according to PROCHECK (Laskowski et al. 1996).

Analysis of alignment properties (PALES)

Based on both secondary and tertiary structure, the trimeric EMILIN1 gC1q domain structure, obtained by homology modelling, was used as a starting point for a structure refinement with RDCs. Both the location of the β-strands in the homology model and the presence of the unstructured region (Gly931-Ser951) are in good agreement with all the NMR evidence. Tertiary structure similarities between the template ACRP30 A used for the modelling and the EMILIN1 gC1q molecule are consistent with the alignment properties of the two proteins using PALES (Zweckstetter and Bax 2000). A prediction of the molecular alignment properties of the ACRP30 structure was performed using the steric-obstruction algorithm (Steric PALES). Supposing the interaction between the macromolecular solute and the alignment medium is mainly steric, this algorithm can predict the alignment tensor parameters together with heteronuclear dipolar couplings (Zweckstetter and Bax 2000) for every known molecular structure. An asymmetric alignment was predicted by Steric PALES for the ACRP30 crystal structure, as expected for a trimeric protein with a C_3 symmetry axis, with an axial and rhombic component of the alignment tensor, D_a and D_r, of –16.6 and –1.5 Hz, respectively, hence a rhombicity $R = 0.09$. The predicted RDC values for ACRP30 are very similar to the experimental RDCs measured for the EMILIN1 gC1q domain. Figure 6 shows, for example, the comparison of calculated and experimental $^{1}D_{NH}$ couplings for the two molecules (a

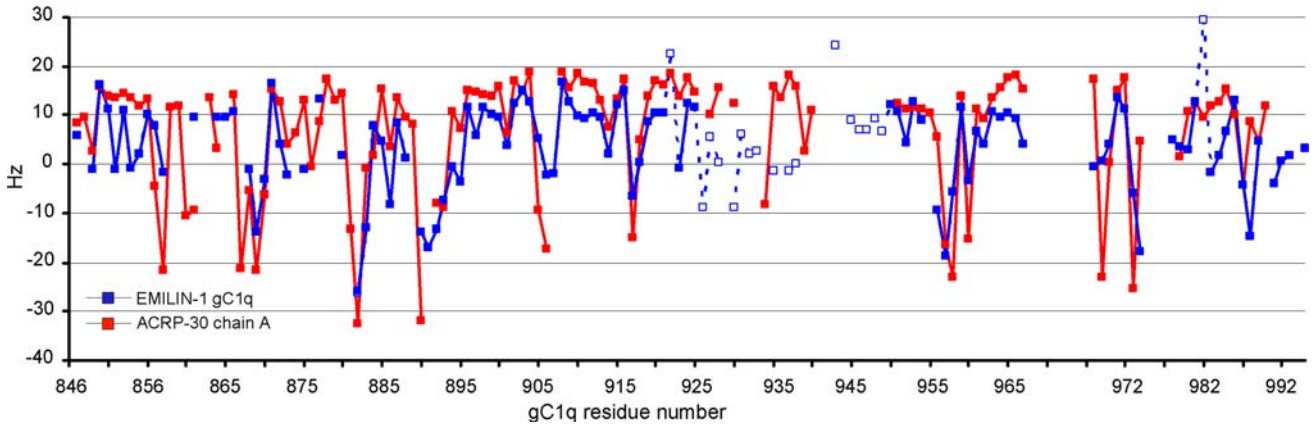

Fig. 6 Comparison of $^1D_{NH}$ residual dipolar couplings measured for EMILIN1 gC1q (*blue filled squares*) and $^1D_{NH}$ residual dipolar couplings predicted for ACRP-30 chain A with PALES (*red filled squares*). Similar values can be recognized in the two molecules. EMILIN1 gC1q RDC values that show large deviations from predicted values, i.e. residues 927–950, 922 and 982, are indicated by open blue squares. The quality of the corresponding experimental data is presented in Fig. 5

single monomer considered for each protein). Except for the fragment Tyr927-Phe950, remarkable similarities can be found in the two patterns, confirming the reliability of the alignment and suggesting a similar orientation for the internuclear vectors in the two proteins. The agreement between predicted ACRP30 and experimental EMILIN1 gC1q couplings was not obvious when comparing different classes of dipolar couplings, namely $^1D_{C'C\alpha}$ and $^1D_{C'N}$ couplings, possibly due to larger experimental errors associated with these values. For these two classes, five gC1q experimental couplings showing large deviations with respect to the corresponding ACRP30 calculated values were rejected, in order to obtain an estimate for the alignment tensor parameters D_a and D_r of EMILIN1 gC1q molecule. Values of $D_a = -14$ Hz, $D_r = -2.3$ Hz, hence a rhombicity $R = 0.19$, were obtained from the histogram distribution of the ensemble of normalized couplings ($^1D_{NH}$, $^1D_{C'N}$ and $^1D_{C'C\alpha}$). The histogram was built by discarding the aforementioned outliers RDCs, as well as RDCs measured for amino acids in the segment Tyr927-Gly945. The heteronuclear (and other) NOE data clearly show that the segment Tyr927-Gly945 is unstructured and the corresponding RDCs were not included in the database for D_a and D_r determination.

Simulated annealing refinement

Before starting any refinement, the trimeric crystal structure-based homology model for the EMILIN1 C1q domain was submitted to a regularization step in order to enforce the planarity of the peptide bonds and tetrahedral geometry at C^α sites. The regularized model was analysed with PALES to evaluate the fitting of experimental $^1D_{NH}$ couplings. The remaining classes of RDCs, namely $^1D_{C'N}$ and $^1D_{C'C\alpha}$, with higher uncertainty, were not included in this preliminary manipulation of the homology model. The PALES algorithm applied to the trimeric regularized EMILIN1 gC1q structure returned an alignment tensor for which the experimental $^1D_{NH}$ couplings have the lowest least-square deviation from the calculated ones, finally based on the values $D_a = -12.5$ Hz and $R = 0.015$. A reduced set of dipolar couplings was employed to optimise the fitting, preserving the threefold symmetry by trying to maintain identical coupling values for the corresponding residues in the three polypeptide chains of the trimer. The fitting was evaluated by checking the quality factor Q, which indicates the agreement between the true protein structure and the coordinates used for calculating predicted couplings (Cornilescu et al. 1998). The initial Q factor of 0.652, calculated after fitting the regularized structure with the reduced set of $^1D_{NH}$ couplings, indicates that the overall molecular folding is compatible with the RDC data although small re-orientations of the backbone internuclear vectors are necessary. Q values lower than 30% normally correspond to a good agreement between a set of coordinates and measured RDCs. The structural changes needed to reach agreement with the dipolar couplings were achieved by means of a simulated annealing protocol implemented in Xplor-NIH (Chou et al. 2000), described in details in the experimental section. Although the number of one-bond RDCs that were measured for NH vectors ($^1D_{NH}$), C′N vectors ($^1D_{C'N}$) and C′C^α vectors ($^1D_{C'C\alpha}$) did not reach the requirement of three values per residue (Chou et al. 2000), further information was added in the refinement dataset from NOEs and from backbone dihedral angle restraints that proved consistent with the homology model. The main difficulty of a RDC-based structure refinement resides in the degeneracy intrinsically associated with RDCs that may cause the structure to be partially inverted during the simulated annealing trajectory and trapped into

local minima. To avoid this the simulated annealing was performed in two stages, both run at very low temperature. The two stages employed the same set of experimental restraints introduced as energy penalty terms with appropriate force constants (dihedral angles, RDCs, NOEs, H-bonds), together with various other terms commonly used in NMR structure calculations (van der Waals, bond lengths, bond and torsion angles, specific Ramachandran potential, improper torsion angles). The final result was an ensemble of ten low-energy structures with limited dispersion, i.e. a C^{α} RMSD around 0.4 Å, and a Q factor (Cornilescu et al. 1998) between 0.245 and 0.273 calculated using only $^{1}D_{NH}$ or the three classes of backbone dipolar couplings, respectively, on a single structure of the ensemble. Table 1 collects the main parameters of the refinement statistics. A comparison between the experimental and back-calculated RDCs is provided in supplementary material. Besides the compactness of the ensemble of structures, other parameters like the ϕ–ψ distribution in the most favourite regions of Ramachandran plot and the backbone average G-factor (Laskowski et al. 1996) were evaluated to check if the results meet the accepted standards of structure determination. The first indicator gave a value of 62%, i.e. within the consensus standard value for NMR determined structures of 73.6 ± 15.5% (Doreleijers et al. 1998). The G-factor, i.e. the log-odds score probability, computed from high-resolution X-ray data, of a specific angle distribution within the final structure family expressed as circular variance (Allen and Johnson 1991) or order parameter (Hyberts et al. 1992), resulted in a value of –1.49 for the overall backbone, below the accepted value range of about –1 for high resolution structures. The structure ensemble resulting at this stage, in spite of being a low resolution one was

Table 1 NMR restraint and refinement statistics[a]

	EMILIN1 gC1q	
	Restrained MD	NAMD Minimization
NMR distance & dihedral restraints		
Distance restraints		
Total NOE	54	54
Intra-residue	0	0
Inter-residue	54	54
Sequential (|i-j| = 1)	41	41
Medium-range (|i-j| < 4)	5	5
Long-range (|i-j > 5)	8	8
Intermolecular	0	0
Hydrogen bonds	27	–
Total dihedral angle restraints	163	163
Phi	84	84
Psi	79	79
RDC restraints		
Total	234	–
$D_{C'CA}$	59	–
$D_{C'N}$	70	–
D_{NH}	105	–
Structure statistics		
Q factor	0.260 ± 0.002	0.367 ± 0.004
Average pairwise r.m.s.d.[b] (Å)		
Heavy	0.57 ± 0.19	0.62 ± 0.17
Backbone	0.34 ± 0.16	0.36 ± 0.17
Restraint violations		
Average number of distances violations (>0.2 Å)	7.7 ± 0.0	28.4 ± 0.7
Average distance violation (Å)	0.5 ± 0.2	0.3 ± 0.3
Average number of H-bond violations (>0.2 Å)	5.9 ± 0.7	–
Average H-bond violation (Å)	0.24 ± 0.04	–
Average number of violated dihedral angles (>|5°|)	14.7 ± 2.1	28.8 ± 2.0
Average dihedral angle violation	10.9 ± 5.6	2.4 ± 3.5

[a] The table lists the available information from NMR evidence that was employed for refinement purposes of the structure obtained by homology modelling. The left column refers to the restrained molecular dynamics refinement using NOE distance, H-bond, dihedral and RDC restraints. The right column refers to the final NAMD refinement using only NOE distance and TALOS dihedral restraints. All values are given on a single subunit basis. The numbers of dihedral angle deviations >|5°| reflect an average phi and psi uncertainty of 18.7 ± 5.6° from TALOS evaluations. Additional qualitative information from NMR data are reported in Fig. 3

[b] Pairwise r.m.s.d. were calculated over ten refined structures

deposited in the PDB (ID code 2OII), because the corresponding functional inference was successfully tested (Verdone et al. 2008). As described in the "Materials and methods" section, a further refinement step was subsequently performed in order to improve the backbone geometry. At the end of this minimization, 70.8% of the backbone torsion angles were in the core region and an additional 21% in allowed regions of Ramachandran map, according to PROCHECK output with a resolution of 2.1 Å (Laskowski et al. 1996). Also the backbone dihedral G-factor improved slightly although still below the accepted threshold (–1.38). A control by WHATCHECK (Hooft et al. 1996) indicates that there are still features that deviate abnormally from the high-resolution structure standards, mainly bad residue packing environments and unsatisfied buried hydrogen bonds. The last-refined structures were deposited in the PDB (ID code 2ka3).

Overall structural model

The structure of the EMILIN1 gC1q domain with the lowest RDC restraint energy, refined after calculating one hundred conformations in each stage of the simulated annealing protocol can be considered the best representation of NMR-restrained homology model for the protein solution structure (Fig. 7). The secondary structure elements for this structure are reported in Figs. 3 and 7c. The original trimeric homology model, built by superimposing the backbone of three identical monomers to the template backbone, underwent a regularization step followed by energy minimization which introduced minor differences among the three polypeptide chains. For the purpose of this study, however, the three polypeptide chains can be considered identical. The simulated annealing steps that followed the first regularization session and the final

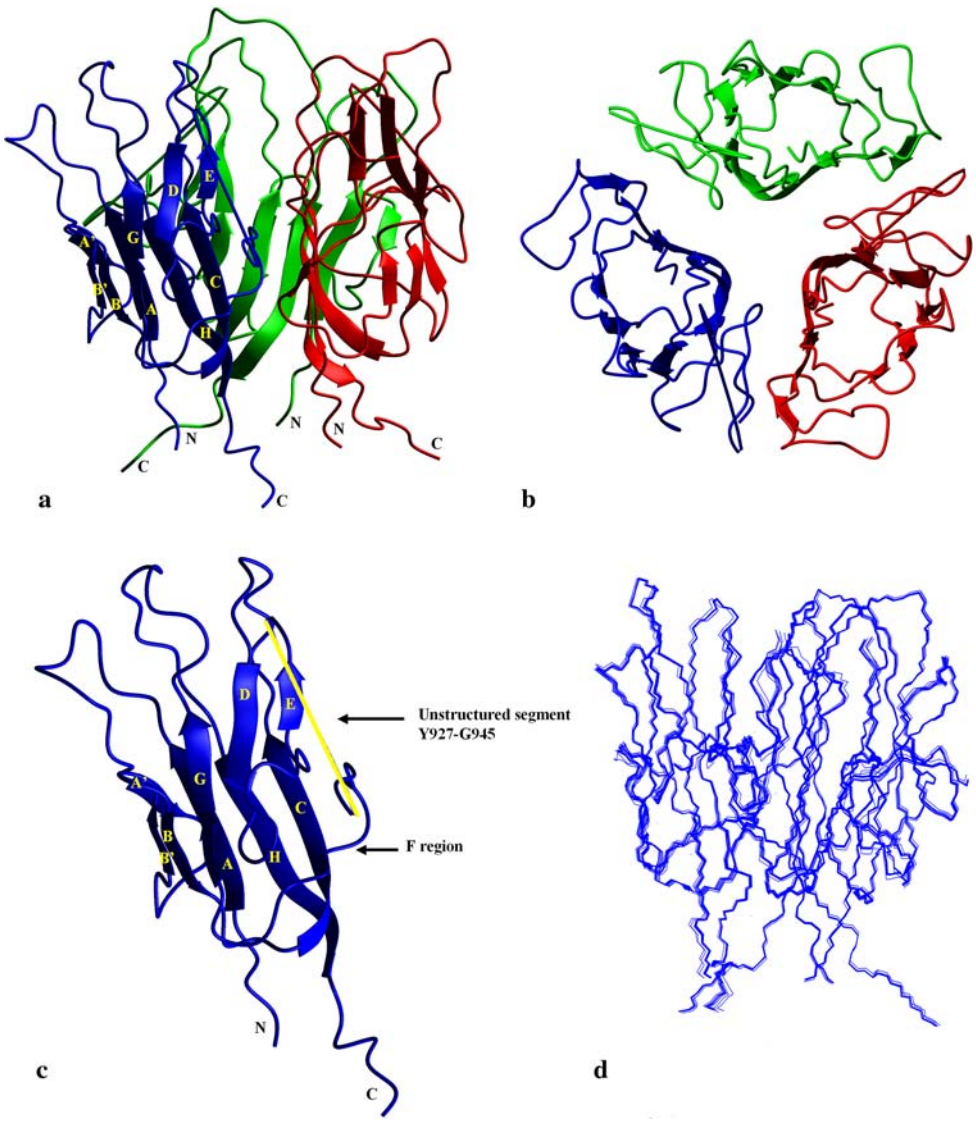

Fig. 7 Structure of the homotrimeric globular EMILIN1 gC1q domain as obtained by homology model refinement with RDCs. The three protomers are shown in ribbon representation and different colours, respectively, as side view and top view, in panels **a** and **b**. Each monomer has the nine-stranded folding topology displayed in panel **c**. The location of region F is indicated by the *arrow*. A yellow line is drawn instead of the unstructured segment Y927-G945 that was not modelled in our calculation. β-strands are labelled according to the C1q/TNF superfamily nomenclature (Shapiro and Scherer 1998). Panel **d** shows the superposition of the C^{α} traces for the ten lowest RDC restraint energy structures, obtained from NMR-restrained refinement of the homology-modeled structure. Drawings were done using MOLMOL (Koradi et al. 1996)

minimization refinement were always performed using an identical set of restraints for the three chains, and the molecular symmetry was enforced by automatically generated inter-subunit distance restraints, deliberately introduced in the Xplor calculations to preserve the C_3 symmetry. Apart from these inter-subunit symmetry-related distance restraints, no other inter-subunit connectivities were imposed. Thus the trimeric association geometry built from the quaternary structure of the ACRP30 homologous protein, although not supported by direct identification of inter-subunit contacts, is in agreement with the RDC data which identify the internuclear vector orientation, and hence implicitly restrict the relative orientation of the monomers.

The subunit arrangement defines a central cavity (Fig. 7b) where a number of inter-subunit hydrophobic contacts are established. Comparing the EMILIN1 gC1q quaternary structure before and after the simulated annealing refinement, a distinct increase of the size of the central cavity is observed, possibly due to topology changes undergone by each monomer during the refinement to optimise the molecular packing. The cavity in the final structure appears as a truncated cone, with the smaller base on the top of the trimer ($\sim$150 Å^2) and a depth of $\sim$16 Å. The overall buried trimer surface calculated with a 1.4 Å molecular probe is 3,096 Å^2, a value which appears lower than any other values calculated for the buried surface of homologous proteins gC1q domains (5,320 Å^2 for ACRP30, 5,490 Å^2 for complement C1q protein, 6,150 Å^2 for collagen VIII and 7,360 Å^2 for collagen X). Besides the cautions imposed by the lack of direct experimental information on the interface between the subunits and the quality limits of the packing environments, the decrease of buried surface in EMILIN1 gC1q may reflect partly the missing contribution of segment 927–945 that could not be included in the model (see below).

The tertiary structure of each monomer in the trimer is formed by nine β-strands arranged in two sheets, giving a β-sandwich folding topology (Fig. 7c). Following the numbering of homologous proteins (Shapiro and Scherer 1998), the identified β-strands are: A (A851-A855), A′ (R870-V871), B′ (Y879-D880), B (V885-T887), C (G892-A898), D (E909-L912), E (D923-S924), G (T961-D965), H (F981-G988). The two sheets of each monomer define a hydrophobic core formed by residues highly conserved among the proteins of the family: Phe852, Tyr879, Phe886, Tyr894, Leu956, Val962 and Gly983. The inner channel is mainly defined by sheet 1 from each monomer formed by strands A, A′, H, C opposed to sheet 2 (strands B, B′, D, E and G) that constitutes the exterior of the molecule. The observation of HCCH TOCSY correlation for residues 854, 855, 880, 888 and 890, attributed to reduced dipolar broadening due to side chain mobility (despite the low mobility of the relative backbone NHs), correlates with the exposure of strands B and B′ and suggests that the partially buried C-terminus end of strand A could be also poorly packed in the channel interior. The central channel geometry is also defined by an unstructured segment including residues Thr947-Leu956. This stretch appears to deviate from an extended conformation and breaks into a reverse turn arrangement due to a diverging (βI-like) turn at residues Ser951 and Leu952. The geometry of the EMILIN1 gC1q molecule in this region is unusual compared to homologous proteins because the unstructured segment Thr947-Leu956 essentially corresponds to the β-strand F that faces strand E, giving rise to a ten-stranded β-sandwich folding topology. The largest deviations from canonical β structure, in spite of structure prediction, ^{15}N and ^{13}C resonance frequencies and hydrogen exchange pattern, concern a stretch of six residues, i.e. Ser951-Leu956, where the reverse turn occurs. The presence of a turn, however, is consistent with the chemical shift evidence at Ser951 and Leu952. The preceding residues, i.e. 947–950, adopt an extended rigid geometry, consistent with ^{15}N{^{1}H} NOE data (Fig. 4), that does not comply, however, with the conformational interval requirements for β structure classification. It is worth noting that, in our structure model, segment Thr947-Leu956 has a well-defined position and participates in inter-subunit interactions even without a β-strand arrangement. A similar characteristic has been previously observed for the crystal structure of the ACRP30 C1q domain, where the melting of the edge strand E to become a loop was reported for the subunit M2 (Shapiro and Scherer 1998).

A very important feature of the EMILIN1 gC1q NMR-based homology model is the presence of an unstructured flexible segment for each monomer, spanning residues Tyr927-Gly945, which could not be modelled due to the lack of experimental restraints. Our structure calculation can only define for each subunit the position of the two residues adjacent to this flexible loop, namely Gly926 and Thr946. The position of these edge-residues allows us to suggest that this unstructured loop should be placed at the apex of the trimeric domain and protrude out from the main globular assembly into the solvent. The distance between the carbonyl of Gly926 and the amide of Thr946, i.e. the shortest path from the end of strand E and the start of the missing strand F region, is around 23 Å and could be spanned by six or seven residues, at minimum. This is nicely consistent with the heteronuclear NOE data (Fig. 4) that illustrate how the local mobility of EMILIN1 gC1q gradually increases from the average extent expected for a residue that moves with the same tumbling rate as the overall trimer (Thr946) to a higher flexibility regime that can be envisaged for a flexible loop, from Glu939 down to Glu930, and then gradually decreases back to the initial

level of the trimer along the fragment 929–925. The absence of local secondary and tertiary structure, the mobility pattern and the high solvent accessibility indicate that it is likely that the three 19-residue unstructured loops could constitute an extension by which the whole EMILIN1 protein interacts with its ligands/receptors. Considering the relative distribution of hydrophobic, aliphatic and charged residues in this segment, charged interactions appear to be more likely than hydrophobic ones. These predictions have been tested by site directed mutagenesis and were reported in a separate communication (Verdone et al. 2008). All the expected inferences were confirmed; in particular it was shown that (i) the $\alpha 4\beta 1$ integrin binding site of EMILIN gC1q has a key residue, Glu933, located in the most mobile region of the unstructured segment 927–945 (Fig. 4); (ii) the disordered strand F region is essential for correct monomer folding and assembly in the EMILIN gC1q trimer, at variance with segment 927–945.

Discussion

In this paper we presented the NMR-restrained homology model of a trimeric protein, the globular C1q domain of EMILIN1. Once backbone resonance assignment was completed, the definition of the tertiary structure was achieved using a hybrid approach that involved homology modelling and structure refinement with experimental RDCs, chemical shift analysis and a limited number of NOE connectivities. Because RDCs are affected by errors and are degenerate, i.e. compatible with more than one internuclear vector orientation, the corresponding restraints were introduced gradually, to avoid large distortions which might have trapped the structure into false minima. The conformation obtained defines both tertiary and quaternary structure for the trimeric EMILIN1 gC1q domain. Among the observed amide NOEs, no connectivity could be attributed to a contact between monomers. The degeneracy of the NMR resonance chemical shifts, on the other hand, indicates that a ternary symmetry must be preserved by the solution structure and the measured extent of $^{15}N\{^1H\}$ NOE is fully consistent with a trimeric species. So the actual experimental restraints for the refinement were based on the RDCs, as well as NOEs and dihedral angles and TALOS chemical-shift-based predictions. The relative weighting between geometric terms and experimental restraints was substantially the same as reported by Chou et al. (2000).

Although the resolution that was finally achieved is less than that typically reached using a high number of restraints in NOE-based NMR structures, the RDC data provided a means of accomplishing a homology model from the NMR evidence. The percentage of residues in the core region of the Ramachandran plot reached for the EMILIN1 gC1q domain, i.e. 62%, falls within the consensus standard of the NMR-determined structures ($73.6 \pm 15.5\%$). A further refinement step using CHARMM v. 27b (MacKerell et al. 1998) with the CMAP correction (MacKerell et al. 2004) improves this percentage up to 70.8%. This is valuable if one takes into account the inherent difficulties of the restraint determination with a 52 kDa protein assembly, and the fact that consensus standards are much more biased over the typical structural accuracy that is obtained from 'traditional' NMR restraints (NOEs, *J*-couplings, exchange, etc.).

The overall fold of the presented structural model reproduces, as expected, many features of the known structural homologues, namely mouse ACRP30, human collagen X, mouse collagen VIII and human complement C1q (Shapiro and Scherer 1998; Bogin et al. 2002; Kvansakul et al. 2003; Gaboriaud et al. 2003). However this structure model describes distinctive structural properties especially the loss of strand F in each domain. The overall folding topology is similar but the disorder of strand F causes an increase in the central cavity formed by the three subunits. The second unusual feature of the final structure regards the presence of a 19-residue region (segment 927–945), for each subunit of the trimer, with 10–11 residues out of 19 protruding from the main globular structure. Our experimental evidence indicates that this segment is unfolded, highly dynamic and highly accessible to solvent, facts that make fragment 927–945 a good candidate for hosting an interaction site; we note that flexible external loops in intrinsically disordered protein regions are common interaction regions (Uversky 2002; Radivojac et al. 2007). The role and relevance of fragment 927–945 were successfully tested by site-directed mutagenesis experiments that, in particular, located at Glu933 the $\alpha 4\beta 1$ integrin binding site of EMILIN gC1q (Verdone et al. 2008).

An interesting feature of the EMLIN1 gC1q model presented here is the conspicuous decrease of the contact surface among the monomers within the trimeric assembly, with respect to the corresponding areas observed in the crystal structures of ACRP30, collagen X, collagen VIII and complement C1q (Shapiro and Scherer 1998; Bogin et al. 2002; Kvansakul et al. 2003; Gaboriaud et al. 2003). This would suggest a decreased stability of the EMILIN1 gC1q homotrimer, although boiling is required to dissociate the complex into monomers (Mongiat et al. 2000). A possible explanation could involve Ca^{2+} ions. The EMILIN1 gC1q preparations should not contain bound Ca^{2+} ions, and mass spectrometry to assess Ca^{2+} presence in the NMR sample failed because of the wide molecular mass distribution range arising from non-uniform deuterium

labelling. Ca^{2+} ions, occurring in human collagen X (Bogin et al. 2002) and complement C1q (Gaboriaud et al. 2003), do not seem to compact the trimeric aggregate. It is worth noting, however, that the two residues involved in metal ion coordination in the known structures map to Gly926 and Asn934 in EMILIN1 gC1q, i.e. two positions related to the disordered fragment 927–945. Another possible reason to account for the contact surface reduction in the EMILIN1 gC1q trimer could be the absence of the strips of partially exposed aromatic residues that are involved in the supramolecular assembly of the homotrimeric structures for collagens VIII and X (Kvansakul et al. 2003; Bogin et al. 2002). None of the corresponding positions in EMILIN gC1q (Ser861, Val908, Leu912, Asp923, Gly925, Cys963, Ala972, Glu976) bears an aromatic side chain (see "Supplementary material"), although three out of eight exhibit aliphatic groups. In ACRP30, four positions of the strip correspond to aromatic residues, whereas the remaining four conserve a hydrophobic character at least. In the human complement C1q, on the other hand, the partially exposed aromatic strip is largely absent, but in this case a heterotrimeric interface is being considered. Thus, beyond the prudence for an inference that should be affected by the explicit absence of segment 927–945 in the model and the lack of direct information on the trimeric interface, a reduced contact surface in EMILIN1 gC1q may be plausible.

As recently reported, EMILIN1 is linked to the pathogenesis of hypertension (Zacchigna et al. 2006) and observed to be functionally involved in cell migration and adhesion (Spessotto et al. 2003, 2006). In particular the latter two functions are related to the C-terminal domain that is addressed by the present study *via* an interaction with $\alpha 4\beta 1$ integrin. EMILIN1 gC1q domain is the only one, among the gC1q structures described to date, that interacts with an integrin. Any structural information on this functionally relevant domain is therefore inherently valuable. Our NMR-restrained homology model, albeit a low-resolution structure, prompted for successful functional experiments (Verdone et al. 2008) that provide novel and original information concerning the peculiar interaction features of the EMILIN1 gC1q domain and the link with the unstructured fragment 927–945 including in part the insertion unique to EMILIN proteins. Accounting for the underlying NMR-based modelling was worth at this stage, to illustrate an example of structure-function correlation within a still ongoing project aimed at improving quality and extent of the results.

Acknowledgments This work was supported by Italian Ministry of Education (MIUR) funding (PRIN and FIRB N. RBRN07BMCT). The assistance of Dr. A. Makek is acknowledged. IDC and SAC were supported by the Wellcome Trust, GV was supported also by EMBO. Finally we wish to acknowledge the contribution of Dr. Anatoly Sharipo (ASLA Ltd.) who untimely passed away recently. His generous commitment was essential to start this and many other projects.

References

Allen FH, Johnson O (1991) Automated conformational analysis from crystallographic data. IV. Statistical descriptors for a distribution of torsion angles. Acta Crystallogr B47:62–67

Bartels C, Xia TH, Billeter M, Güntert P, Wüthtrich K (1995) The program XEasy for computer-supported NMR spectral-analysis of biological macromolecules. J Biomol NMR 6:1–10

Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE (1990a) Practical aspects of proton–carbon–carbon–proton three-dimensional correlation spectroscopy of ^{13}C-labelled proteins. J Magn Reson 87:620–627

Bax A, Clore GM, Gronenborn AM (1990b) ^{1}H–^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new 3-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched protein. J Magn Reson 88:425–431

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Methods Enzymol 339:127–174

Bogin O, Kvansakul M, Rom E, Singer J, Yayon A, Hohenester E (2002) Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. Structure 10:165–173

Boyd J, Soffe N (1989) Selective excitation by pulse shaping combined with phase modulation. J Magn Reson 85:406–413

Bressan GM, Castellani I, Colombatti A, Volpin D (1983) Isolation and characterization of a 115,000-dalton matrix-associated glycoprotein from chick aorta. J Biol Chem 258:13262–13267

Brunner E (2001) Residual dipolar coupling in protein NMR. Concepts Magn Reson 13:238–259

Chou JJ, Li S, Bax A (2000) Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings. J Biomol NMR 18:217–227

Chou JJ, Gaemers S, Howder B, Louis JM, Bax A (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J Biomol NMR 21:377–382

Christian S, Ahorn H, Novatchkova M, Garin-Chesa P, Park JE, Eisenhaber F, Rettig WJ, Lenter MC (2001) Molecular cloning and characterization of EndoGlyx-1, an EMILIN-like multisubunit glycoprotein of vascular endothelium. J Biol Chem 276:48588–48595

Colombatti A, Doliana R, Bot S, Canton A, Mongiat M, Mungiguerra G, Paron-Cilli S, Spessotto P (2000) The EMILIN protein family. Matrix Biol 19:289–301

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:10571–10572

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Danussi C, Spessotto P, Petrucco A, Wassermann B, Sabatelli P, Montesi M, Doliana R, Bressan GM, Colombatti A (2008) EMILIN1 causes structural and functional defects in lymphatic vasculature. Mol Cell Biol 28:4026–4039

Davis AL, Keeler J, Laue ED, Moskau D (1992) Experiments for recording pure absorption heteronuclear correlation spectra using pulsed field gradients. J Magn Reson 98:207–216

Doliana R, Mongiat M, Bucciotti F, Giacomello E, Deutzmann R, Volpin D, Bressan GM, Colombatti A (1999) EMILIN, a

component of the elastic fiber and a new member of the C1q/tumor necrosis factor superfamily of proteins. J Biol Chem 274:16773–16781

Doliana R, Bot S, Mungiguerra G, Canton A, Cilli SP, Colombatti A (2001) Isolation and characterization of EMILIN2, a new component of the growing EMILINs family and a member of the EMI domain-containing superfamily. J Biol Chem 276:12003–12011

Doreleijers JF, Rullmann JAC, Kaptein R (1998) Quality assessment of NMR structures: a statistical survey. J Mol Biol 281:149–164

Gaboriaud C, Juanhuix J, Gruez A, Lacroix M, Darnault C, Pignol D, Verger D, Fontecilla-Camps JC, Arlaud GJ (2003) The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. J Biol Chem 278:46974–46982

Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR technique applied to a 31 kDa protein. J Magn Reson 96: 432–440

Grzesiek S, Bax A (1993) The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. J Am Chem Soc 115:12593–12594

Hayward CP, Hassell JA, Denomme GA, Rachubinski RA, Brown C, Kelton JG (1995) The cDNA sequence of human endothelial cell multimerin. A unique protein with RGDS, coiled-coil, and epidermal growth factor-like domains and a carboxyl terminus similar to the globular domain of complement C1q and collagens type VIII and X. J Biol Chem 270:18246–18251

Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272

Hyberts SG, Goldberg MS, Havel TF, Wagner G (1992) The solution structure of Eglin c based on measurements of many NOE and coupling constants. Protein Sci 1:736–751

Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 151:283–312

Kay LE, Gardner KH (1997) Solution NMR spectroscopy beyond 25 kDa. Curr Opin Struct Biol 7:722–731

Kay LE, Keifer P, Saarinen T (1992) Pure absorbtion gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlation in H_2O samples of proteins. J Magn Reson Ser B 101:333–337

Kishore U, Gaboriaud C, Waters P, Shrive AK, Greenhough TJ, Reid KB, Sim RB, Arlaud GJ (2004) C1q and tumor necrosis factor superfamily: modularity and versatility. Trends Immunol 25:551–561

Koradi R, Billeter M, Wüthtrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55

Kvansakul M, Bogin O, Hohenester E, Yayon A (2003) Crystal structure of the collagen alpha1(VIII) NC1 trimer. Matrix Biol 22:145–152

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WEIII, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

MacKerell AD Jr, Feig M, Brooks CLIII (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126:698–699

Marion D, Wüthrich K (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of proton–proton spin–spin coupling constants in proteins. Biochem Biophys Res Commun 113:967–974

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989a) Overcoming the overlap problem in ^{1}H-NMR spectra of larger proteins by use of 3-dimensional heteronuclear ^{1}H–15 N Hartmann–Hahn multiple quantum coherence and nuclear Overhauser multiple quantum coherence spectroscopy-application to Interleukin- 1-β. Biochemistry 28:6150–6156

Marion D, Kay LE, Sparks SW, Torchia DA, Bax A (1989b) 3-Dimensional heteronuclear NMR of ^{15} N-labeled proteins. J Am Chem Soc 111:1515–1517

Mongiat M, Mungiguerra G, Bot S, Mucignat MT, Giacomello E, Doliana R, Colombatti A (2000) Self-assembly and supramolecular organization of EMILIN. J Biol Chem 275:25471–25480

Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar coupling from simplified two-dimensional NMR spectra. J Magn Reson 131:373–378

Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected heteronuclear correlation spectroscopy. J Magn Reson 93:151–170

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradivic CZ, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:439–456

Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc Natl Acad Sci USA 95:13585–13590

Salzmann M, Wider G, Pervushin K, Senn H, Wüthrich K (1999) TROSY-type triple-resonance experiments for sequential NMR assignment of large proteins. J Am Chem Soc 121:844–848

Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J Biomol NMR 18:303–309

Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broadband decoupling: WALTZ-16. J Magn Reson 52:335–338

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators: application to spin decoupling. J Magn Reson 77:274–293

Shapiro L, Scherer PE (1998) The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. Curr Biol 12:335–338

Spessotto P, Cervi M, Mucignat MT, Mungiguerra G, Sartoretto I, Doliana R, Colombattti A (2003) Beta 1 integrin-dependent cell adhesion to EMILIN1 is mediated by the gC1q domain. J Biol Chem 278:6170–6177

Spessotto P, Bulla R, Danussi C, Radillo O, Cervi M, Monami G, Bossi F, Tedesco F, Doliana R, Colombatti A (2006) EMILIN1 represents a major perivascular element determining human trophoblast invasion of the uterine wall. J Cell Sci 119: 4574–4584

States DJ, Haberkorn RA, Ruben DJ (1982) A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. J Magn Reson 48:286–292

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar ^{1}H–^{15}N and ^{1}H–^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 4:732–738

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci USA 92:9279–9283

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756

Verdone G, Colebrooke SA, Boyd J, Viglino P, Corazza A, Doliana R, Mungiguerra G, Colombatti A, Esposito G, Campbell ID (2004) Sequence-specific backbone NMR assignments for the C-terminal globular domain of EMILIN1. J Biomol NMR 29: 91–92

Verdone G, Doliana R, Corazza A, Colebrooke SA, Spessotto P, Bot S, Bucciotti F, Capuano A, Silvestri A, Viglino P, Campbell ID, Colombatti A, Esposito G (2008) The solution structure of the EMILIN1 gC1q domain reveals a disordered insertion necessary for interaction with the $\alpha 4\beta 1$ integrin. J Biol Chem 283:18947–18956

Wang XY, Marquardt JL, Wingfield P, Stahl SJ, Lee-Huang S, Torchia D, Bax A (1998) Simultaneous measurement of ^{1}H–^{15}N, ^{1}H–$^{13}C'$, and ^{15}N–$^{13}C'$ dipolar couplings in a perdeuterated 30 kDa protein dissolved in a dilute liquid crystalline phase. J Am Chem Soc 120:7385–7386

Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4:171–180

Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222:311–333

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140

Yang D, Tolman JR, Goto NK, Kay LE (1998) An HNCO-based pulse scheme for the measurement of ^{13}C–$^{1}H\alpha$ one-bond dipolar couplings in ^{15}N, ^{13}C labelled proteins. J Biomol NMR 12:325–332

Zacchigna L, Vecchione C, Notte A, Cordenonsi M, Dupont S, Maretto S, Cifelli G, Ferrari A, Maffei A, Fabbro C, Braghetta P, Marino G, Selvetella G, Aretini A, Colonnese C, Bettarini U, Russo G, Soligo S, Adorno M, Bonaldo P, Volpin D, Piccolo S, Lembo G, Bressan GM (2006) Emilin1 links TGF-beta maturation to blood pressure homeostasis. Cell 124:929–942

Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 122:3791–3792