Journal of Molecular Graphics and Modelling 29 (2011) 1015–1024

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

A second Ig-like domain identified in dystroglycan by molecular modelling and dynamics

Maria Cristina De Rosa[a,*], Davide Pirolli[b], Manuela Bozzi[b], Francesca Sciandra[a], Bruno Giardina[b], Andrea Brancaccio[a,**]

[a] CNR – Istituto di Chimica del Riconoscimento Molecolare, c/o Istituto di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Largo F. Vito 1, I-00168 Rome, Italy
[b] Istituto di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Largo F. Vito 1, I-00168 Rome, Italy

ARTICLE INFO

Article history:
Received 18 February 2011
Received in revised form 19 April 2011
Accepted 21 April 2011
Available online 5 May 2011

Keywords:
Dystroglycan
Ig-like fold
Protein modelling
Molecular dynamics
Dystroglycanopathies

ABSTRACT

Dystroglycan (DG) is a cell surface receptor which is composed of two subunits that interact noncovalently, namely α- and β-DG. In skeletal muscle, DG is the central component of the dystrophin–glycoprotein complex (DGC) that anchors the actin cytoskeleton to the extracellular matrix. To date only the three-dimensional structure of the N-terminal region of α-DG has been solved by X-ray crystallography. To expand such a structural analysis, a theoretical molecular model of the murine α-DG C-terminal region was built based on folding recognition/threading techniques. Although there is no a significant (<30%) sequence homology with the N-terminal region of α-DG, protein fold recognition methods found a significant resemblance to the α-DG N-terminal crystallographic structure. Our *in silico* structural prediction identified two subdomains in this region. Amino acid residues ~500–600 of α-DG were predicted to adopt an immunoglobulin-like (Ig-like) β-sandwich fold. Such modeled domain includes the β-DG binding epitope of α-DG and, confirming our previous experimental results, suggests that the linear epitope (residues 550–565) assumes a β-strand conformation. The remaining segment of the α-DG C-terminal region (residues 601–653) is organized in a coil–helix–coil motif. A 20-ns molecular dynamics simulation in explicit water solvent provided support to the predicted Ig-like model structure. The identification of a second Ig-like domain in DG represents another important step towards a full structural and functional description of the α/β DG interface. Preliminary characterization of a novel recombinant peptide (505–600) encompassing this second Ig-like domain demonstrates that it is soluble and stable, further corroborating our *in silico* analysis.

1. Introduction

Dystroglycan (DG) is a widely expressed adhesion molecule composed of two subunits, α and β-DG, originating from post-translational cleavage of a single precursor [1]. α-DG is a highly glycosylated peripheral protein that interacts with several extracellular matrix components, including laminin, agrin and perlecan in muscle, and neurexin in brain. α-DG retains contact with the membrane interacting noncovalently with the β subunit, which is a transmembrane protein whose cytoplasmatic tail binds dystrophin and acts as a scaffold for various adaptor proteins [2]. In skeletal muscle, DG is part of the dystrophin–glycoprotein complex (DGC), a multimeric transmembrane protein complex that connects the actin cytoskeleton to the extracellular matrix, conferring structural stability to the sarcolemma during contraction [3]. In fact, mutations in each component of the DGC can make muscle fibers more susceptible to damage and lead to a number of severe neuromuscular disorders, such as Duchenne muscular dystrophy and limb-girdle muscular dystrophies [4]. No natural mutations have been identified in the DG gene yet, but an altered glycosylation of α-DG, due to genetic defects in specific glycosyltransferases, is associated with a subgroup of congenital muscular dystrophies known as "secondary dystroglycanopathies". In these pathologies, α-DG is characterized by a reduction of its glycosylation both in muscle and brain, where its binding to extracellular matrix molecules is drastically perturbed [5]. Furthermore, α-DG is also the cellular receptor for *Mycobacterium leprae*, the causative agent of leprosy, and for Old World arenaviruses [6,7].

Due to its central role in connecting cells to the extracellular matrix and its involvement in different pathological processes, the solution and analysis of the structure of α-DG represents an important issue. Primary structure analysis and electron microscopy showed that α-DG has a dumbbell-like structure organized in two globular domains, the N- and C-terminal domains, connected by an

* Corresponding author. Tel.: +39 06 3057612; fax: +39 06 3053598.
** Corresponding author. Tel.: +39 06 30155135; fax: +39 06 3053598.
E-mail addresses: mariacristina.derosa@icrm.cnr.it (M.C. De Rosa), andrea.brancaccio@icrm.cnr.it (A. Brancaccio).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.04.008

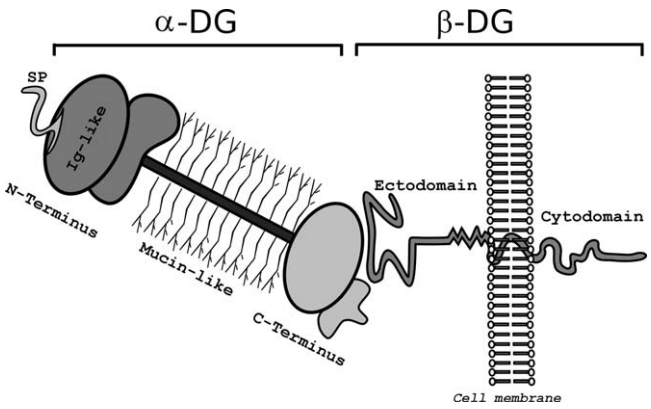

Fig. 1. *Domain organization of the DG complex*. The DG gene encodes a unique polypeptide (that we have defined as pre-DG, see [13]) that is cleaved into two subunits, α-DG and β-DG, interacting through noncovalent interactions. α-DG has a dumbbell-like structure featuring two terminal globular domains connected by a hyperglycosylated and elongated central mucin-like region. α-DG is not homogeneously glycosylated: whilst the mucin-like portion is highly glycosylated, the two globular domains are only scarcely modified by sugars. The α-DG N-terminal portion points towards the extracellular space whereas the C-terminal forms an interface with β-DG; β-DG consists of an extracellular domain, a transmembrane region and a C-terminal cytosolic domain. The α/β-DG interface is formed by the α-DG C-terminal domain and the β-DG extracellular domain. The various domains or regions of the two subunits are indicated, SP refers to the signal peptide that is lost upon ER-targeting of the precursor protein.

elongated central mucin-like region that contains highly glycosylated sequences rich in prolines, serines and threonines [8] (see the synoptic DG domain scheme reported in Fig. 1 and further details within its legend).

Our structure at 2.3-Å resolution of the N-terminal domain of α-DG, the first so far, was solved by X-ray crystallography of a murine recombinant fragment. This three-dimensional (3D) structure revealed the presence of two autonomous subdomains connected by a flexible linker (PDB ID: 1U2C) [9]. The N-terminal module, spanning the amino acids 60–158, shows an Ig-like fold, whereas the C-terminal module, spanning the amino acids 180–303, appears to be very similar to the ribosomal protein S6. The N-terminal domain has been suggested to functionally serve as an essential recognition site for the putative glycosyltransferase LARGE during the maturation process of DG and it has been proposed to be processed into the cell (possibly undergoing massive proteolysis) by a convertase-like activity within the endoplasmic reticulum (ER) [10].

The C-terminal domain of α-DG is involved in the interaction with the β-subunit [11,12]. The β-DG binding site was mapped between the amino acids 550–565 and two residues, G563 and P565, were identified as crucial for the interaction and for the proper processing of the DG precursor [13]. In muscle, the C-terminal domain of α-DG also interacts with biglycan, a small leucine-rich repeat proteoglycan [14]. Interestingly, whilst binding of α-DG to laminin, agrin and perlecan requires O-linked glycosylation of its mucin-like domain, the interactions of α-DG with β-DG and biglycan are independent of glycosylation, underlining the functional importance of the core protein for α-DG function.

Unfortunately, the difficulty in preparing high amounts of soluble recombinant C-terminal domain of α-DG have so far prevented its crystallization and 3D structural analysis. The only hints on the α-DG C-terminal domain structure come from a sequence alignment-based approach, which suggested a similarity with cadherin domains as well as a structural similarity with its own N-terminal domain [15–17]. Such results somehow anticipate and corroborate the ones presented in this study, where we have further and significantly expanded the structural analysis of the C-terminal domain of α-DG, obtaining a reliable model of its 3D-structure using computational structure prediction.

Although the increasing number of available protein structures in the Protein Data Bank has improved the possibility of predicting the structure of a protein via homology modelling, for the majority of sequences this remains impossible. This is the case of α-DG which has no closely related homologues with a known structure, precluding us from any strategy based on homology modelling. In this situation, fold recognition structure prediction methods, which search for distant homology between a target protein of unknown structure and proteins with known structure, can be extremely useful [18]. In order to predict a putative fold of the C-terminal region of α-DG we used two different approaches: the I-TASSER approach [19], which combines the methods of threading, *ab initio* modelling, and structural refinement to predict a protein structure, and a combined HHPRED [20]/MODELLER [21] approach, which integrates the methods of fold recognition with a model building procedure based on the satisfaction of spatial restraints derived from templates.

2. Methodology

2.1. Model building

PSI-BLAST analyses using the murine α-DG C-terminal region protein sequence revealed the absence of homologous proteins of known atomic structure. We therefore used a threading approach, which combines three-tridimensional fold recognition by sequence alignment with template crystal structures and model structure refining. Two different approaches were used: (a) the I-TASSER server (available at http://zhanglab.ccmb.med.umich.edu/I-TASSER/) [19], which uses variants of the threading approach based on sequence profile–profile alignments combined to *ab initio* modelling by iterative implementation of the Threading ASSEmbly Refinement (TASSER). This computational modelling program has proven to be one of the most effective programs at accurately predicting a protein three-dimensional structure blindly from its primary sequence [22,23]; (b) HHPRED (available on a web server http://toolkit.tuebingen.mpg.de/hhpred) [20,24], which uses Hidden Markov Models [25] for template search. Thereafter, the resulting query-template alignment from HHPRED was submitted to the program MODELLER [21] as implemented in Discovery Studio (Accelrys Inc.). Fifty models, optimized by a short simulated annealing refinement protocol available in MODELLER, were generated and their consistency was evaluated on the basis of the PDF (probability density function) violations provided by the program. MODELLER 9 calculations were performed on the workstation HP XW8600 running Red Hat Enterprice Linux 5. The best models among those predicted by the two computational approaches were selected, and candidate models were checked using the programs PROCHECK [26], VERIFY3D [27] and ProSA-Web (Protein Structure Analysis), available at https://prosa.services.came.sbg.ac.at/prosa.php [28]. Visualization and molecular graphics were done using Discovery Studio (Accelrys Inc.).

2.2. Molecular dynamics simulations

The α-DG C-terminal region models were submitted to molecular dynamics simulation using GROMACS v. 4.0.7 [29] and employing the GROMOS96 force field [30]. The structure was immersed in a cubic box with periodic boundary conditions and was solvated with explicit SPC water molecules. The system was then neutralized by 3 Cl^- counterions that were added at random positions to the bulk solvent. The box dimensions (10.2 nm × 10.2 nm × 10.2 nm) were set to allow at least 0.9 nm

between the protein and the box faces on each side. The final system consisted of 1712 protein atoms surrounded by 35,000 water molecules. Before running a simulation, the system was energy minimized for 1000 iterations of steepest descents and then equilibrated for 20 ps, during which the protein atoms were restrained. All restraints were then removed and the temperature was gradually increased from 0 K to 300 K in 50 ps. Next, the molecular dynamics were run for 20 ns at 300 K, and the data were collected every 5 ps. The protein was coupled to an external temperature bath using a v-rescale algorithm with a coupling temperature constant τT of 0.1 ps. van der Waals interactions were modeled using 6–12 Lennard–Jones potentials with a 1.4 nm cutoff. Long-range electrostatic interactions were calculated using Particle Mesh Ewald method, with a cutoff for the real space term of 1.2 nm. Covalent bonds were constrained using the LINCS algorithm. The analysis of the trajectories was carried out using the standard GROMACS tools g_rms, g_rmsf and g_hbond. For comparison, a 20 ns MD simulation of the N-terminal of α-DG (1U2C) was performed using the same simulation protocol.

2.3. Recombinant expression of α-DG(505–600)

The full-length cDNA encoding for murine DG [13,31] was used as a template to generate by PCR a DNA construct, corresponding to the Ig-like C-terminal region of α-DG, α-DG(505–600). The primers used were: forward 5′-CCC**GTCGAC**AGGGTAGATGCCTGGGTGGGAACC-3′ and reverse 5′-CCC**GAATTC**TTGTGGGCGCTTGTGAACATGGAT-3′, *SalI* and *EcoRI* restriction sites are in bold type. The DNA construct obtained was purified and cloned into a bacterial vector which is appropriate to express the protein as thioredoxin fusion product, also containing an N-terminal 6His tag and a thrombin cleavage site [32]. The recombinant fusion protein was expressed in *E. coli* BL21(DE3) Codon Plus RIL strain and purified using nickel affinity chromatography. The fragment of interest was obtained upon thrombin cleavage. Tricine/SDS–PAGE was used to check the purity of the recombinant protein under analysis [33].

2.4. Limited proteolysis, solid-phase assays and fluorescence analysis

Limited proteolysis was carried out as previously described [34]. Namely, 0.1 mg/ml of α-DG(505–600) and β-DG(654–750) were incubated with 20 nM trypsin at 25 °C. Aliquots of the reaction mixtures were collected at increasing times and the proteolysis reactions stopped by boiling the samples in the presence of sample buffer for SDS–PAGE analysis. For solid-phase binding assays, recombinant β-DG(654–750) was biotinylated in 5 PBS buffer at pH 7.4, with 0.5 mg ml^{-1} sulfo-*N*-hydroxylsuccinimido-biotin (S-NHS-biotin, Pierce®). The reaction was carried out for 30 min on ice and in the dark and dialyzed overnight against 10 mM Tris/HCl, 150 mM NaCl, pH 7.4. The optimal dilution for signal detection was determined by dot blot analysis and revealed by enhanced chemioluminescence (Pierce®). α-DG(505–600) and α-DG(485–630) recombinant peptides were coated onto a microtitre plate, whereas biotinylated β-DG(654–750) was used as a soluble ligand at increasing concentrations.

Fluorescence measurements and analysis were carried out as already described elsewhere [35].

3. Results

3.1. Domain(s) identification within the α-DG C-terminal region

Identifying candidate proteins for homology modelling has not proven to be feasible in the case of DG domains, due to the absence of homologous proteins (i.e. displaying a significant degree of sequence identity, >30%) of known atomic structure. To obtain a reliable full-length structural model of the α-DG C-terminal region, we applied two different fold recognition approaches (I-TASSER server and HHPRED/MODELLER). The following protein structures, PDB ID: 1L3W (C-cadherin ectodomain) [36], PDB ID: 1U2C (α-DG N-terminal region) [9], and PDB ID: 1Q55 (EP-cadherin) [37] were chosen by I-TASSER as templates in the modelling, all sharing an immunoglobulin-like fold. 1U2C includes the Ig-like domain of the N-terminal region of α-DG [9] (this template was chosen by HHPRED/MODELLER too) and the other two share a cadherin-like fold, characterized by a prominent β strand secondary structure (Fig. 1S).

To select the models we relied on the *C*-score, a confidence score employed by I-TASSER to estimate the predicted model quality. The *C*-score calculation is based on the significance of the threading template alignments and the convergence parameters of the structure assembly simulations. Typically, a good predicted model is obtained when the estimated level of confidence (*C*-score) is comprised between −5 and 2.

Moreover, both the structural similarity and accuracy of the model are checked using the TM-score and root mean square deviation (RMSD) parameters. The correct topology of the models is obtained for all the structures with TM-scores >0.5, whilst TM-score values <0.17 indicate a random similarity [38].

The best model of the α-DG C-terminal region produced by I-TASSER had a *C*-score equal to −0.67, and its closest structure in the PDB is the one of the α-DG N-terminal region (PDB ID: 1U2C; TM-score 0.77). Despite little amino acid sequence similarity (23% sequence identity), the I-TASSER structural model of the α-DG C-terminal region shows an excellent superimposition with the α-DG N-terminal region, implying common architectural features. According to I-TASSER, the α-DG C-terminal region consists of an Ig-like domain (residues 500–600) and a coil–helix–coil region (residues 601–653).

Similar results were obtained using the HHPRED/MODELLER approach (Fig. 1S). The HHpred server at http://toolkit.tuebingen.mpg.de/hhpred was used to find remotely homologous proteins with known structure. The server generated a list of templates and query template alignments by HMM (hidden Markov matrix) profiling. The N-terminal region of α-DG (PDB ID: 1U2C) was selected as structural template because it scored the highest in the query-template alignment. We then used the pairwise alignment generated by HHpred to generate 50 different models by MODELLER, which were carefully analyzed for energy value and probability density function (PDF) violations. The model with the lowest PDF value was selected as representative of the HHPRED/MODELLER approach. A structural comparison of the two predicted models, together with 1U2C, is shown in Fig. 2A–C, which highlights how the overall structural model generated with the HHPRED/MODELLER approach is in excellent agreement with the one generated by I-TASSER. The Cα-RMSD between the two models is 1.89 Å over residues 485–653 (C-terminal region) and 1.45 Å over residues 500–600 (Ig-like domain).

To evaluate the effect of template choice on the model, we generated structural models of the α-DG C-terminal region using 1L3W and 1Q55 as template structures. A model using multiple templates was also built. In all cases the region comprising residues 500–600 consists of an Ig-like domain (Table 1S), whereas there was no reasonably good alignment with 1L3W and 1Q55 templates for the 601–653 region.

3.2. Model validation

The overall stereochemical quality of the models was assessed by PROCHECK. The model structure obtained from the

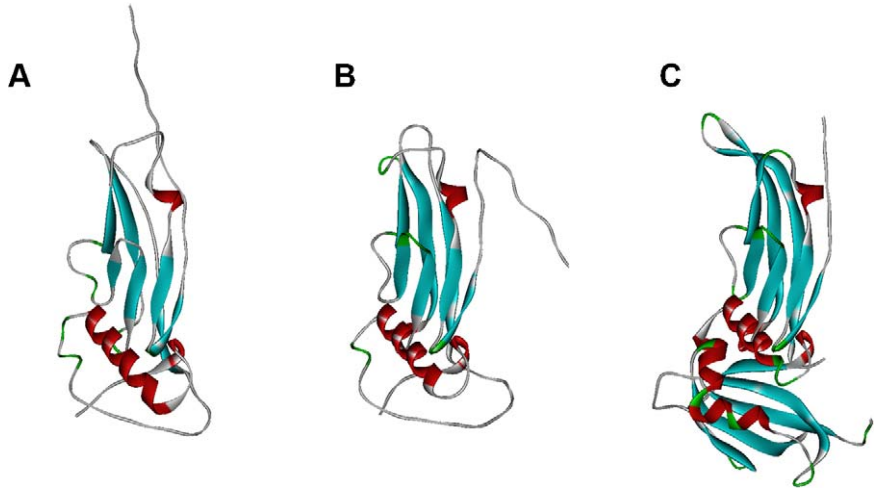

Fig. 2. *3D-structures of α-DG domains.* Structural models of the α-DG C-terminal domain generated by I-TASSER (A) and HHPRED/MODELLER (B) whilst (C) refers to the crystallographic structure (PDB ID: 1U2C) of the N-terminal domain of α-DG [9]. The color schemes on each model indicate the secondary structure elements: *red*, helix; *light blue*, strand; *green* turn; and *gray*, coil. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

HHPRED/MODELLER procedure features a high percentage of residues in the core and allowed region of the Ramachandran plot (94.5 and 4.8% respectively), with only few residues localized in the generously allowed region (0.7%), and no residue found in the disallowed region of the diagram, as compared to the PDB ID: 1U2C template (89%, 11%, 0% and 0%, respectively). On the contrary, the Ramachandran plot of the I-TASSER model showed 80.0% residues in the most favorable region, 15.20% in the allowed region, 2.8% in the generously allowed region and 2.10% in the disallowed region.

The ProSA test was applied to the final models to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. In order to verify whether the interaction energy of each residue with the remainder of the protein is negative (typically positive values correspond to problematic or erroneous parts of a model) the protein structures were submitted to the ProSA web server available at https://prosa.services.came.sbg.ac.at/prosa.php. The ProSA analysis of the final models and the 1U2C template shows that almost all the residues have negative values of interaction energy, whilst only very few residues display positive values (Fig. 3A). Furthermore, the low ProSA *z*-score values obtained (−5.39 for the I-TASSER and −4.92 for the HHPRED/MODELLER model, respectively) confirm the good quality of these models. Fig. 3B reports the *z*-score values calculated using the ProSA-web server.

The reliability of the modeled fold was also checked with VERIFY-3D, which evaluates the compatibility of a given residue in a certain three-dimensional environment. A score below zero for a given residue means that the conformation adopted by that residue in the model is not compatible with its surrounding environment. As shown in Fig. 3C, the VERIFY-3D scores of our models are always positive for the I-TASSER model and are similar to those obtained with the template structure of 1U2C in the region corresponding to the Ig-like domain. On the contrary, the VERIFY-3D analysis suggests a lower confidence ahead of residue 600, especially for the ∼625–630 stretch of residues, of the HHPRED/MODELLER model (negative scores).

3.3. Analysis of the C-terminal Ig-like domain of α-DG and comparison with our previous mutagenesis data

As mentioned in the introductory section, the N-terminal domain of α-DG is the only portion of DG whose high-resolution structure is known [9], and it displays an Ig-like fold from residue 60 to 158. We were able to identify a second Ig-like motif in the α-DG C-terminal region. In both I-TASSER and HHPRED/MODELLER models, the structure is a typical β-sandwich Ig fold containing seven β-strands [39]. Similarly to 1U2C, it belongs to the I-set of the Ig superfamily, and consists of two tightly packed β-sheets, one composed of 4 (β1, β7, β6, β3) and the other of 3 (β2, β5, β4) strands, which are connected by six loops (Fig. 4). In the first β-sheet, the strands β1 and β7 are arranged in a parallel and β7, β6 and β3 in an anti-parallel fashion, whereas the second β-sheet consists of anti-parallel strands (Fig. 4). In the I-TASSER model, residues 521, 522 and 523 in the β2–β3 loop form a turn of α-helix, whereas two helices are present into the HHPRED/MODELLER model, one between β2 and β3 (residues 520–523), the other between β5 and β6 (567–569).

Interestingly, as also shown in the enlarged inset of Fig. 4, the amino acids Pro565 and Gly563, which play a crucial role in the binding with β-DG [13], are located next and within the β5 strand (in the β2–β5–β4 sheet), respectively. The buried position of Gly563 and Pro565 suggests that their role mainly consists in providing structural stability to the Ig-like domain, without a direct involvement of their side-chains in the formation of contacts building the α/β-DG interface. It is known from previous experimental work that alanine substitution of only one of the two crucial amino acids does not reduce significantly the binding affinity towards the β-DG N-terminal domain (K_D = 4.7 μM for the mutant G563A and K_D = 5.5 μM for the mutant P565A) [13], whereas the double mutant G563A/P565A is unable to bind the β-DG N-terminal domain (at least in the concentration range explored).

Gly563 belongs to the β5-strand (Fig. 4) whilst Pro565 is the first residue following this strand. In the mutant G563A, the presence of a foreign $-CH_3$ might perturb the contacts that the β5-strand form with the hydrophobic core of the domain, destabilizing the entire strand. Such an effect of disturbance might become more evident when the Pro residue is concomitantly mutated. Therefore, a role of the β5 strand, and possibly of the entire β2–β5–β4 sheet, could be that of providing an access platform to which the highly flexible (natively unfolded, see [3] and references within) β-DG N-terminal domain could adapt its conformation to form the α/β-DG interface through multiple weak interactions. However, it should also be noted that our mutagenesis analysis showed that Trp551, Phe554 and Asn555, all belonging to the β4 strand, when mutated to Ala did not affect significantly the interaction with the β-DG recombinant ectodomain [31], thus suggesting a specific and more independent role of the β5 strand in the interaction with β-DG (see below).

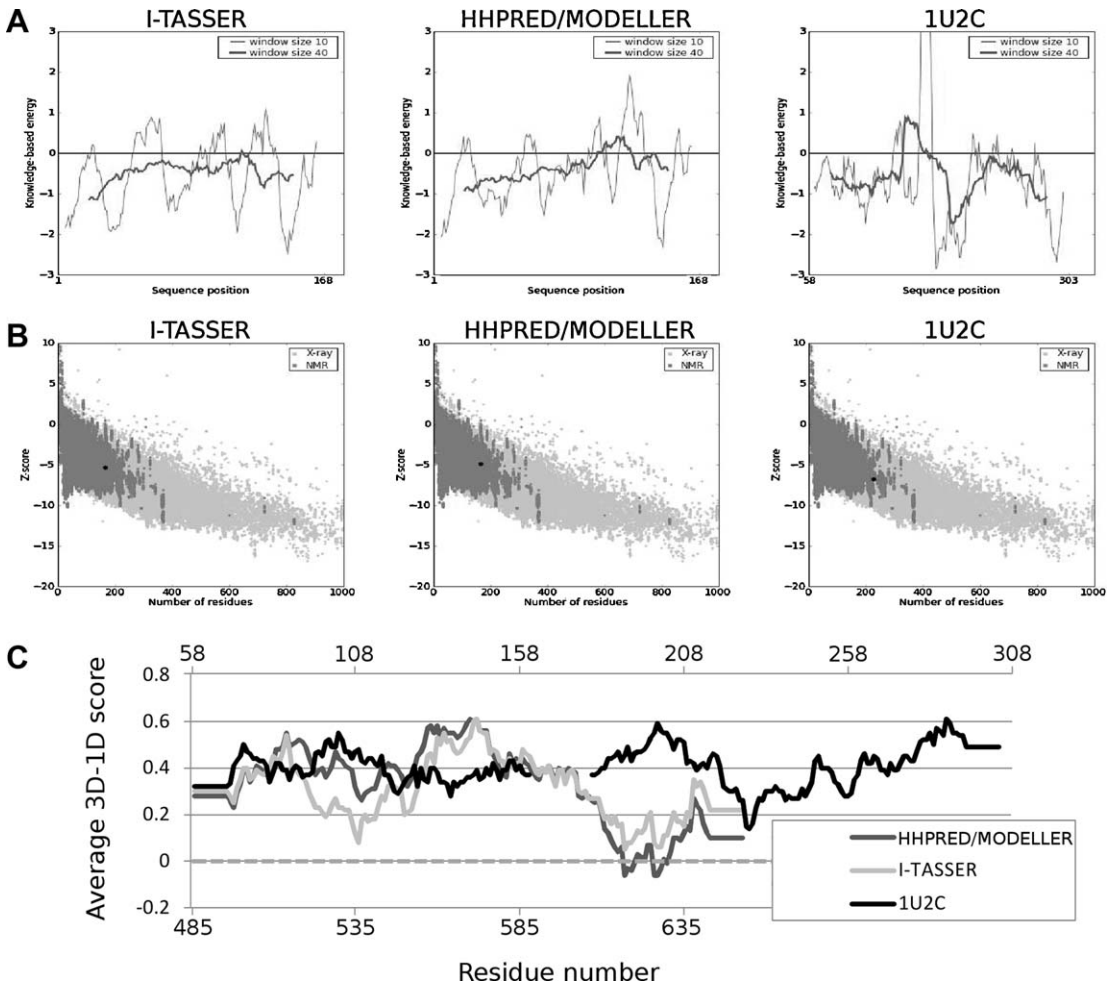

Fig. 3. *Quality assessment of the modeled α-DG C-terminal domain in comparison with the N-terminal structure.* (A) Energy plots from Prosa-web of the I-TASSER model, the HHPRED/MODELLER model and 1U2C. Plot of energies as a function of amino acid sequence position (*thin line*). Residue energies averaged over a sliding window are plotted (*thick line*) as a function of the central residue in a 40 residue-sized window. Positive values correspond to problematic or erroneous parts of the input structure. (B) ProSA-web *z*-scores of all protein chains in PDB determined by X-ray crystallography (*light*) or NMR spectroscopy (*dark*) with respect to their length. The *z*-score of the I-TASSER model, HHPRED/MODELLER model and 1U2C are highlighted as large *black* dots. (C) VERIFY-3D plots of I-TASSER and HHPRED/MODELLER models compared to the 1U2C structure [9]. The top *x*-scale values refer to the 1U2C structure.

3.4. Molecular dynamics and secondary structure analysis

In order to evaluate the stability of the putative Ig-like domain in the C-terminal region of α-DG, we followed the evolution of the predicted structures in molecular dynamics (MD) simulations. Both the I-TASSER and HHPRED/MODELLER models were submitted to MD carried out in explicit solvent with the GROMOS96 force field at 300 K. Four snapshots were extracted from the MD trajectory at regular intervals of 5 ns and the programs PROCHECK, ProSA-web and VERIFY3D were used to validate the structures. No significant variation in the values calculated for the modeled structures was observed over the course of the simulations. In particular, the PROCHECK Ramachandran plot analysis shows that more than 95% of the amino acids fall into the core/allowed Ramachandran plot regions. VERIFY3D and ProSA-web evaluation scores are rated good and similar to those obtained with the modeled starting structures (Table 1).

The stability and behaviour of each model were assessed by analyzing the root-mean-square deviation (RMSD) with respect to the starting structure, the root-mean-square fluctuations (RMSF), the radius of gyration and the secondary structure, which was assigned according to the rules defined by the Kabsch and Sander's DSSP program [40]. In order to evaluate the stability of the predicted Ig-like domain, molecular dynamics simulation of the N-terminal of α-DG was also performed.

At a first glance, the Cα-RMSD profiles, shown in Fig. 5, display a similar behaviour for the entire α-DG C-terminal region in the two models (Fig. 5A). A closer inspection of the RMSD indicates that its variation is mainly due to the coil–helix–coil motif, whilst the Ig-like domain is rather stable in both simulations (Fig. 5B). However it is worth noting that for the I-TASSER model the values of RMSD are smaller than those for the HHPRED/MODELLER throughout the whole simulation time of 20 ns, indicating that the MD simulation preserves the starting structure generated by I-TASSER better than that generated by HHPRED/MODELLER. Moreover, the RMSD values did not show any large fluctuation compared with the crystal structure of the α-DG N-terminal Ig-like domain (Fig. 5B).

The deviations described above indicate how far the protein has moved from the starting structure. The Cα root-mean-square fluctuations (RMSF) with respect to the average MD conformation are used as a mean of describing flexibility differences among residues. A region with a large RMSF experiences more marked changes in position relative to its average position, and is therefore more flexible. A region with a low RMSF shows more limited movements during the MD simulation in relation to its average position, and is therefore less flexible. Fig. 5C shows the RMSF

Table 1
Quality assessment of modeled α-DG C-terminal region at different snapshots (every 5 ns) of the MD simulations.

Model/time (ns)	PROCHECK				ProSA-web *z*-score	VERIFY-3D score
	Core (%)	Allowed (%)	Generously allowed (%)	Disallowed (%)		
I-TASSER/0	80.0	15.2	2.8	2.1	−5.39	>0.05
I-TASSER/5	74.5	23.4	2.1	0.0	−6.42	>0.02
I-TASSER/10	70.3	27.6	1.4	0.7	−6.41	>0.09
I-TASSER/15	76.6	20.0	2.1	1.4	−5.95	>0.18
I-TASSER/20	68.3	28.3	2.1	1.4	−5.79	>0.07
HHPRED/MODELLER/0	94.5	4.80	0.7	0.0	−4.92	>−0.06
HHPRED/MODELLER/5	73.8	21.4	3.4	1.4	−5.08	>0.19
HHPRED/MODELLER/10	75.2	21.4	1.4	2.1	−5.17	>0.06
HHPRED/MODELLER/15	73.1	24.1	1.4	1.4	−5.45	>0.04
HHPRED/MODELLER/20	78.6	18.6	0.7	2.1	−5.66	>0.04

for the I-TASSER and the HHPRED/MODELLER model simulations. There was no significant difference in the RMSF between the I-TASSER and the HHPRED/MODELLER models. The RMSF profile also shows that all the β strand structures were consistently stable, with RMSF values around 1.0 Å, whilst the loops connecting the β strands and the coil–helix–coil motif (601–653) were in general much more flexible. A similar behaviour was observed with the crystallographic 1U2C structure, which however displayed significantly higher RMSF values in the 6-residue longer β6–β7 loop (Fig. 5C). Consistent with the RMSD analysis, the RMSFs of the I-TASSER model are slightly smaller than those of the HHPRED/MODELLER model in the β strands. The radius of gyration (R_g), which is an indicator of how far the atoms in a molecule are from its center of mass, was plotted as a function of time (Fig. 5D). The value of the radius of gyration of the Ig-like domain remains essentially identical to its starting value, with a mean (over the length of the trajectory) value of 1.38 nm and 1.29 nm (I-TASSER and HHPRED/MODELLER, respectively) and a standard deviation of only 0.01 nm and 0.02 nm, respectively. Remarkably, the plot of the radius of gyration for the N-terminal Ig-like domain is very similar to that observed for the I-TASSER model (Fig. 5D). The time evolution of the secondary structure elements is depicted in Fig. 5E–G. The secondary structure plots also reveal that β sheets are extremely stable during the 20 ns simulations of both crystallographic and *in silico* structures of Ig-like domains (Fig. 5E–G). The small helix within the β2–β3 loop is lost during the I-TASSER run whereas it is more persistent during the HHPRED/MODELLER one. Some significant differences between the two plots are observed in the extreme C-terminal of α-DG, which may suggest that there is something amiss with the model structure of the 601–653 region. It is worthwhile to note that, in both molecular modelling approaches, the quality of the alignment was significantly higher for the Ig-like domain as compared with the extreme C-terminal portion of α-DG. As observed in a previous section, the same VERIFY-3D analysis, used for identifying regions of improper folding, suggested low confidence in the 625–630 region of α-DG (Fig. 3C).

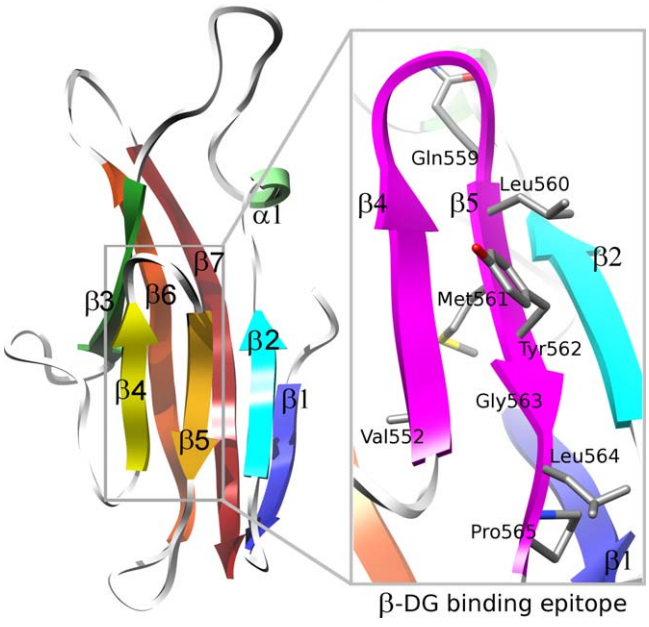

Fig. 4. *The Ig-like domain identified within the α-DG C-terminal region.* The structure reported is the one referring to the I-TASSER model and different colours have been used to highlight the various strands forming the β-sandwich. In the enlarged inset, the β-DG binding epitope is shown in *magenta* and the side chains referring to Glu559, Leu560, Met561, Tyr562, Leu564 and Pro565 protruding from strand β5 are displayed in stick representation and coloured by atom type (the image was produced using the UCSF Chimera program).

In order to analyze the packing interactions between the β-sheets of the predicted Ig-like domain, we monitored the separation of the sheets during the MD simulation. We observed an average distance of 10.1 ± 0.6 Å (I-TASSER model) and 11.8 ± 0.6 Å (HHPRED/MODELLER) between the Cα positions of residues Met561 (strand β5) and Phe591 (strand β6) which maintain the hydrophobic core. The average distance between the corresponding residues in 1U2C (Leu115 and Phe150, respectively) was 13.2 ± 0.7 Å (12.4 Å in the crystal structure).

These values are comparable with those reported for Ig-like domains lacking the disulfide bridge between the β-sheets, where the cysteine residues are replaced by hydrophobic residues that maintain the hydrophobic core [41]. Distances between 6.1–7.1 and 7.2–11.7 Å respectively for domains with and without the canonical disulfide bridge are reported [41].

Such differences in compactness of the α-DG Ig-like domain may provide insight into the differences observed in Fig. 5B.

The secondary structural elements of the Ig-like domain in the predicted model, in comparison to those of 1U2C and of a structural neighbour of 1U2C such as the cysteine peptidase inhibitor (ICP) from *Leishmania mexicana* (PDB ID: 2C32) [42], are displayed in Fig. 6. The alignment, in which the strands are highlighted, shows that the overall organization of the seven β strands is conserved within the three Ig-like proteins, the relevant difference being the length of the β2–β3 and β4–β5 interconnecting loops that are significantly longer in the predicted Ig-like of DG and 2C32, respectively.

On the whole, the most consistent finding emerging from molecular dynamics simulations is the overall energetic stability of the predicted Ig-like domain.

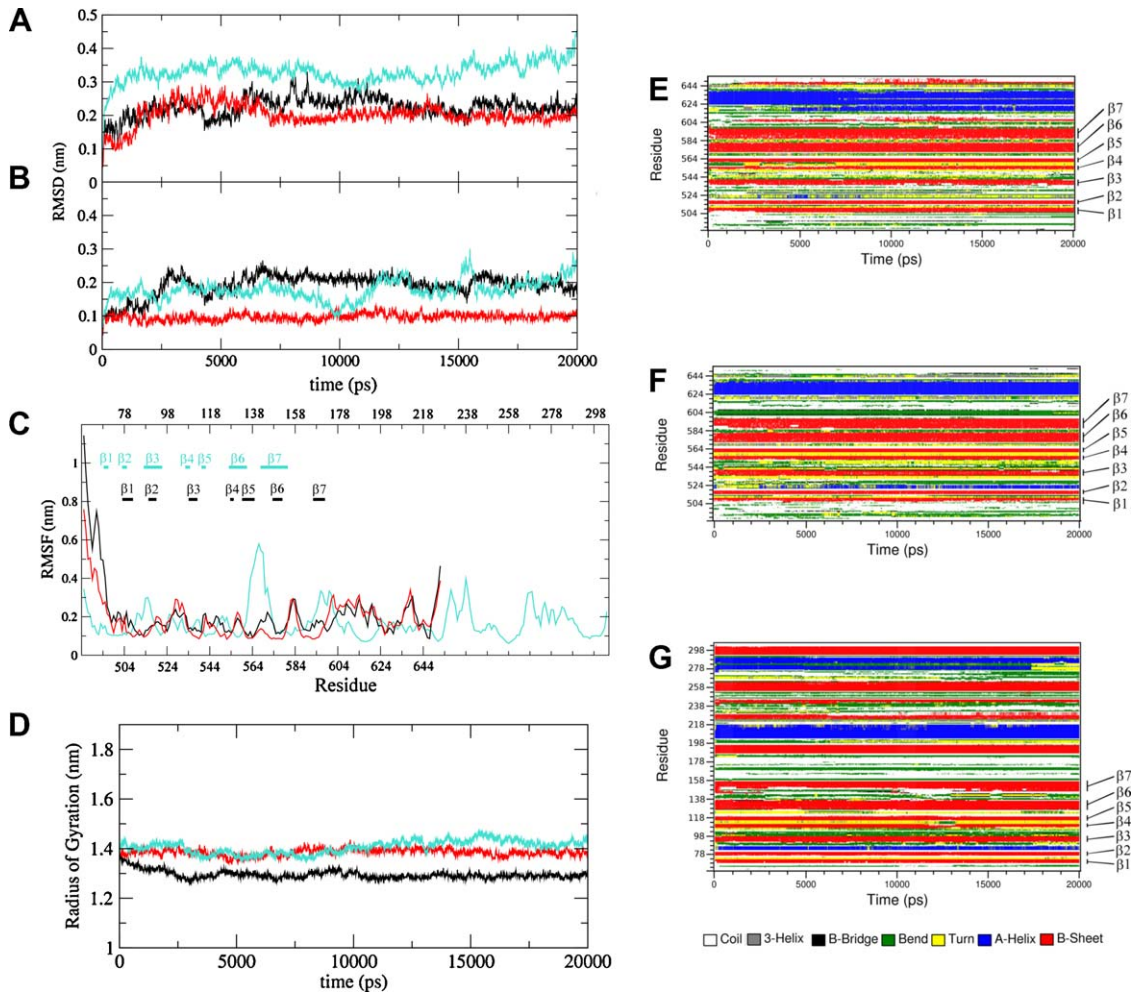

Fig. 5. *Analysis of MD trajectories*. Time series of the Cα atoms RMSD from the starting structure for the entire C-terminal region (A) and for the Ig-like domain (B). Residue-based Cα RMSF relative to the average structure (C). Time series of the radius of gyration of Cα atoms of the Ig-like domain (D). Plots of the I-TASSER model are shown in *red*, those for the HHPRED/MODELLER model in *black*. Plots of the 1U2C structure also are shown for comparison (*turquoise*). Classification of the MD trajectory in terms of secondary structure elements generated by DSSP [39] for the I-TASSER model (E), for the HHPRED/MODELLER model (F) and for 1U2C structure (G). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.5. *Expression of a novel recombinant peptide spanning the second α-DG Ig-like domain*

Based on the new information obtained from the model herein described, we have amplified and cloned a DNA construct that allowed the expression and purification in *E. coli* of a novel recombinant peptide spanning the α-DG C-terminal Ig-like domain (between positions 505 and 600) (Fig. 7A). The peptide has been obtained upon thrombin cleavage (see the details in Section 2) and a preliminary characterization indicates that it might well represent an isolated and stable protein domain, as indicated: (i) by a limited proteolysis test in comparison with the natively unfolded ectodomain of β-DG (Fig. 7B and C), (ii) by a fluorescence emission spectrum (see inset in Fig. 7E, where the spectrum shows a maximum wavelength at 340 nm, characteristic of buried aromatics), as well as an unfolding behaviour (Fig. 7E), typical of a polypeptide chain with a compact tertiary structure. Specifically, the denaturation curve of α-DG(505–600), recorded as a function of increasing guanidine hydrochloride concentration, is highly similar to the unfolding transition measured for the α-DG N-terminal Ig-like domain [35], with the same midpoint denaturant concentration of 1.6 ± 0.2 M; such a result corroborates the *in silico* data, pointing to the presence of a second Ig-like domain in the C-terminal region of α-DG.

In addition, the peptide retains the ability of binding recombinant β-DG in solid-phase binding assays (Fig. 7D), indicating that the moieties mandatory for β-DG binding are unaffected in this new peptide, that will be employed in further structural studies.

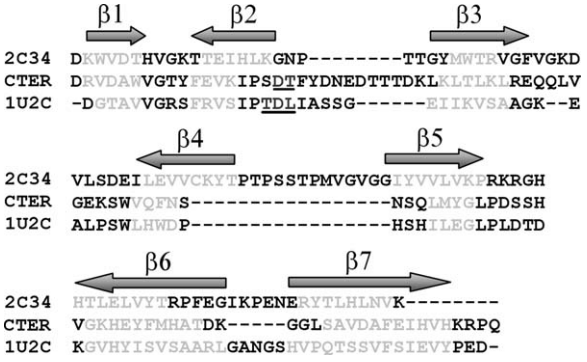

Fig. 6. *β-strands organization within the DG Ig-like domains*. The position and orientation of the β-strand elements, reported as arrows above the sequences, are evidenced in a multiple alignment of primary structures of PDB ID: 1U2C (N-terminal domain of α-DG), PDB ID: 2C34 (a close structural neighbour of 1U2C) and of the newly modeled Ig-like domain within the C-terminal portion of α-DG (reported as CTER). The residues belonging to β-strands are reported in light gray. The stretches of residues forming short helices are underlined.

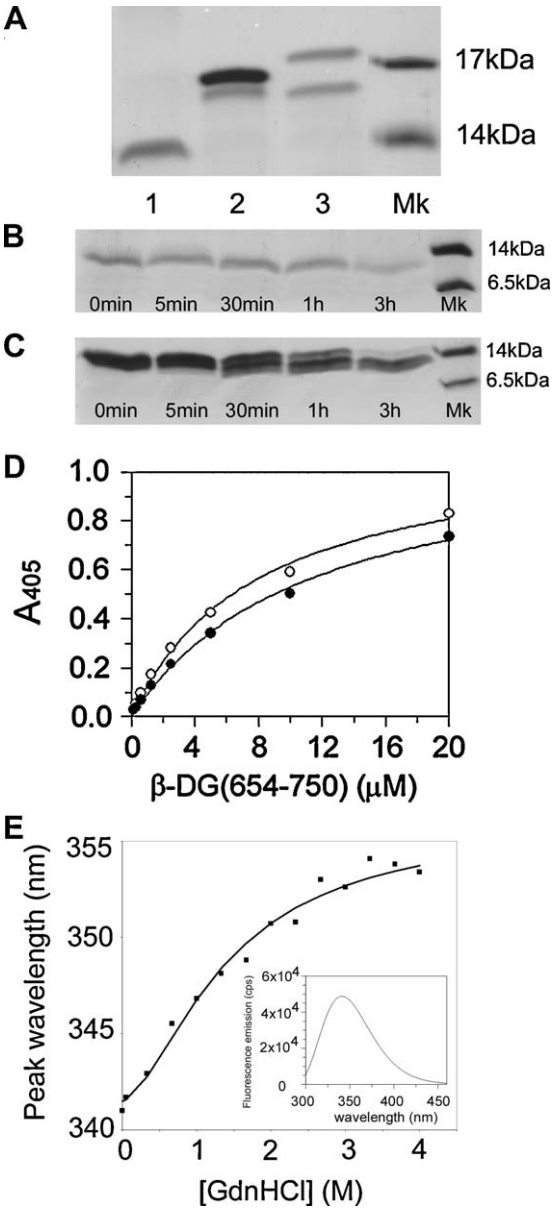

Fig. 7. *Recombinant expression and preliminary characterization of the novel Ig-like domain*. (A) The electrophoretic mobility of the purified Ig-like C-terminal domain of α-DG, α-DG(505–600) (lane 1), is compared to those of larger α-DG C-terminal constructs, such as α-DG(485–620) (lane 2), and α-DG(485–630) (lane 3), in which a certain extent of C-terminal degradation was always observable [11]. Mk, molecular mass markers. (B) and (C) refer to a SDS–PAGE analysis of α-DG(505–600) and β-DG(654–750) respectively, carried out upon limited proteolysis with 20 nM trypsin. Aliquots of the reaction mix were collected at increasing times and analyzed by SDS–PAGE. Trypsin digests the natively unfolded β-DG(654–750), and already in 30 min there is an evident smearing caused by multiple fragmentations, whereas has no relevant effects on the α-DG(505–600) band, whose intensity is weakly reduced after 3 h of incubation. Mk, molecular mass markers. (D) Solid-phase binding assay, α-DG(505–600) (*full circles*) and α-DG(485–630) (*open circles*) both tested for binding towards soluble biotinylated β-DG(654–750). Each continuous line corresponds to a representative experiment (from a set of at least three experiments producing similar results), and was obtained by fitting experimental data to a single class of equivalent binding sites equation. The apparent K_D values measured, for α-DG(485–630) ~6 uM and α-DG(505–600) ~8 uM, are in full agreement with our previous experiments [12]. (E) Unfolding behaviour of α-DG(505–600) measured by fluorescence maximum emission peak changes as a function of guanidine hydrochloride concentration. The excitation wavelength was 280 nm, the protein concentration was 150 nM. Data have been fit as described in Ref. [35]. A fluorescence maximum shift from 340 nm, typical of buried tryptophan residues, to 353 nm, characteristic of totally exposed tryptofans, clearly indicates that an unfolding transition is taking place. Inset of panel E: fluorescence emission spectrum of α-DG(505–600) in the absence of guanidine hydrochloride, showing a shape and position of the maximum typical of a folded polypeptide.

4. Discussion

4.1. A second Ig-like domain can be reliably modeled within α-DG

The presence of an additional Ig-like domain (belonging to the immunoglobulin fold superfamily, which reaches out to include also the cadherin-like domains) in α-DG was postulated by Ponting and co-workers, who proposed the presence of two cadherin-like domains (CDHL) within the two globular portions of α-DG as the output of a thorough alignment analysis [15]. However, in a computational investigation that followed, it was proposed that such domains would not display the typical calcium-binding sites of the cadherin domains [16,17], an evidence that was confirmed also by our crystallographic work on the N-terminal of α-DG, showing the presence of an Ig-like fold between 58 and 160 a.a. positions with no calcium binding sites [9].

Therefore, we decided to apply a computational method in order to obtain a 3D-model of the C-terminal portion of α-DG and to further verify whether an Ig-like structural arrangement was to be finally confirmed. In the absence of a crystallographic structure, computational modelling tools can be applied to obtain a structural model able to address some important questions related to protein function.

"Threading" is a three-dimensional structure prediction technique especially useful when there is not enough sequence similarity between the target sequence and a known three-dimensional structure and, therefore, comparative modelling cannot be applied. In this study, two different threading approaches have been used which displayed remarkable reliability and agreement in protein structure prediction, I-TASSER and a combined HHPRED/MODELLER approach. Whereas the first one uses variations of the threading approach based on sequence profile–profile alignments combined to *ab initio* modelling, the latter performs a fold recognition analysis using the database of Hidden Markov Models (HMM) domains of known structure from the protein databank (PDB), followed by an automated model-building procedure based on MODELLER. Both the approaches indicated that the N-terminal and C-terminal domains of α-DG share a similar three-dimensional fold (RMSD = 2.1 Å), despite possessing low sequence identity (23%).

4.2. Reference to our previous biochemical work on the α/β interface: importance of the β5-strand for β-DG binding

Several prominent features of the model herein described are in line with our previous biochemical work. First of all, it is worth noting that the recombinant α-DG(485–600) peptide displayed an unfolding behaviour, as measured by intrinsic fluorescence, reminiscent of an Ig-like structural arrangement [35]. In addition, the presence of an Ig-like domain is compatible with some of the experimental binding data that we have so far collected on the α/β-DG interface. It is important to note that the affinity (within the low micromolar range) towards the recombinant β-DG N-terminal domain of the recombinant construct α-DG(485–651), spanning the entire α-DG C-terminal domain, is very similar to the one displayed by α-DG(485–600) [12], a peptide devoid of the coil–helix–coil motif. This confirms that the entire α-DG C-terminal domain could be divided into two functionally and structurally distinct regions, the Ig-like domain, containing the β-DG binding epitope, and the coil–helix–coil motif.

As described in Section 3, our model gives an important *in silico* support to these experimental findings, showing how the primary structure elements that we believe are the most important for recognizing β-DG belong to a specific β-strand (i.e. β5), that could be therefore considered as the most important functional hot-spot within the whole Ig-like domain. Curiously, this β5 strand is very

well conserved in the N-terminal Ig-like domain of α-DG, that does not bind β-DG [43], and Gly563 and Pro565 are located in identical topological position, with respect to the β5 strand, in both the conserved strands (see Fig. 6) as well as in all the known DG sequences available so far, including DG-like proteins of invertebrates [13]. It is likely that the whole of "β5 scaffold" (including the strand and the following loop), rather than the specific side-chains of Gly563 and Pro565 only, is crucial both for the overall stability of the α-DG C-terminal Ig-like domain and for establishing contacts with the β-DG ectodomain.

It is clear that our new 3D model could be of paramount importance for directing the next recombinant steps necessary to fully unravel the fine details of the α/β binding mechanism. In particular, the identification of specific β-strand(s) harbouring the β-DG binding epitope will be the basis for the rational design of new peptides, or for directing further mutagenesis, aiming at the functional modulation of the dystroglycan complex.

4.3. Further insight for DG domains analysis

Another interesting aspect of our work is that the identification of a second Ig-like domain within DG could be phylogenetically relevant. Based on primary structure similarities, it is reasonable to assume that such domain organization of DG is common to the majority of DG orthologues so far analyzed, and in particular to vertebrate DGs. Besides, the very presence of two Ig-like domains within α-DG could suggest an ancestral exon shuffling event, anticipating the insertion of the giant intron (ranging from ∼10 to 20 Kb in mammals) that is intertwined within the sequence coding for the first Ig-like, whose exact mechanism and/or consequences need to be carefully investigated [44,45].

In addition, the growing biological importance of DG Ig-like domains is further highlighted by a recent work showing the first case of primary mutation of *dag1* in zebrafish that is affecting the *core protein* of DG but not its glycosylation shell [46]. In fact, the mutated residue (Val567 to Asp) is very well conserved, being Val or Ile, within all the DG sequences so far identified [13], and located within the β7-strand of the second Ig-like domain that we have herein identified.

From an exquisitely experimental point of view, the outcome of our analysis is very promising. Indeed, we demonstrated that the first part of what was previously considered as belonging to the C-terminal of α-DG (i.e. approx residues 485–500) is likely to be mostly unstructured and organized in a fashion that is reminiscent of the 30–60 portion that anticipates the Ig-like N-terminal domain of α-DG [9]. Especially, in our set of recombinantly expressed α-DG C-terminal peptides, this disorganized and elongated stretch of residues could reduce the stability and solubility of the protein, a behaviour that we have repeatedly observed and has precluded further structural analyses on such peptides.

The analysis hereby described will be informative for the rational design of our next round of recombinant constructs; Fig. 7 reports a preliminary analysis on a novel 505–600 peptide. As already mentioned, in this peptide we have significantly shortened the N-terminal portion that does not belong to the Ig-like domain, as suggested by both models (see Fig. 2B and C), cutting out an elongated stretch that was instead present in the 485–600 peptide [12].

We are currently undertaking the production of an entirely novel panel of recombinant peptides by testing slightly different domain-borders, in search of the best soluble candidate that would better preserve the secondary and tertiary structural elements revealed by the model, and that would be then expressed and purified in higher amounts for crystallization studies or direct submission to high-field NMR, in isotope-labeled form, for structure determination in solution.

Acknowledgements

We are grateful to Dr. Maria Giulia Bigotti (Rome) for critical reading of the manuscript. Computational resources for MD simulations were provided by CASPUR – Inter-University Consortium for Supercomputing application for University and Research (Standard HPC Grant 2010). FS is the recipient of a fellowship from the Association Française contre les Myopathies (AFM).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.04.008.

References

[1] O. Ibraghimov-Beskrovnaya, J.M. Ervasti, C.J. Leveille, C.A. Slaughter, S.W. Sernett, K.P. Campbell, Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix, Nature 355 (1992) 696–702.
[2] C.J. Moore, S.J. Winder, Dystroglycan versatility in cell adhesion: a tale of multiple motifs, Cell Commun. Signal. 8 (2010) 3.
[3] M. Bozzi, S. Morlacchi, M.G. Bigotti, F. Sciandra, A. Brancaccio, Functional diversity of dystroglycan, Matrix Biol. 28 (2009) 179–187.
[4] R.D. Cohn, K.P. Campbell, Molecular basis of muscular dystrophies, Muscle Nerve 23 (2000) 1456–1471.
[5] D.E. Michele, R. Barresi, M. Kanagawa, F. Saito, R.D. Cohn, J.S. Satz, J. Dollar, L. Nishino, R.I. Kelley, H. Somer, V. Straub, K.D. Mathews, S.A. Moore, K.P. Campbell, Post-translational disruption of dystroglycan–ligand interactions in congenital muscular dystrophies, Nature 418 (2002) 417–422.
[6] A. Rambukkana, H. Yamada, G. Zanazzi, T. Mathus, J.L. Valzer, P.D. Yurchenco, K.P. Campbell, V.A. Fischetti, Role of α-dystroglycan as a Schwann cell receptor for *Mycobacterium leprae*, Science 282 (1998) 2076–2079.
[7] W. Cao, M.D. Henry, P. Borrow, H. Yamada, J.H. Elder, E.V. Ravkov, S.T. Nichel, R.W. Compans, K.P. Campbell, M.B. Oldstone, Identification of α-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus, Science 282 (1998) 2079–2081.
[8] A. Brancaccio, T. Schulthess, M. Gesemann, J. Engel, Electron microscopic evidence for a mucin-like region in chick muscle α-dystroglycan, FEBS Lett. 368 (1995) 139–142.
[9] D. Bozic, F. Sciandra, D. Lamba, A. Brancaccio, The structure of the N-terminal region of murine skeletal muscle α-dystroglycan discloses a modular architecture, J. Biol. Chem. 279 (2004) 44812–44816.
[10] M. Kanagawa, F. Saito, S. Kunz, T. Yoshida-Moriguchi, R. Barresi, Y.M. Kobayashi, J. Muschler, J.P. Dumanski, D.E. Michele, M.B. Oldstone, K.P. Campbell, Molecular recognition by LARGE is essential for expression of functional dystroglycan, Cell 117 (2004) 953–964.
[11] M. Bozzi, G. Veglia, M. Paci, F. Sciandra, B. Giardina, A. Brancaccio, A synthetic peptide corresponding to the 550–585 region of α-dystroglycan binds β-dystroglycan as revealed by NMR spectroscopy, FEBS Lett. 22 (2001) 210–214.
[12] F. Sciandra, M. Schneider, B. Giardina, S. Baumgartner, T.C. Petrucci, A. Brancaccio, Identification of the β-dystroglycan binding epitope within the C-terminal region of α-dystroglycan, Eur. J. Biochem. 268 (2001) 4590–4597.
[13] F. Sciandra, M. Bozzi, S. Morlacchi, A. Galtieri, B. Giardina, A. Brancaccio, Mutagenesis at the α–β interface impairs the cleavage of the dystroglycan precursor, FEBS J. 276 (2009) 4933–4945.
[14] M.A. Bowe, D.B. Mendis, J.R. Fallon, The small leucine-rich repeat proteoglycan biglycan binds to α-dystroglycan and is upregulated in dystrophic muscle, J. Cell Biol. 148 (2000) 801–810.
[15] N.J. Dickens, S. Beatson, C.P. Ponting, Cadherin-like domains in α-dystroglycan, α/ε-sarcoglycan and yeast and bacterial proteins, Curr. Biol. 12 (2002) 197–199.
[16] L. Cao, X. Yan, C.W. Borysenko, H.C. Blair, C. Wu, L. Yu, CHDL: a cadherin-like domain in proteobacteria and cyanobacteria, FEMS Microbiol. Lett. 251 (2005) 203–209.
[17] P. Hulpiau, F. van Roy, Molecular evolution of the cadherin superfamily, Int. J. Biochem. Cell Biol. 41 (2009) 349–369.
[18] D. Fischer, Servers for protein structure prediction, Curr. Opin. Struct. Biol. 16 (2006) 178–182.
[19] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics 9 (2008) 40.
[20] J. Söding, A. Biegert, A. Lupas, The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res. 33 (2005) W244–W248.
[21] A. Sali, T. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol 234 (1993) 779–815.
[22] J. Battey, J. Kopp, L. Bordoli, R. Read, N. Clarke, T. Schwede, Automated server predictions in CASP7, Proteins 69 (Suppl. 8) (2007) 68–82.
[23] Y. Zhang, Template-based modeling and free modeling by I-TASSER in CASP7, Proteins 69 (Suppl. 8) (2007) 108–117.
[24] J. Söding, Protein homology detection by HMM–HMM comparison, Bioinformatics 21 (2005) 951–960.

[25] R. Hughey, A. Krogh, Hidden Markov models for sequence analysis: extension and analysis of the basic method, Comput. Appl. Biosci. 12 (1996) 95–107.
[26] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton, Stereochemical quality of protein structure coordinates, Proteins 12 (1992) 345–364.
[27] R. Luthy, J.U. Bowie, D.D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85.
[28] M. Wiederstein, M. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407–410.
[29] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, H. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[30] U. Stocker, W.F. van Gunsteren, Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data, Proteins 40 (2000) 145–153.
[31] M. Bozzi, F. Sciandra, P. Torreri, E. Pavoni, T.C. Petrucci, B. Giardina, A. Brancaccio, Concerted mutation of Phe residues belonging to the β-dystroglycan ectodomain strongly inhibits the interaction with α-dystroglycan in vitro, FEBS J. 273 (2006) 4929–4943.
[32] R.A. Kammerer, T. Schulthess, R. Landwehr, A. Lustig, D. Fischer, J. Engel, Interaction of agrin with laminin requires a coiled-coil conformation of the agrin-binding site within the laminin γ1 chain, J. Biol. Chem. 273 (1998) 10602–10608.
[33] H. Schägger, G. von Jagow, Tricine–sodium dodecyl sulfate–polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem. 166 (1987) 368–379.
[34] A. Brancaccio, T. Schulthess, M. Gesemann, J. Engel, The N-terminal region of α-dystroglycan is an autonomous globular domain, Eur. J. Biochem. 246 (1997) 166–172.
[35] E. Di Stasio, P. Bizzarri, F. Misiti, E. Pavoni, A. Brancaccio, A fast and accurate procedure to collect and analyze unfolding fluorescence signal: the case of dystroglycan domains, Biophys. Chem. 107 (2004) 197–211.
[36] T. Boggon, J. Murray, S. Chappuis-Flament, E. Wong, B. Gumbiner, L. Shapiro, C-cadherin ectodomain structure and implications for cell adhesion mechanisms, Science 296 (2002) 1308–1313.
[37] W. He, P. Cowin, D. Stokes, Untangling desmosomal knots with electron tomography, Science 302 (2003) 109–113.
[38] Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality, Proteins 57 (2004) 702–710.
[39] P. Bork, L. Holm, C. Sander, The immunoglobulin fold. Structural classification, sequence patterns and common core, J. Mol. Biol. 242 (1994) 309–320.
[40] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[41] D.M. Halaby, A. Poupon, J. Mornon, The immunoglobulin fold family: sequence analysis and 3D structure comparisons, Protein Eng. 12 (1999) 563–571.
[42] B. Smith, N. Picken, G. Westrop, K. Bromek, J. Mottram, G. Coombs, The structure of *Leishmania mexicana* ICP provides evidence for convergent evolution of cysteine peptidase inhibitors, J. Biol. Chem. 281 (2006) 5821–5828.
[43] E. Di Stasio, F. Sciandra, B. Maras, F. Di Tommaso, T.C. Petrucci, B. Giardina, A. Brancaccio, Structural and functional analysis of the N-terminal extracellular region of β-dystroglycan, Biochem. Biophys. Res. Commun. 266 (1999) 274–278.
[44] M.M. Buljan, A. Bateman, The evolution of protein domain families, Biochem. Soc. Trans. 37 (2009) 751–755.
[45] T. Kawashima, S. Kawashima, C. Tanaka, M. Murai, M. Yoneda, H.N. Putnam, D.S. Rokhsar, M. Kanehisa, N. Satoh, H. Wada, Domain shuffling and the evolution of vertebrates, Genome Res. 19 (2009) 1393–1403.
[46] V. Gupta, G. Kawahara, S.R. Gundry, A.T. Chen, W.I. Lencer, Y. Zhou, L.I. Zon, L.M. Kunkel, A.H. Beggs, The zebrafish *dag1* mutant: a novel genetic model for dystroglycanopathies, Hum. Mol. Genet. 20 (2011) 1712–1725.