



Covaceuszach, S., Bozzi, M., Bigotti, M. G., Sciandra, F., Konarev, P. V., Brancaccio, A., & Cassetta, A. (2017). Structural flexibility of human - dystroglycan. *FEBS Open Bio*, 7(8), 1064-1077.  
<https://doi.org/10.1002/2211-5463.12259>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1002/2211-5463.12259](https://doi.org/10.1002/2211-5463.12259)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1002/2211-5463.12259> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

## Structural flexibility of human $\alpha$ -dystroglycan

Sonia Covaceuszach<sup>1,†</sup>, Manuela Bozzi<sup>2,3,†</sup>, Maria Giulia Bigotti<sup>4</sup>, Francesca Sciandra<sup>3</sup>, Petr Valeryevich Konarev<sup>5,6</sup>, Andrea Brancaccio<sup>3,4</sup> and Alberto Cassetta<sup>1</sup>

1 Istituto di Cristallografia, CNR, Trieste Outstation, Italy

2 Istituto di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Roma, Italy

3 Istituto di Chimica del Riconoscimento Molecolare, CNR c/o Università Cattolica del Sacro Cuore, Roma, Italy

4 School of Biochemistry, University of Bristol, UK

5 Shubnikov Institute of Crystallography of Federal Scientific Research Centre, "Crystallography and Photonics" of Russian Academy of Sciences, Moscow, Russia

6 National Research Centre "Kurchatov Institute", Moscow, Russia

### Keywords

conformational stability; muscular dystrophy; small-angle X-ray scattering; structural flexibility; X-ray crystal structure;  $\alpha$ -Dystroglycan

### Correspondence

A. Brancaccio, Istituto di Chimica del Riconoscimento Molecolare (ICRM), CNR c/o Università Cattolica del Sacro Cuore, Largo Francesco Vito 1, 00168 Roma, Italy  
Fax: +39 (06) 305-3598  
Tel: +39 (06) 301-55135  
E-mail: andrea.brancaccio@icrm.cnr.it  
S. Covaceuszach, Istituto di Cristallografia, CNR, Basovizza, SS14, km 163,5 Trieste 34012 Trieste, Italy  
Tel: +39 (040) 375-7520  
E-mail: sonia.covaceuszach@ts.ic.cnr.it

<sup>†</sup>These authors contributed equally to this work.

(Received 8 June 2017, accepted 9 June 2017)

doi:10.1002/2211-5463.12259

Dystroglycan (DG), composed of  $\alpha$  and  $\beta$  subunits, belongs to the dystrophin-associated glycoprotein complex.  $\alpha$ -DG is an extracellular matrix protein that undergoes a complex post-translational glycosylation process. The bifunctional glycosyltransferase like-acetylglucosaminyltransferase (LARGE) plays a crucial role in the maturation of  $\alpha$ -DG, enabling its binding to laminin. We have already structurally analyzed the N-terminal region of murine  $\alpha$ -DG ( $\alpha$ -DG-Nt) and of a pathological single point mutant that may affect recognition of LARGE, although the structural features of the potential interaction between LARGE and DG remain elusive. We now report on the crystal structure of the wild-type human  $\alpha$ -DG-Nt that has allowed us to assess the reliability of our murine crystallographic structure as a  $\alpha$ -DG-Nt general model. Moreover, we address for the first time both structures in solution. Interestingly, small-angle X-ray scattering (SAXS) reveals the existence of two main protein conformations ensembles. The predominant species is reminiscent of the crystal structure, while the less populated one assumes a more extended fold. A comparative analysis of the human and murine  $\alpha$ -DG-Nt solution structures reveals that the two proteins share a common interdomain flexibility and population distribution of the two conformers. This is confirmed by the very similar stability displayed by the two orthologs as assessed by biochemical and biophysical experiments. These results highlight the need to take into account the molecular plasticity of  $\alpha$ -DG-Nt in solution, as it can play an important role in the functional interactions with other binding partners.

Dystroglycan (DG) is a heterodimeric transmembrane glycoprotein that is a part of the multimeric dystrophin–glycoprotein complex and plays a crucial role in the

association of cells with the basement membranes [1]. DG links the basal lamina with the cytoskeleton by bridging the intracellular dystrophin to a plethora of

### Abbreviations

DG, dystroglycan;  $D_{max}$ , maximum size of the particle; DSF, differential scanning fluorimetry; ECM, extracellular matrix; EOM, ensemble optimization method; h $\alpha$ -DG-Nt, human  $\alpha$ -DG-Nt; Ig-like, immunoglobulin-like; LARGE, like-acetylglucosaminyltransferase; MMexp, experimental molecular mass of the solute; m $\alpha$ -DG-Nt, murine  $\alpha$ -DG-Nt; NSD, normalized spatial discrepancy;  $R_g$ , radius of gyration; rmsd, root mean square deviation; S6 domain, small subunit ribosomal protein S6 of *Thermus thermophilus*; SAXS, small-angle X-ray scattering;  $T_m$ , melting temperatures;  $V_p$ , excluded volume of the hydrated particle;  $\alpha$ -DG-Nt, N-terminal region of  $\alpha$ -DG;  $\alpha$ -DG,  $\alpha$ -dystroglycan;  $\beta$ -DG,  $\beta$ -dystroglycan.

extracellular matrix (ECM) proteins, that is, laminin, agrin, and perlecan, thus offering stability to tissues. DG is highly expressed in skeletal muscle where it was first discovered [2] and where it confers structural stability to the sarcolemma during contraction, but it is also strongly expressed in heart, in brain, and in peripheral nerves, where it is involved in various physiological processes [3]. Moreover, DG has been also associated with Old World arenaviral infections, acting as a receptor for virus anchoring [4].

Dystroglycan is encoded by a single gene (*DAG1*) [2] and the corresponding precursor is proteolytically cleaved within the endoplasmic reticulum, resulting in the formation of the extracellular  $\alpha$ -dystroglycan ( $\alpha$ -DG) and the transmembrane  $\beta$ -dystroglycan ( $\beta$ -DG). In their mature forms,  $\alpha$ - and  $\beta$ -DG are linked together through noncovalent interactions involving the N-terminal and the C-terminal regions of  $\beta$ -DG and  $\alpha$ -DG, respectively.  $\alpha$ -DG undergoes a complex, and is still not fully understood, glycosylation post-translational process, which involves several enzymes at various stages of  $\alpha$ -DG maturation [5] in both endoplasmic reticulum and Golgi apparatus. A correct  $\alpha$ -DG glycosylation has been shown to be critical for its physiological functions. In a family of neuromuscular diseases called secondary dystroglycanopathies, the hypoglycosylated forms of  $\alpha$ -DG, resulting from defective enzymes responsible for  $\alpha$ -DG glycosylation, display limited binding capabilities toward laminin with severe implications for health [6]. A low degree of  $\alpha$ -DG glycosylation has also been found in rare diseases caused by single point mutations hitting the DG gene [7–9]. In recent years, it has been discovered that the  $\text{Ca}^{2+}$ -dependent interaction of  $\alpha$ -DG with its main binding partner in ECM, the LG domain, is specifically mediated by a novel polysaccharide [10]. The  $\alpha$ -DG glycosylation biosynthetic pathways involve a kinase and several glycosyltransferases [5], among which the bifunctional glycosyltransferase like-acetylglucosaminyltransferase (LARGE) that adds the repeating disaccharide unit ( $-\alpha3\text{-GlcA-}\beta3\text{-Xyl-}$ ) to a glycan anchored at the site defined by Thr317 and Thr319 [11]. Indeed, it has been reported that the elongation of the glycan operated by LARGE requires the presence of the N-terminal region of  $\alpha$ -DG ( $\alpha$ -DG-Nt), which is thought to act as a recognition site for LARGE before being processed by a furin-like proprotein convertase [12,13]. It has been proposed that  $\alpha$ -DG-Nt would be able to bind other partners in the ECM [14], but the biological implications of these potential interactions remain elusive [15].

We aimed at characterizing the biophysical and biochemical bases behind the biological function of  $\alpha$ -DG, and our efforts have been focused on the comprehension

of the molecular determinants that modulate its binding abilities [16]. An electron microscopy study showed that  $\alpha$ -DG assumes a dumbbell-like shape, with two globular N-terminal and C-terminal regions at the extremes of a mucin-like region [17]. Furthermore, the crystal structure of the murine N-terminal region disclosed a modular architecture composed by an immunoglobulin-like (Ig-like) domain and a second domain similar to the small subunit ribosomal protein S6 of *Thermus thermophilus* (S6 domain) [18]. Although the experimental structure of the C-terminal region of  $\alpha$ -DG is still unknown, homology modeling suggests that its fold is likely to be a second Ig-like structure [19]. According to the crystallographic structure, murine  $\alpha$ -DG-Nt ( $m\alpha$ -DG-Nt) assumes an overall rather compact fold, with the Ig-like and S6 domains interacting with each other and linked together by a flexible loop [18]. Such an organization has been observed in the crystal structure of the murine  $\alpha$ -DG missense pathological mutant T190M, which displays the same fold, mutual orientation, and interaction between the Ig-like and S6 domains observed for WT  $\alpha$ -DG-Nt [20]. Despite such high degree of structural similarity, it has been proposed that T190M might have a reduced ability to assist LARGE in its glycan elongation action, and the resulting  $\alpha$ -DG hypoglycosylation leads to a form of limb-girdle muscular dystrophy [7]. The comparison of the WT and T190M crystallographic models ruled out any effect of the T190M mutation on the Ig-like and S6 domains folding, as earlier proposed by a computational study [21]. Therefore, we considered whether the murine crystallographic structure could be reliably used as a general structural model for  $\alpha$ -DG-Nt. In addition, we have explored  $\alpha$ -DG-Nt propensity for plasticity in solution by small-angle X-ray scattering (SAXS) analysis of both murine and human proteins. The high-resolution crystal structure of human  $\alpha$ -DG-Nt ( $h\alpha$ -DG-Nt) has also been determined, in order to assess the structural similarities between the two orthologs, which are 93% identical in their amino acid sequence. The combination of the high- and low-resolution structural data (respectively, in the crystals and in solution) with biochemical and biophysical experiments proves that the murine and human proteins share a highly conserved overall architecture as well as a striking parallel structural flexibility in solution.

## Results

### Conformational stability of the N-terminal domains of murine and human $\alpha$ -dystroglycan

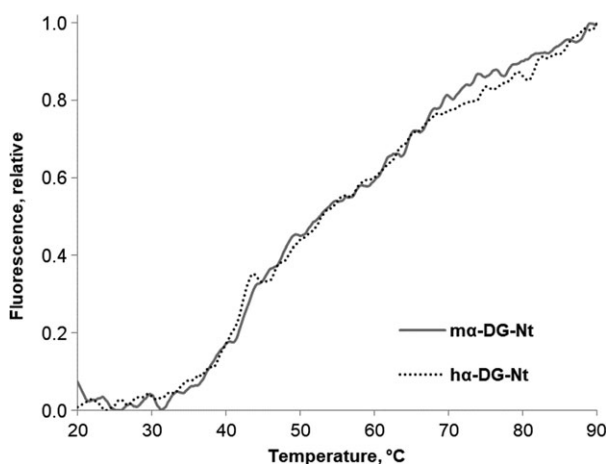
Conformational stabilities of murine and human  $\alpha$ -DG-Nt were compared by biochemical and

biophysical experiments, that is, differential scanning fluorimetry (DSF) and limited proteolysis assays.

In the DSF assay, thermally induced protein denaturation is monitored by the increase in SYPRO Orange fluorescence upon exposure of hydrophobic patches during protein unfolding. The comparison of the resulting thermal unfolding profiles and melting temperatures ( $T_m$ ) can be used to infer differences in the conformation and therefore in the thermal stability of the proteins [22].

Figure 1 shows the changes in fluorescence emission of the murine and human proteins upon thermal unfolding in the presence of the dye.

Both the murine and human variants showed very low background fluorescence in the pretransition region that is quite flat. Their denaturation curves are almost superimposable and suggest a two-transition unfolding process typical of protein containing two independently folded domains. The  $T_m$  values for the first transition obtained from the Boltzmann sigmoid fitting of the data ( $43.76 \pm 0.08$  °C and  $43.94 \pm 0.11$  °C for murine and human variants, respectively) did not show any significant difference, supporting the hypothesis that the two proteins share a very similar conformational stability in solution. A similar thermal stability of the two proteins is also clear from the inspection of the second transition, although in this case the  $T_m$  could not be calculated because it was not possible to reach the aggregation region even increasing the final temperature to the upper range value of the instrument.



**Fig. 1.** Thermal denaturation assay by DSF. Comparison of the thermal unfolding curves for murine and human  $\alpha$ -DG-Nt. The increase in fluorescence emission at 516 nm indicates the association of SYPRO Orange with exposed hydrophobic residues as the protein unfolds. Experiments were performed in triplicate: A single representative curve is shown for clarity.

Limited proteolysis was also used to probe conformational stability, assuming that proteolytic recognition sites become accessible upon unfolding. Indeed, this technique is widely used to examine flexible and exposed regions, considering that proteolysis occurs exclusively at ‘hinges and fringes’ [23] and conformational parameters such as accessibility and segmental mobility correlate quite well with exposed proteolytic sites [24]. Limited proteolysis experiments (data not shown) with a panel of seven different proteases did not reveal any significant difference in conformational stability or flexibility between  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt.

### Crystallographic structure of human $\alpha$ -dystroglycan N-terminal domain

The crystallographic structure of  $h\alpha$ -DG-Nt has been determined at a resolution of 1.8 Å. Upon completion of the crystallographic refinement, the final  $R$ -factor was 0.163 ( $R_{\text{free}} = 0.195$ ), with residues 52–60, 163–179, and 305–315 missing in the final model as no reliable electron density could be detected for these regions. A lower-quality electron density was also observed for the flexible regions encompassing residues 89–91 and 181–185. The region comprising residues 159–162 shows signs of multiple conformations, but any attempt to model it during the refinement did not improve the  $2F_o - F_c$  and  $F_o - F_c$  density maps, nor the refinement quality indicators.

The Ig-like domain (residues 62–160) and the S6 domain (residues 182–305) assume the same relative orientation as observed in  $m\alpha$ -DG-Nt [18] (Fig. S1A, B) with a root mean square deviation (rmsd) between the murine and  $h\alpha$ -DG-Nt crystallographic models equal to 0.468 Å (calculated on 225 common  $C_{\alpha}$ s).

Differences between the primary structures of human and murine  $\alpha$ -DG (see Table 1 for alignment) were easily identified in  $h\alpha$ -DG-Nt, according to the  $2F_o - F_c$  and  $F_o - F_c$  density maps (see Fig. S2 for selected examples).

The differences in the primary structures are all mapped on the protein surface of  $h\alpha$ -DG-Nt (Fig. 2 and Fig. S3), with residues fully or partially exposed to the bulk solvent.

The residues that are different in  $h\alpha$ -DG-Nt and  $m\alpha$ -DG-Nt are clustered in four small patches (Table 2) that are longitudinally distributed along one edge of the proteins (Fig. 2). Such an uneven distribution in  $\alpha$ -DG-Nt seems to be an intrinsic property of this protein region (Fig. S4). Patches P1 and P2 are located on the Ig-like domain, whereas patches P3 and P4 are on the S6 domain (Fig. S3). Patches P2 and P3 face each other in a large cleft lined by the  $\beta$ -strands B, D, and E of the

**Table 1.** Sequence alignment.

<i>Homo sapiens</i>		$\beta A'$	$\beta B$	$\beta C$	$\beta D$	
<i>Mus musculus</i>						
<i>Homo sapiens</i>	52	SVLSDLHEAVPTVVGIPDGTAVVGRSFRVTIPTDLIASSGDIKVSAGKEALPSWLHWD	111			
<i>Mus musculus</i>	50	SVLSDFQEAVPTVVGIPDGTAVVGRSFRVSIPTDLIASSGEIIKVSAGKEALPSWLHWD	109			
		*****:*****:*****:*****:*****				
<i>Homo sapiens</i>		$\beta E$	$\beta F$	$\beta G$		
<i>Mus musculus</i>						
<i>Homo sapiens</i>	112	SQSHLEGLPLDTDKGVHYISVSATRLGANGSHIPQTSSVFSIEVYPEDHSELQSVRTAS	171			
<i>Mus musculus</i>	110	PHSHILEGLPLDTDKGVHYISVAARLGANGSHVPQTSSVFSIEVYPEDHNEPQSVRAAS	169			
		.: ** *****:*****:*****:*****: * ** *: **				
<i>Homo sapiens</i>		$\beta 1$	$\beta 2$			
<i>Mus musculus</i>						
<i>Homo sapiens</i>	172	PDPGEVVSSACAADEPVTVLTVILDADLTGMTPKQRIDLLHRMRSFSEVELHNMKLVFVV	231			
<i>Mus musculus</i>	170	SDPGEVVPSACAADEPVTVLTVILDADLTGMTPKQRIDLLNRMQSFSEVELHNMKLVFVV	229			
		. ***** . *****:*****:*****:*****: ** : *****				
<i>Homo sapiens</i>		$\beta 3$	$\beta 4$	$\alpha$	$\alpha$	$\alpha$
<i>Mus musculus</i>				$\alpha$	$\alpha$	$\alpha$
<i>Homo sapiens</i>	232	NNRLEFDMSAFMAGPGNAKKVVENGALLSWKLGCSLNQNSVPDIHGVEAPAREGAMSAQLG	291			
<i>Mus musculus</i>	230	NNRLEFDMSAFMAGPGNAKKVVENGALLSWKLGCSLNQNSVPDIRGVE TPAREGAMSAQLG	289			
		*****:*****:***:*****				
<i>Homo sapiens</i>		$\beta 5$				
<i>Mus musculus</i>						
<i>Homo sapiens</i>	292	YPVVGWHIANKKPLPKRVRRIH	315			
<i>Mus musculus</i>	290	YPVVGWHIANKKPTLPKRLRRIH	313			
		*****.*****:*****				

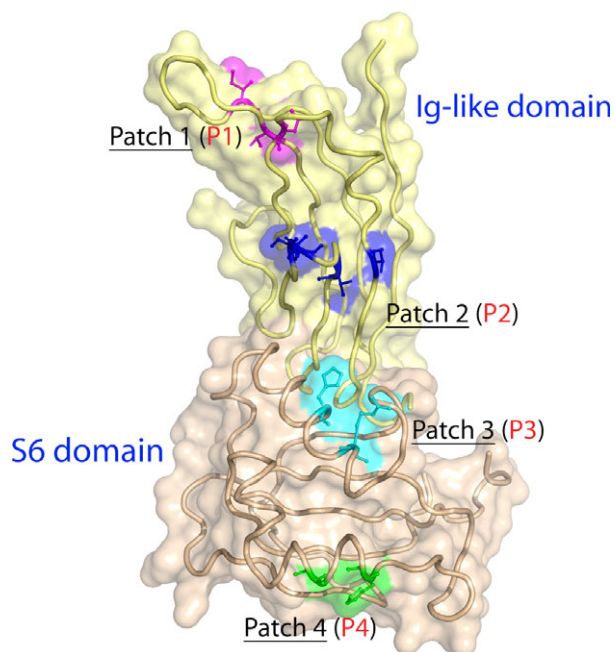
Ig-like domain and by the  $\alpha$ -helices H2 and H3 of the S6 domains (see [18] and Fig. S3B). P2 and P3 patches do not show any explicit mutual interaction.

While the global rmsd is quite low, small but significant deviations between superposed  $C_{\alpha}$ s (around 0.7–1.2 Å against average values of  $\sim 0.1$ –0.3 Å and a maximum-likelihood error estimate of 0.233 Å) are observed in the zones encompassing residues 111–114, 134–145, and 155–162. The zone defined by residues 134–145, which includes N- and C-terminal regions of  $\beta$ -strands F and G and the turn connecting them [18], does not show any remarkable structural variation and the  $C_{\alpha}$ s deviations above the average are probably due to the intrinsic dynamics of the turn connecting the two strands. Apart from the flexible linker (residues 161–181) connecting the Ig-like and the S6 domains, the zone encompassing residues 111–114 shows the highest deviations between superposed  $C_{\alpha}$ s (Fig. S2A), which is probably due to the very different nature of the residues occupying the same topological position in the two proteins (Pro110 in  $m\alpha$ -DG-Nt and Ser112 in  $h\alpha$ -DG-Nt). It is well known that prolines reduce the

conformational freedom of the proteins backbone [25], and it is likely that its substitution in the corresponding position of  $h\alpha$ -DG-Nt with Ser112 may affect the local main chain conformation. Indeed, in  $h\alpha$ -DG-Nt, residues 112–114 assume a helix- $3_{10}$  conformation, instead of the turn observed in  $m\alpha$ -DG-Nt. We do not observe any relevant backbone deviation between  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt for the residues being part of patches P3 and P4. Nonetheless, the interacting network of residues spatially closer to His212 and Arg215 (patch P3) is different from that observed in the corresponding region of  $m\alpha$ -DG-Nt (Fig. 3A–B). This discrepancy is likely due to the different chemical nature of the positively charged His and Arg amino acids ( $h\alpha$ -DG-Nt) with respect to the polar Asn and Gln residues ( $m\alpha$ -DG-Nt). This notion is further supported by the comparison of the electrostatic potentials of  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt, which are locally different around residues 212 and 215 (Fig. 3A–B), while overall being quite similar (Fig. S1A–B).

The  $h\alpha$ -DG-Nt and  $m\alpha$ -DG-Nt crystal structures display the most significant structural differences in





**Fig. 2.** P1–P4 mapping on  $h\alpha$ -DG-Nt structure and accessible surface. The  $h\alpha$ -DG-Nt model is depicted as ribbon representation, with the Ig-like and S6 domains colored in pale yellow and in pale pink, respectively. Residues belonging to the patches P1–P4 are depicted as a stick-and-ball model, and their solvent-accessible surfaces are mapped on the  $h\alpha$ -DG-Nt surface. Residues and surfaces belonging to different patches are colored differently.

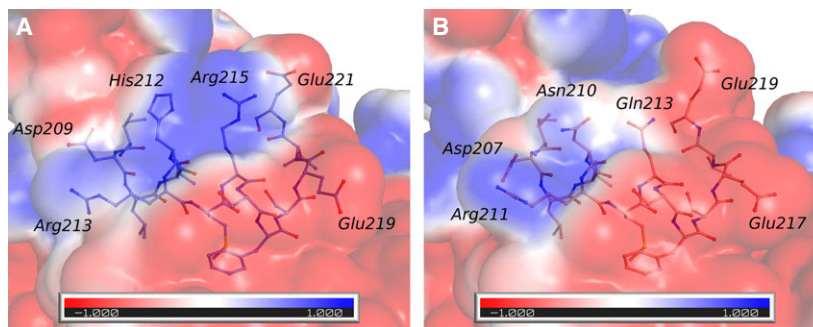
the flexible linker (see Fig. S2A,B) connecting the Ig-like and S6 domains. According to the  $h\alpha$ -DG-Nt refined model, the conformation of the N-terminal part of the linker, the only one reliably modeled in both structures, differs from that observed in  $m\alpha$ -DG-Nt crystal structure. This finding is not surprising, being the linker very flexible [20], probably playing a pivotal role in  $\alpha$ -DG structural plasticity as discussed in the next paragraphs. As mentioned above, the linker is also one of the  $\alpha$ -DG zones with the highest sequence variability among different species (see [26] and Table S1). Besides, the presence of Leu164 in  $h\alpha$ -DG-Nt (not modeled in the  $h\alpha$ -DG-Nt crystallographic structure) instead of Pro162 in the corresponding position of  $m\alpha$ -DG-Nt may influence  $h\alpha$ -DG-Nt conformational variability with respect to  $m\alpha$ -DG-Nt. Indeed, residues 159–161 assume a helix- $3_{10}$  conformation in  $h\alpha$ -DG-Nt, whereas they display a turn/coil conformation in  $m\alpha$ -DG-Nt.

### Association state and overall parameters of the N-terminal domains of murine and human $\alpha$ -dystroglycan in solution

Small-angle X-ray scattering experiments were performed to compare the conformations in solution of  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt.

**Table 2.** Amino acid differences between  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt.

Patch 1 (P1)			Patch 2 (P2)			Patch 3 (P3)			Patch 4 (P4)		
Res	$m\alpha$ -DG-Nt	$h\alpha$ -DG-Nt	Res	$m\alpha$ -DG-Nt	$h\alpha$ -DG-Nt	Res	$m\alpha$ -DG-Nt	$h\alpha$ -DG-Nt	Res	$m\alpha$ -DG-Nt	$h\alpha$ -DG-Nt
92	Asp	Glu	81	Ser	Thr	212	Asn	His	275	Arg	His
136	Ala	Thr	112	Pro	Ser	215	Gln	Arg	279	Thr	Ala
145	Val	Ile	113	His	Gln						
			116	Ile	Thr						

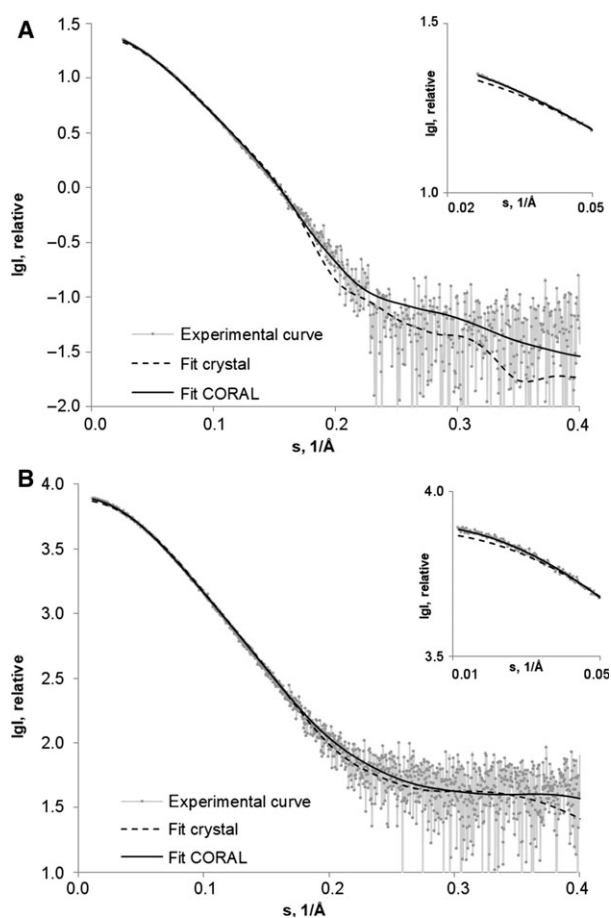


**Fig. 3.** Electrostatic potential maps of  $m\alpha$ -DG-Nt and  $h\alpha$ -DG-Nt. The electrostatic potential (kBT/ec) is mapped on the human and murine  $\alpha$ -DG-Nt-accessible surfaces. Molecular models are represented as ribbons with selected residues depicted as stick-and-ball. (A)  $h\alpha$ -DG-Nt, 208–221 stretch. (B)  $m\alpha$ -DG-Nt, 206–219 stretch.

**Table 3.** Overall parameters calculated from SAXS data analysis.

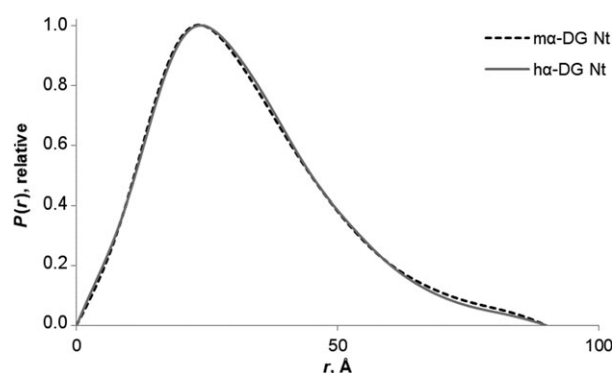
Samples	Concentration (mg·mL <sup>-1</sup> )	$R_g$ (Å)	$D_{max}$ (Å)	$V_p$ (Å <sup>3</sup> )	MM <sub>exp</sub> (kDa)	$\chi_{crystal}$	$\chi_{CORAL}$	$\chi_{ab\ initio}$	$\chi_{EOM}$
m $\alpha$ -DG-Nt	0.22–3.38	25.2 ± 0.04	90 ± 3	44 000 ± 2000	25.1 ± 3.0	2.87	0.84	0.81	0.63
h $\alpha$ -DG-Nt	0.24–4.2	25.1 ± 0.04	90 ± 3	44 000 ± 2000	28.8 ± 3.0	2.25	1.17	1.14	1.06

Notations: MM<sub>exp</sub>, experimental molecular mass of the solute;  $\chi_{crystal}$ ,  $\chi_{CORAL}$ ,  $\chi_{ab\ initio}$ , and  $\chi_{EOM}$ , discrepancy (chi-square value) for the fit from the crystallographic structures with the missing regions reconstructed by CORAL keeping fixed the two domains, from rigid body modeling using CORAL, from *ab initio* modeling using DAMMIN, and from EOM, respectively.



**Fig. 4.** Experimental X-ray scattering data and obtained fits for (A) m $\alpha$ -DG-Nt and (B) h $\alpha$ -DG-Nt. Experimental SAXS patterns, scattering calculated from the crystallographic models ('fit crystal', continuous line), scattering calculated from rigid body models obtained by CORAL ('fit CORAL', dashed line). The plots display the logarithm of the scattering intensity as a function of momentum transfer. The zoomed regions of these graphs at low angles are presented in the insert.

The protein solutions were analyzed at different concentrations (Table 3), and in both cases, no systematic changes with the solute concentration could be observed, although the murine protein showed a certain degree of aggregation, likely due to the higher concentration of the stock solution.



**Fig. 5.** Distance distribution functions for m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt. Overlaid of the  $p(r)$  distance distribution functions calculated from the experimental SAXS data of m $\alpha$ -DG-Nt (black dotted line) and h $\alpha$ -DG-Nt (gray solid line).

The experimental SAXS curves, obtained at the highest concentration of m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt, are displayed in Fig. 4A,B, respectively, and the computed  $p(r)$  distance distribution functions are displayed in Fig. 5; the overall parameters extracted from the SAXS data are summarized in Table 3 (additional details on SAXS structural parameters are reported in Table S2).

The molecular mass (MM) of the proteins, estimated from the relative forward scattering intensities ( $s = 0$ , with  $s$  the scattering vector), suggests that both proteins are monomeric in solution at all conditions tested and is in good agreement with the value estimated from the primary sequences (around 28.5 kDa). This is further corroborated by excluded volume of the hydrated protein molecules ( $V_p$ ), consistent with the empirical finding for globular proteins that the hydrated volume expressed in nm<sup>3</sup> should numerically be about twice the MM in kDa. The experimental radius of gyration ( $R_g$ ) and maximum size ( $D_{max}$ ; Table 3) point to an elongated shape of the proteins, and the two  $p(r)$  functions that nicely overlap (Fig. 5) display an asymmetric tail, typical of elongated particles.

It is interesting to note that in both cases, the scattering curves computed by the CRYSOLO program [27]

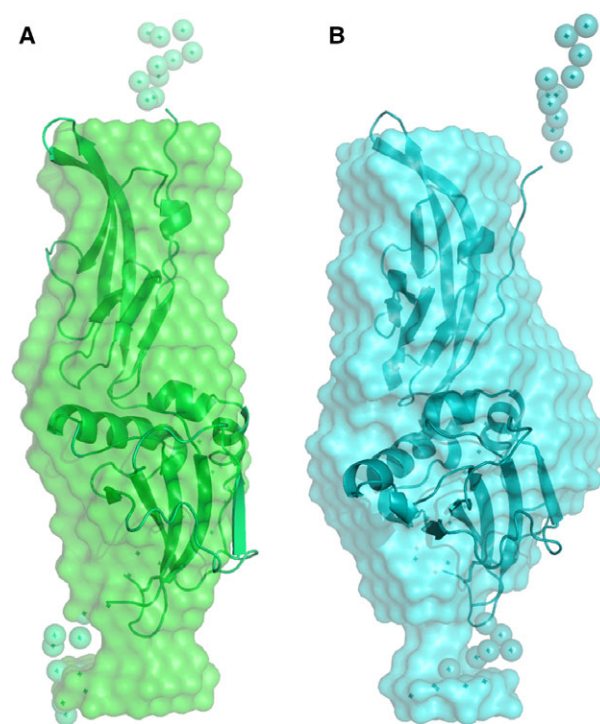
based on the crystallographic models (PDB ID 1U2C and 5LLK for m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt, respectively) give a poor fit to the respective experimental data (data not shown). Even upon reconstruction of the missing regions (around 10 amino acids at both N-terminal and C-terminal regions and the missing linker between the two domains that are kept fixed) using CORAL [28], the fit is not improved ( $\chi_{\text{crystal}}$  in Table 3 and 'fit crystal' in Fig. 4 with the zoomed portions at low angles in the inserts). It can be thus concluded that both murine and human  $\alpha$ -DG-Nt are monomeric in solution, even at relatively high concentrations, but they show a significantly more extended conformation than in the crystallographic models.

### Molecular shape reconstruction of the N-terminal domains of murine and human $\alpha$ -dystroglycan in solution

Two different strategies have been employed to reconstruct the macromolecular shapes of the two proteins in solution.

At first, low-resolution three-dimensional models of m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt were reconstructed from the experimental X-ray scattering data using the *ab initio* modeling program DAMMIN [29] (Fig. 6), with all models providing an excellent fit to the experimental data (Table 3,  $\chi_{\text{ab initio}}$ ). The final DAMMIN models are the result of analyzing and averaging 10 independent solutions. The normalized spatial discrepancy (NSD) value, which describes the similarity between the different models produced by the program [30], is low for both the murine and human models ( $0.478 \pm 0.029$  and  $0.547 \pm 0.014$ , respectively), indicating that the multiple solutions built by the program are very similar to each other. The visible similarity in the shapes of the two models suggests that at low resolution, no differences could be detected.

A second approach to molecular shape reconstruction consisted in a rigid body modeling of the two proteins from the scattering data conducted by using the program CORAL [28]. High-resolution models of individual Ig-like and S6 domains in the corresponding crystal structures were combined with different conformations of flexible dummy residue linkers. The relative orientations of the two domains and the reconstruction of the flexible missing loops linking the Ig-like and the S6 domains were optimized. The CORAL models nicely overlay with the scattering curves ( $\chi_{\text{CORAL}}$  in Table 3 and 'fit CORAL' in Fig. 4 with the zoomed portions at low angles in the inserts) and well superpose into the SAXS envelopes of the *ab initio* models (Fig. 6). It is interesting to notice that, probably due



**Fig. 6.** Structural models of m $\alpha$ -DG-Nt (A) and h $\alpha$ -DG-Nt (B). Averaged and filtered *ab initio* beads models as obtained by DAMMIN (green and cyan semitransparent surfaces) superimposed on the typical CORAL models (green and cyan cartoon representations for the folded Ig-like and S6 domains, spheres for the restored missing fragments).

to packing stabilizing interactions of the crystal lattice, the two crystallographic structures are more compact than the respective conformations in solution, whose more elongated shapes are most evident in the asymmetric tail at the higher  $r$  of their  $p(r)$  distributions (Fig. 5). As a quantitative measure of structure compactness, the distances between centers of masses of the Ig-like and S6 domains in the crystallographic models (29.7 and 29.9 Å in the human and murine models, respectively) have been compared to those of the CORAL models (33.6 and 34.4 Å in the human and murine models, respectively), confirming the existence of more extended conformations in solution. The most straightforward explanation is that the solution structures of the N-terminal regions of murine and human  $\alpha$ -DG display a conformation that is more flexible than the one inferred from their crystal structures.

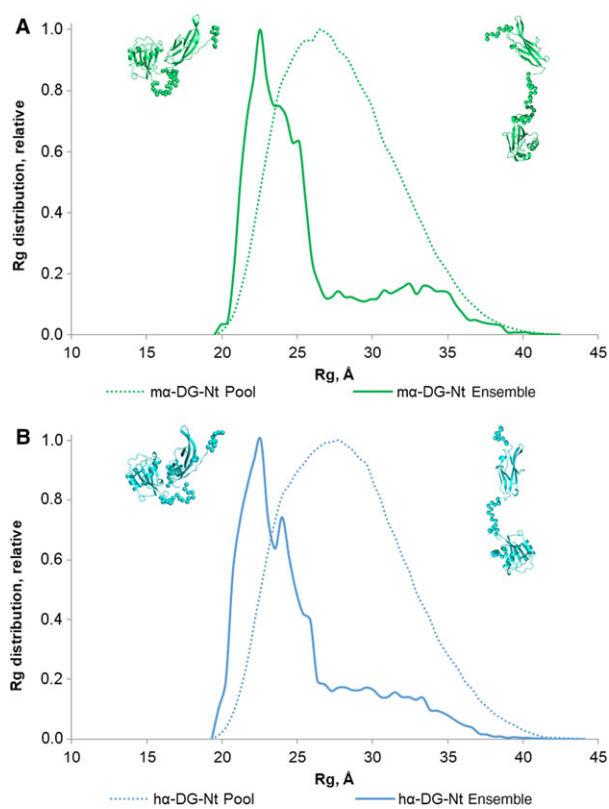
### Interdomain flexibility of the N-terminal domains of murine and human $\alpha$ -dystroglycan in solution

More extended m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt solution structures are in agreement with the evidence of a



disordered region linking the Ig-like and S6 domains, as suggested by the crystallographic analysis. Indeed, linker flexibility could allow variability in the relative orientation of the individual domains, resulting in structural plasticity.

An analysis of the interdomain flexibility and size distribution of possible multiple configurations in solution was conducted by using the ensemble optimization method (EOM) [31], obtaining typical optimized ensembles that fit well the measured scattering data (Table 3). The EOM analyses of the murine and h $\alpha$ -DG-Nt are presented in Fig. 7 as a size distribution, plotting the  $R_g$  of the structures forming the random pool and the selected ensembles against their relative frequencies.



**Fig. 7.**  $R_g$  distributions for the EOM models of m $\alpha$ -DG-Nt (A) and h $\alpha$ -DG-Nt (B). The distributions for the initial random pools of models are shown as green and blue dot lines; green and blue solid lines correspond to the selected ensembles. The representative conformations are shown near the distributions: compact conformation on the left and extended conformation on the right of the respective distributions. Folded domains are depicted as cartoons (m $\alpha$ -DG-Nt in green and h $\alpha$ -DG-Nt in blue); linkers and the N-terminal and C-terminal reconstructed regions are represented by spheres.

The  $R_g$  distributions of these ensembles (Fig. 7, solid lines) are very similar to each other and nearly as broad as the distribution of randomly generated models (Fig. 7, dashed lines), supporting the hypothesis of a certain degree of interdomain flexibility. Moreover, these  $R_g$  distributions are both characterized by a bimodal profile. Indeed, the predominant fractions (around 60%) represent relatively compact models with  $R_g$  of about 21–25 Å, while a very small fraction (around 15%) of models with  $R_g$  of about 30–35 Å is due to more elongated configurations: two molecular structures representative of different conformations are shown in Fig. 7 (compact conformations on the left and extended conformations on the right). This interdomain flexibility may confer structural plasticity, which in turn may represent the molecular basis regulating  $\alpha$ -DG maturation and/or modulating the interactions with other physiological binding partners.

## Discussion

The aim of this study was to investigate the comprehensive structure of  $\alpha$ -DG-Nt in solution. By exploring for the first time the conformational landscape of human and mouse  $\alpha$ -DG-Nts in solution by means of SAXS experiments, we disclosed unexpected shared features of  $\alpha$ -DG-Nt that may help to elucidate the molecular basis of the physiological and pathological functions of  $\alpha$ -DG. Besides, the analysis of the SAXS and crystallographic data of the two orthologs validates previous results and the structure of m $\alpha$ -DG-Nt as a fully descriptive model for h $\alpha$ -DG-Nt. This notion is further supported by the comparison of the conformational stability of the two orthologs as assessed by biochemical and biophysical experiments.

The h $\alpha$ -DG-Nt crystal structure displays the same fold of m $\alpha$ -DG-Nt, with few structural differences. It is interesting to note that when mapping the not conserved amino acids on the high-resolution 3D structure, they cluster around four distinct patches, which span along only one edge of the  $\alpha$ -DG-Nt longest axis (Figs 2 and S4). Such clustering might have a biological significance; that is, these patches might represent ‘hotspots’ for transient species-specific protein–ligand interactions. On the other hand, the conserved surface involving especially the S6 domain (Fig. S1) may be of functional relevance for LARGE recognition along the  $\alpha$ -DG maturation pathway. Moreover, it is interesting to note that the patch involving residues 112 and 113 in h $\alpha$ -DG-Nt just follows Asp111, whose mutation to Asn has been related to pathological  $\alpha$ -DG hypoglycosylation [8], suggesting that this residue is involved in

the complex mechanism leading to the mature, fully glycosylated  $\alpha$ -DG.

In order to explore the molecular structure of the  $\alpha$ -DG-Nt in solution, we have undertaken a SAXS study. SAXS is the technique of election for low-resolution structural studies in solution, especially valuable for flexible systems, whose conformational variability description is precluded to crystallography [32]. While SAXS cannot infer the molecular structure at the atomic level like X-ray crystallography can mostly do, it offers the unique opportunity to obtain, albeit at lower resolution, a structural model in solution, free of the packing forces that may instead influence a crystallographic model [33]. Indeed, while packing effects are not expected to affect the compact folds of the single domains, they could in this case have an impact on the relative orientation of the Ig-like and the S6 domains. Mutual domains orientation may be assisted by the flexible 20-residue-long linker connecting them, which may influence the overall conformation of the  $\alpha$ -DG-Nt. Combined with the crystallographic models of murine and h $\alpha$ -DG-Nt s, SAXS analysis may provide a reliable low-resolution model of  $\alpha$ -DG at near-physiological conditions [34].

The SAXS study presented here points to structural models significantly different from those observed in the murine and h $\alpha$ -DG-Nt crystal structures. According to their solution structures obtained by SAXS data analysis exploiting the respective crystal structures, m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt both assume a remarkably less compact structure in solution than that observed in the crystal structures. The comparison of m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt  $p(r)$  distributions pinpoints to a more extended conformation, a feature that is notably similar for the two orthologs in solution (Fig. 5). Furthermore, the m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt SAXS models nicely fit on the low-resolution molecular envelopes obtained by DAMMIN *ab initio* method. Employing the program CORAL [28], apart from rigid body fitting of the Ig-like and S6 domains, we have been able to reconstruct the missing parts of both m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt. While the N- and C-terminal zones of the Ig-like and S6 domains, respectively, assume extended conformations, the linker connecting the two subdomains displays a more compact structure, rather similar in the two proteins despite important differences in their amino acid sequences. It is interesting to note that by comparing the crystal structures with the CORAL models, the S6 domain appears to be rotated by about 90° to each other around the Ig-like domain. Even if this observation might suggest some functional implications for this assembly, it must be taken into account that the CORAL models have some intrinsic

limitations due to the assumption of the models to be rigid. Indeed to overcome this bias and to gain further information on the conformational variability in m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt in solution, the interdomain flexibility has been assessed by using the EOM method [31]. According to our analysis, murine and h $\alpha$ -DG-Nt share a common behavior in solution, as shown by the similarity in their bimodal  $R_g$  distribution curves, with comparable maxima and slightly different frequencies. Indeed, both m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt in solution are partitioned among few principal populations, which differ in their compactness (Fig. 7). More extended conformations seem to be present, but their relative abundance is low when compared with the most frequent. The common conformational characteristics of murine and h $\alpha$ -DG-Nt suggest that in solution  $\alpha$ -DG may have a more complex behavior than expected only on the basis of its crystal structures.

According to the present SAXS study, the structural plasticity of the  $\alpha$ -DG-Nt seems to be a general property of this protein, as m $\alpha$ -DG-Nt and human  $\alpha$ -DG-Nt show common structural features. Such unexpected conformational variability of the  $\alpha$ -DG-Nt is of great interest and it may play a functional role in its ability to interact with different partners, either inside the cell along its maturation pathway or at the level of the ECM, or in both. It has been proposed that  $\alpha$ -DG-Nt can assist the bifunctional glycosyltransferase LARGE in its complex enzymatic actions, a function that may require  $\alpha$ -DG-Nt to assume different, functionally relevant conformational states. It is also tempting to speculate that the functional flexibility of the  $\alpha$ -DG-Nt was positively selected as it conferred a strong advantage for the multistep maturation pathway. Along this pathway, concerted connections must be established between a plethora of glycosyltransferases and regulatory enzymes that extensively decorate the mucin-like region within the Golgi lumen [35]. Further biochemical and structural work is warranted in order to assess the possibility of direct interactions of  $\alpha$ -DG-Nt with some of these enzymes. Furthermore, the conformational variability in  $\alpha$ -DG highlights the consolidated notion that  $\alpha$ -DG displays a distinct structural modularity, in line with the recent analysis of  $\alpha$ -DG conserved domain organization in metazoan [36]. The autonomous folding modular nature of the entire  $\alpha$ -DG-Nt [26] prompted its possible use as a serum biomarker in DMD patients [37] or also in human uterine fluid to determine uterine receptivity [38].

In line with these results, molecular plasticity of  $\alpha$ -DG in solution should be considered and investigated to enhance our understanding of the molecular basis

of the physiological and pathological role of a central component of the dystrophin–glycoprotein complex.

## Materials and methods

### Cloning, expression, and purification

As previously reported for the murine  $\alpha$ -DG N-terminal region, m $\alpha$ -DG-Nt(50–313)R168H [18], we have cloned the DNA fragment encoding for its human counterpart, h $\alpha$ -DG-Nt(52–315), into the bacterial vector pHis-Trx, for the expression of the protein as a thioredoxin fusion product, with an N-terminal His<sub>6</sub> tag and a thrombin cleavage site. We also introduced within h $\alpha$ -DG-Nt(52–315) the additional mutation R170H, in order to make the protein more proteolytically resistant [22]. The recombinant h $\alpha$ -DG-Nt(52–315) R170H was obtained as previously reported [20]. Namely, it was expressed as fusion protein in *Escherichia coli* BL21 (DE3) Codon Plus RIL, purified by nickel affinity chromatography, cleaved by thrombin, and further purified by anion-exchange and gel filtration chromatography.

### Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry experiments were carried out on a CFX96 Touch Real-time PCR instrument (Bio-Rad, Hercules, CA, USA). Measurements were taken using an excitation wavelength of 470–505 nm and an emission wavelength of 540–700 nm. Data were acquired using a temperature gradient from 20 to 90 °C in 0.2 °C·min<sup>-1</sup> increments. The samples contained 0.5 mg·mL<sup>-1</sup> murine and h $\alpha$ -DG-Nt proteins in 20 mM Tris, 150 mM NaCl pH 7.5, and 90× SYPRO Orange (Sigma-Aldrich, St. Louis, MO, USA) in a total volume of 25  $\mu$ L. The melting curves represent the fluorescence increase arising from the association of SYPRO Orange with exposed hydrophobic residues as the protein unfolds with increasing temperature [33]. Experiments were performed in triplicate. Fluorescence data were analyzed and the  $T_m$ , represented by the inflection points of the transition curves, were calculated using the Boltzmann sigmoid fit [22].

### Limited proteolysis

m $\alpha$ -DG-Nt and h $\alpha$ -DG-Nt were subjected to limited proteolysis at 37 °C at a final concentration of 30  $\mu$ M in 25 mM Tris pH 7.5, 150 mM NaCl buffer. A panel of proteases from Proti-Ace kits (Hampton Research, Aliso Viejo, CA, USA), that is, bromelain, proteinase K, subtilisin, thermolysin, trypsin, endoproteinase Glu-C, and clostripain (endoproteinase Arg-C), were tested at a final concentration of 2  $\mu$ g·mL<sup>-1</sup>. The reactions were stopped after 1, 5, 10, 20, 40, and 60 min by adding SDS sample buffer to aliquots of the reaction mixtures. The samples were analyzed by performing 15% SDS/PAGE and Coomassie staining.

### Crystallization, data collection, structure solution, and refinement

Attempts to grow crystals of h $\alpha$ -DG-Nt by using the hanging-drop vapor diffusion method from conditions similar to those previously reported for both wild-type m $\alpha$ -DG-Nt and its mutant T190M [18,20] were not successful. While exploring new crystal growth conditions by using commercial high-throughput screening kits (100 + 100 nL of protein and precipitant solution at both 277 and 297 K), we also attempted cross-seeding methods starting from already-grown crystals of m $\alpha$ -DG-Nt mutant. While the screenings did not reveal new crystallization conditions, cross-seeding, based on well-established protocols [39,40], resulted in the growth of well-shaped crystals. Precipitant conditions (0.6–1.4 M sodium citrate buffer) and pH (6.8–7.2) were screened by mixing 1  $\mu$ L of protein solution (10 mg·mL<sup>-1</sup> h $\alpha$ -DG-Nt in 150 mM NaCl, 25 mM Tris, pH 7.5) with 1  $\mu$ L of precipitant solution; the drops were equilibrated at 277 K for 3–6 days before seeding. Fully grown crystals were obtained in 2 weeks after seeding in the optimized growth conditions (0.8 M sodium citrate buffer, pH 7.2). Both streak-seeding and microseeding methods were attempted with seeds stock prepared following the manufacturer's protocol (Hampton Research, HR2-320 User Guide). The best-shaped crystals were obtained by using the streak-seeding method. Repeated streak seeding (two to three times), following the same protocol but at optimal precipitant and pH conditions, increased crystal dimensions and improved their diffraction quality.

Data collections were carried out at the XRD1 beamline at ELETTRA (Trieste, Italy) [41,42] using a Pilatus-2M (Dectris Ltd., Baden, Switzerland) detector and the wavelength of 1.00 Å. Data collection was carried out at 100 K after having quickly dipped the crystals into a cryoprotectant solution (25% ethylene glycol added to the precipitant solution) and frozen in liquid nitrogen. Indexing, integration, and data reduction of the diffraction data were carried out by the xds program [43]. Data reduction statistics of the h $\alpha$ -DG-Nt dataset are reported in Table 4.

The structure solution of h $\alpha$ -DG-Nt was obtained by Patterson search methods, using the PHASER software, implemented in the PHENIX crystallographic package [44]. The crystal structure of WT m $\alpha$ -DG-Nt (PDB ID: 1U2C [18]) was used as a search template. Rigid body refinement was initially carried out, followed by a simulated annealing step. The following cycles of the crystallographic refinement included positional refinement and translation-libration-screw (TLS) model parameterization before the individual B-factors refinement. All the refinement cycles were carried out by using PHENIX.REFINE [45] and were alternated with the manual rebuilding of the structure by using the COOT software [46]. Solvent molecules were included in the final model by using the automatic search protocol available in PHENIX.REFINE and manually checked before being included

**Table 4.** X-ray diffraction: data collection and model refinement statistics.

Data collection	
PDB_ID	5LLK
Space group	H3
Unit cell parameters (Å); <i>a</i> , <i>c</i>	71.8, 144.0
Molecules per asymmetric unit	1
Wavelength (Å)	1.000
Resolution (Å)	48.0–1.80 (1.85–1.80) <sup>a</sup>
Total observations	175 979 (12 160)
Unique reflections	25 220 (1766)
$R_{\text{merge}}^b$	0.046 (1.070)
$R_{\text{meas}}^c$	0.049 (1.164)
CC <sub>1/2</sub>	1.000 (0.868)
$\langle I/\sigma(I) \rangle$	21.60 (1.61)
Completeness (%)	99.0 (94.5)
Redundancy	7.0 (6.9)
Refinement	
Resolution (Å)	36–1.80 (1.864–1.80) <sup>a</sup>
Number of reflections (work set/test set)	23 910/2419
$R_{\text{work}}^d$	0.1641 (0.3862)
$R_{\text{free}}^e$	0.1947 (0.3907)
Number of non-H atoms	
Protein	1702
Waters	129
Organic (ethylene glycol)	8
Average isotropic B-factors (Å <sup>2</sup> )	
Protein	44.0
Ligand	53.6
Solvents	44.9
Rmsd from ideality	
Bond length (Å)	0.01
Angle (deg)	1.01
Ramachandran plot	
Favored regions (%)	99.0
Allowed regions (%)	1.0
Disallowed regions (%)	0.0

<sup>a</sup>Values in parentheses are given for the highest resolution shell.

<sup>b</sup> $R_{\text{merge}} = \sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$ .

<sup>c</sup> $R_{\text{meas}} = \sum_{hkl} [n/(n-1)]^{1/2} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$ .

<sup>d</sup> $R_{\text{work}} = \sum_{\text{work set}} |F_{\text{obs}} - F_{\text{calc}}| / \sum_{\text{work set}} F_{\text{obs}}$ .

<sup>e</sup> $R_{\text{free}} = \sum_{\text{test set}} |F_{\text{obs}} - F_{\text{calc}}| / \sum_{\text{test set}} F_{\text{obs}}$ .

in the model. Occupation of residues Asp160, His161, Ser162, Ala184, and Asp185 that display a poor electron density and a high  $B_{\text{iso}}$  was also refined. Protein stereochemistry was monitored throughout the refinement process and during manual rebuilding with MolProbity [47]. Statistics of the crystallographic refinement are reported in Table 4. Coordinates and structure factors have been deposited in the PDB, with accession number 5LLK. Molecular diagrams were created with Pymol [48] and the STRIDE web server [49] has been used for  $\alpha$ -DG-Nt secondary structure assignments. Protein structures superposition and rmsd estimation were carried out by ProFit (Martin, A.C.R., <http://www.bioinf.org.uk/software/profit>).

## Small-angle X-ray scattering (SAXS)

Small-angle X-ray scattering data for  $\alpha$ -DG-Nt were collected on the BM29 beamline [50] at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) as  $10 \times 1$  s exposure time using a Pilatus 1M detector, sample detector distance of 2.87 m and wavelength of 0.99 Å. SAXS measurements for  $\alpha$ -DG-Nt were taken on the P12 beamline EMBL SAXS-WAXS at PETRAIII/DESY (Hamburg, Germany) [51] as  $20 \times 0.05$  s exposures time using a Pilatus 2M detector, sample detector distance of 3.00 m and wavelength of 1.24 Å. Scattering profiles for the collected frames were compared to detect radiation damage.

Measurements were taken at six different concentrations (the ranges are reported in Table 3; additional details on SAXS structural parameters are reported in Table S2) in 20 mM Tris, 150 mM NaCl, pH 7.5.

After normalization to the intensity of the transmitted beam, frames were merged for each sample. Subtraction of the buffer's contribution to the scattering and further processing steps were performed using PRIMUS [52] from the ATSAS 2.6.0 program package [28]. The forward scattering  $I(0)$  and the  $R_g$  were evaluated using the Guinier approximation [53], assuming that at very small angles ( $s < 1.3/R_g$ ), the intensity is represented as:

$$I(s) = I(0) \times e^{-\frac{(s \times R_g)^2}{3}}$$

Pair distance distribution functions of the particles  $p(r)$  and the maximum sizes  $D_{\text{max}}$  were computed using GNOM [54]. MM was estimated by comparison of the calculated forward scattering  $I(0)$  of the samples with that of the standard solution of bovine serum albumin (MM 66 kDa).  $V_p$  was calculated using the Porod approximation [55]:

$$V_p = \frac{2\pi^2 I(0)}{\int I_{\text{exp}}(s) s^2 ds}$$

The program DAMMIN [29] was employed to construct low-resolution *ab initio* beads models of murine and  $\alpha$ -DG-Nt that best fit the scattering data. It employs a simulated annealing procedure to build a compact bead configuration inside a sphere with the diameter  $D_{\text{max}}$  that fits the experimental data  $I_{\text{exp}}(s)$  to minimize the discrepancy:

$$\chi^2 = \frac{1}{(N-1)} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - c I_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2$$

Ten independent DAMMIN runs were performed for each scattering profile in the 'slow' mode, using default parameters and no symmetry assumptions (P1 symmetry). The models resulting from independent runs were superimposed using the program SUPCOMB [30], and aligned models were averaged using DAMAVER [56] to generate a consensus three-dimensional shape.

A simulated annealing protocol implemented in CORAL [28] was employed to find the optimal positions and



orientations of the available high-resolution models of the Ig-like and S6 domains of murine and human  $\alpha$ -DG-Nt. In addition, the program also generated approximate clash-free conformations of the missing portions of polypeptide chain (around 10 amino acids at both the N-terminal and C-terminal regions and the missing linker between the Ig-like and S6 domains). The model fit of the X-ray crystal structures (PDB ID: 1U2C for  $m\alpha$ -DG-NT and 5LLK for  $h\alpha$ -DG-Nt) against the SAXS data was calculated using CRY SOL [27].

Analysis of the interdomain flexibility and size distribution of possible conformers, consistent with the measured scattering data for murine and human  $\alpha$ -DG-Nt, was conducted using the EOM [31]. This method selects an ensemble of possible conformations from a pool of 10 000 randomly generated models constructed from rigid domains linked by randomly generated flexible linkers. The program CRY SOL is used to calculate the theoretical scattering profiles of these models, and a genetic algorithm, GAJOE, is used to select an ensemble of conformations, whose combined scattering profiles best fit the experimental data. The crystal structures of the Ig-like and S6 domains of murine and human  $\alpha$ -DG-Nt were used as rigid bodies for the analysis of the scattering data, employing ensemble optimization. Linkers between the domains and missing N-terminal and C-terminal stretches were represented as a flexible chain of dummy residues.

### Sequence alignment

Murine and human  $\alpha$ -DG sequences were aligned in MUSCLE 3.8 [57] available at EMBL/EBI (<http://www.ebi.ac.uk/Tools/msa>).

### Acknowledgements

We are grateful to Dr D. Lamba (Istituto di Cristallografia, CNR, Trieste, Italy) for the helpful discussions and for critically reading the manuscript. We are grateful to the ELETTRA, to the EMBL Hamburg, and to the ESRF staff for their help and assistance during the data collection. We acknowledge the support of the EMBL EU funding for recording at EMBL Hamburg SAXS beamline. We acknowledge financial support for travel to the ESRF through BAG MX-1640 and MX-1750 for recording at ESRF SAXS beamline. This work was supported by a Wellcome Trust Career Re-entry Fellowship to MGB and by an AFM (French Telethon) Grant (No. 20009) to AB.

### Data accessibility

Structural data are available in PDB database under the accession number 5LLK.

### Author contributions

SC, MB, AB, and AC conceived and designed the experiments. SC, MB, MGB, FS, PVK, and AC performed the experiments. SC, MB, MGB, FS, PVK, AB, and AC analyzed the data. SC, MB, AB, and AC wrote the manuscript. All the authors have approved the manuscript.

### References

- 1 Ervasti JM and Campbell KP (1991) Membrane organization of the dystrophin-glycoprotein complex. *Cell* **66**, 1121–1131.
- 2 Ibraghimov-Beskrovnya O, Ervasti JM, Leveille CJ, Slaughter CA, Sernett SW and Campbell KP (1992) Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* **355**, 696–702.
- 3 Barresi R and Campbell KP (2006) Dystroglycan: from biosynthesis to pathogenesis of human disease. *J Cell Sci* **119**, 199–207.
- 4 Cao W, Henry MD, Borrow P, Yamada H, Elder JH, Ravkov EV, Nichol ST, Compans RW, Campbell KP and Oldstone MBA (1998) Identification of  $\alpha$ -dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science* **282**, 2079–2081.
- 5 Yoshida-Moriguchi T and Campbell KP (2015) Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology* **25**, 702–713.
- 6 Kanagawa M and Toda T (2016)  $\alpha$ -Dystroglycanopathy. In *Translational Research in Muscular Dystrophy* (Takeda S, Miyagoe-Suzuki Y and Mori-Yoshimura M, eds), pp. 21–38. Springer, Japan.
- 7 Hara Y, Balci-Hayta B, Yoshida-Moriguchi T, Kanagawa M, Beltrán-Valero de Bernabé D, Gündesli H, Willer T, Satz JS, Crawford RW, Burden SJ *et al.* (2011) A dystroglycan mutation associated with limb-girdle muscular dystrophy. *N Engl J Med* **364**, 939–946.
- 8 Dong M, Noguchi S, Endo Y, Hayashi YK, Yoshida S, Nonaka I and Nishino I (2015) *DAG1* mutations associated with asymptomatic hyperCKemia and hypoglycosylation of  $\alpha$ -dystroglycan. *Neurology* **84**, 273–279.
- 9 Geis T, Marquard K, Rödl T, Reihle C, Schirmer S, von Kalle T, Bornemann A, Hehr U and Blankenburg M (2013) Homozygous dystroglycan mutation associated with a novel muscle–eye–brain disease-like phenotype with multicystic leucodystrophy. *Neurogenetics* **14**, 205–213.
- 10 Briggs DC, Yoshida-Moriguchi T, Zheng T, Venzke D, Anderson ME, Strazzulli A, Moracci M, Yu L, Hohenester E and Campbell KP (2016) Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat Chem Biol* **12**, 810–814.

- 11 Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, Zhu Z, Burden SJ, Oldstone MBA and Campbell KP (2011) Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proc Natl Acad Sci* **108**, 17426–17431.
- 12 Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MBA *et al.* (2004) Molecular recognition by LARGE is essential for expression of functional dystroglycan. *Cell* **117**, 953–964.
- 13 Heng S, Paule SG, Li Y, Rombauts LJ, Vollenhoven B, Salamonsen LA and Nie G (2015) Posttranslational removal of  $\alpha$ -dystroglycan N terminus by PC5/6 cleavage is important for uterine preparation for embryo implantation in women. *FASEB J* **29**, 4011–4022.
- 14 Hall H, Bozic D, Michel K and Hubbell JA (2003) N-terminal  $\alpha$ -dystroglycan binds to different extracellular matrix molecules expressed in regenerating peripheral nerves in a protein-mediated manner and promotes neurite extension of PC12 cells. *Mol Cell Neurosci* **24**, 1062–1073.
- 15 Sciandra F, Bozzi M, Bigotti MG and Brancaccio A (2013) The multiple affinities of  $\alpha$ -dystroglycan. *Curr Protein Pept Sci* **14**, 626–634.
- 16 Bozzi M, Morlacchi S, Bigotti MG, Sciandra F and Brancaccio A (2009) Functional diversity of dystroglycan. *Matrix Biol* **28**, 179–187.
- 17 Brancaccio A, Schulthess T, Gesemann M and Engel J (1995) Electron microscopic evidence for a mucin-like region in chick muscle  $\alpha$ -dystroglycan. *FEBS Lett* **368**, 139–142.
- 18 Bozic D, Sciandra F, Lamba D and Brancaccio A (2004) The structure of the N-terminal region of murine skeletal muscle alpha-dystroglycan discloses a modular architecture. *J Biol Chem* **279**, 44812–44816.
- 19 De Rosa MC, Pirolli D, Bozzi M, Sciandra F, Giardina B and Brancaccio A (2011) A second Ig-like domain identified in dystroglycan by molecular modelling and dynamics. *J Mol Graph Model* **29**, 1015–1024.
- 20 Bozzi M, Cassetta A, Covaceuszach S, Bigotti MG, Bannister S, Hübner W, Sciandra F, Lamba D and Brancaccio A (2015) The structure of the T190M mutant of murine  $\alpha$ -dystroglycan at high resolution: insight into the molecular basis of a primary dystroglycanopathy. *PLoS ONE* **10**, e0124277.
- 21 Bhattacharya S, Das A, Ghosh S, Dasgupta R and Bagchi A (2014) Hypoglycosylation of dystroglycan due to T192M mutation: a molecular insight behind the fact. *Gene* **537**, 108–114.
- 22 Ericsson UB, Hallberg BM, Detitta GT, Dekker N and Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298.
- 23 Hubbard SJ, Beynon RJ and Thornton JM (1998) Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures. *Protein Eng Des Sel* **11**, 349–359.
- 24 Thornton JM, Edwards MS, Taylor WR and Barlow DJ (1986) Location of “continuous” antigenic determinants in the protruding regions of proteins. *EMBO J* **5**, 409–413.
- 25 MacArthur MW and Thornton JM (1991) Influence of proline residues on protein conformation. *J Mol Biol* **218**, 397–412.
- 26 Brancaccio A, Schulthess T, Gesemann M and Engel J (1997) The N-terminal region of  $\alpha$ -dystroglycan is an autonomous globular domain. *Eur J Biochem* **246**, 166–172.
- 27 Svergun D, Barberato C and Koch MHJ (1995) CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768–773.
- 28 Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV and Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342–350.
- 29 Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879–2886.
- 30 Kozin MB and Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33–41.
- 31 Bernadó P, Mylonas E, Petoukhov MV, Blackledge M and Svergun DI (2007) Structural characterization of flexible proteins using small-angle x-ray scattering. *J Am Chem Soc* **129**, 5656–5664.
- 32 Koch MHJ, Vachette P and Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* **36**, 147–227.
- 33 Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P *et al.* (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**, 429–440.
- 34 Putnam CD, Hammel M, Hura GL and Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191–285.
- 35 Colley KJ (1997) Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology* **7**, 1–13.
- 36 Adams JC and Brancaccio A (2015) The evolution of the dystroglycan complex, a major mediator of muscle integrity. *Biol Open* **4**, 1163–1179.

- 37 Crowe KE, Shao G, Flanigan KM and Martin PT (2016) N-terminal  $\alpha$  dystroglycan ( $\alpha$ DG-N): a potential serum biomarker for Duchenne muscular dystrophy. *J Neuromuscul Dis* **3**, 247–260.
- 38 Heng S, Vollenhoven B, Rombauts LJ and Nie G (2015) A high-throughput assay for the detection of  $\alpha$ -dystroglycan N-terminus in human uterine fluid to determine uterine receptivity. *J Biomol Screen* **21**, 408–413.
- 39 Ducruix A and Giegé R (1992) *Crystallization of Nucleic Acids and Proteins*. IRL Press, Oxford.
- 40 Bergfors TM (2009) *Protein Crystallization*. International University Line, La Jolla, California.
- 41 Bernstorff S, Busetto E, Gramaccioni C, Lausi A, Olivi L, Zanini F, Savoia A, Colapietro M, Portalone G, Camalli M *et al.* (1995) The macromolecular crystallography beamline at ELETTRA. *Rev Sci Instrum* **66**, 1661–1664.
- 42 Lausi A, Polentarutti M, Onesti S, Plaisier JR, Busetto E, Bais G, Barba L, Cassetta A, Campi G, Lamba D *et al.* (2015) Status of the crystallography beamlines at Elettra. *Eur Phys J Plus* **130**, 1–8.
- 43 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
- 44 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
- 45 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH and Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–367.
- 46 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
- 47 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Biol Crystallogr* **66**, 12–21.
- 48 Schrödinger LLC (2010) *The PyMOL molecular graphics system, version 1.3r1*. Schrodinger LLC, Portland, OR.
- 49 Heinig M and Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* **32**, W500–W502.
- 50 Pernot P, Round A, Barrett R, De Maria Antolinos A, Gobbo A, Gordon E, Huet J, Kieffer J, Lentini M, Mattenet M *et al.* (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* **20**, 660–664.
- 51 Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, Franke D, Mark D, Zengerle R, Cipriani F *et al.* (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr* **48**, 431–443.
- 52 Konarev PV, Volkov VV, Sokolova AV, Koch MHJ and Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282.
- 53 Guinier A (1939) La diffraction des rayons X aux tres petits angles : applications a l'etude de phenomenes ultramicroscopiques. *Ann Phys (Paris)* **12**, 161–237.
- 54 Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**, 495–503.
- 55 Porod G (1982) General theory. In *Small Angle X-ray Scattering* (Glatter O and Kratky O, eds) pp. 17–51. Academic Press, London.
- 56 Volkov VV and Svergun DI (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* **36**, 860–864.
- 57 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

## Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Table S1.** Multiple alignment of selected mammalian sequences of the N-terminal region of  $\alpha$ -DG.

**Table S2.** SAXS structural parameters.

**Figure S1.** Electrostatic potential maps of ma-DG-Nt and ha-DG-Nt.

**Figure S2.** Superimposition of the ha-DG-Nt and ma-DG-Nt selected regions.

**Figure S3.** Amino acid differences between mouse and human  $\alpha$ -DG-Nt mapped onto the h $\alpha$ -DG-Nt accessible surface.

**Figure S4.** Amino acid differences mapped onto the h $\alpha$ -DG-Nt accessible surface.