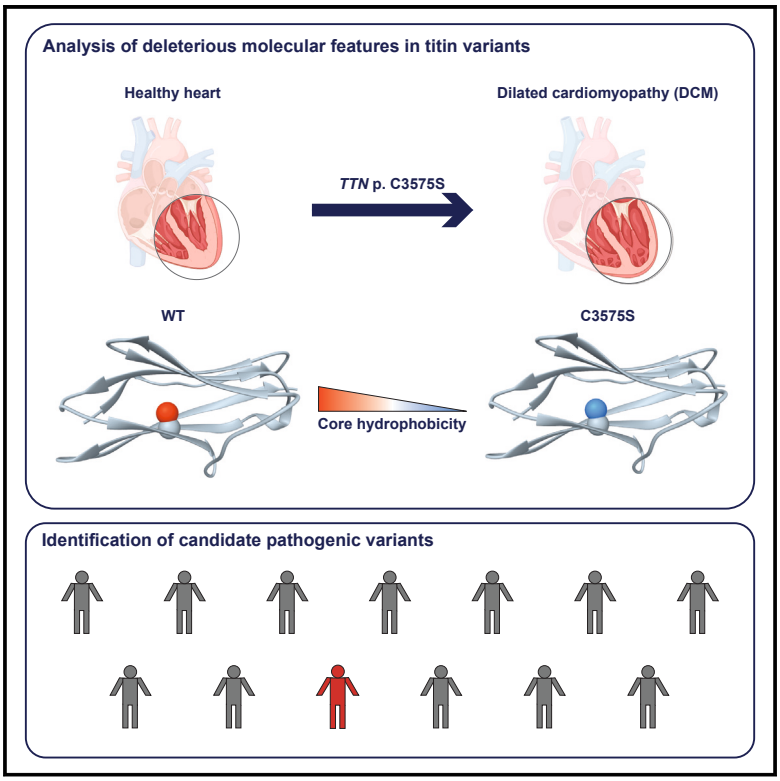
Cell Reports

Article

Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy

Graphical abstract



Highlights

- DCM mutation C3575S destabilizes titin domain I21 without affecting its crystal structure
- Destabilization of I21 C3575S domain is not linked to loss of a disulfide bond
- C3575S severely compromises the stability of the hydrophobic core of the I21 domain
- Variants that reduce core hydrophobicity of titin domains are linked to DCM

Authors

Ines Martinez-Martin, Audrey Crousilles, Juan Pablo Ochoa, ..., David de Sancho, Matthias Wilmanns, Jorge Alegre-Cebollada

Correspondence

ines.martinez@cnic.es (I.M.-M.), jalegre@cnic.es (J.A.-C.)

In brief

Distinguishing pathogenic missense variants from benign polymorphisms is extremely challenging for genes the size of titin. Martinez-Martin et al. have characterized the deleterious physical-chemical features of titin DCM-causing mutation C3575S and used this information to develop a method able to pinpoint pathogenic variants in immunoglobulin domains.

Martinez-Martin et al., 2023, Cell Reports ■■, 113490 ■■, 2023
https://doi.org/10.1016/j.celrep.2023.113490

CellPress

Please cite this article in press as: Martinez-Martin et al., Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy, Cell Reports (2023), https://doi.org/10.1016/j.celrep.2023.113490

Cell Reports

CellPress
OPEN ACCESS

Article

Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy

Ines Martinez-Martin,[1,8,*] Audrey Crousilles,[2] Juan Pablo Ochoa,[1,3,4] Diana Velazquez-Carreras,[1] Simon A. Mortensen,[2] Elias Herrero-Galan,[1] Javier Delgado,[5] Fernando Dominguez,[1,3] Pablo Garcia-Pavia,[1,3] David de Sancho,[6,7] Matthias Wilmanns,[2] and Jorge Alegre-Cebollada[1,*]

[1]Centro Nacional de Investigaciones Cardiovasculares (CNIC), 28029 Madrid, Spain
[2]European Molecular Biology Laboratory (EMBL), Hamburg Unit, 22607 Hamburg, Germany
[3]Heart Failure and Inherited Cardiac Diseases Unit, Department of Cardiology, Hospital Universitario Puerta de Hierro Majadahonda, IDIPHIM, CIBERCV, 28222 Madrid, Spain
[4]Health in Code, 15008 A Coruña, Spain
[5]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain
[6]Polimero eta Material Aurreratuak: Fisika, Kimika eta Teknologia, Kimika Fakultatea, UPV/EHU, 20018 Donostia-San Sebastian, Euskadi, Spain
[7]Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastian, Euskadi, Spain
[8]Lead contact
*Correspondence: ines.martinez@cnic.es (I.M.-M.), jalegre@cnic.es (J.A.-C.)
https://doi.org/10.1016/j.celrep.2023.113490

SUMMARY

The underlying genetic defect in most cases of dilated cardiomyopathy (DCM), a common inherited heart disease, remains unknown. Intriguingly, many patients carry single missense variants of uncertain pathogenicity targeting the giant protein titin, a fundamental sarcomere component. To explore the deleterious potential of these variants, we first solved the wild-type and mutant crystal structures of I21, the titin domain targeted by pathogenic variant p.C3575S. Although both structures are remarkably similar, the reduced hydrophobicity of deeply buried position 3575 strongly destabilizes the mutant domain, a scenario supported by molecular dynamics simulations and by biochemical assays that show no disulfide involving C3575. Prompted by these observations, we have found that thousands of similar hydrophobicity-reducing variants associate specifically with DCM. Hence, our results imply that titin domain destabilization causes DCM, a conceptual framework that not only informs pathogenicity assessment of gene variants but also points to therapeutic strategies counterbalancing protein destabilization.

INTRODUCTION

Dilated cardiomyopathy (DCM) is a cardiac disease characterized by left ventricle dilatation and systolic dysfunction that cannot be explained solely by abnormal loading conditions or coronary artery disease. Common complications include arrhythmias and heart failure,[1] making DCM the most frequent cause of heart transplantation.[2] In the last two decades, clinical management of DCM patients and their families has benefited from identification of causative genetic variants.[3] For instance, non-carrier relatives can be assured they are not at higher risk of developing DCM than the general population, which has a profound impact in their life planning.[4] However, at the present time the genetic origin of DCM cannot be identified in ~60%–70% of families, who therefore cannot take advantage of genetic-based improved care. Instead, genetic analysis of these otherwise genotype-negative individuals typically detects variants for which there is not enough evidence to support pathogenicity, the so-called variants of uncertain significance (VUSs).[5] VUSs have no use in the clinic; hence, there is a pressing need for strategies able to discriminate pathogenic variants from clinically innocuous, benign polymorphisms.

Truncating variants in the titin gene (*TTN*) leading to titin loss of function are a major cause of familial DCM, accounting for approximately 25% of cases.[6,7] Titin is a structural protein that bridges the Z-disk and M-line of sarcomeres in myocytes acting as a molecular spring and scaffold for a variety of protein interactions.[8,9] The elastic I-band part of titin is composed of serially linked immunoglobulin-like (Ig) domains and random coil regions, whereas the rigid A-band contains both Ig and fibronectin type III (FnIII) domains (Figure 1A).[10] Titin is regulated by alternative splicing, which gives rise to isoforms of different length that modulate myocyte stiffness.[11] The long and compliant N2BA and the shorter and more rigid N2B are the two major isoforms in the adult human heart, in which they are present at a 30:70 proportion, respectively.[8]

Remarkably, missense variants in titin are found frequently in DCM individuals.[6,12] However, in the absence of further information, whether these variants are pathogenic or not is impossible to ascertain since benign polymorphisms in titin are also common in the general population.[13] As a consequence, most titin missense variants have remained classified as VUSs.[12,14] Indeed, only three titin missense variants (p.A178D, p.W976R, and p.C3575S) have

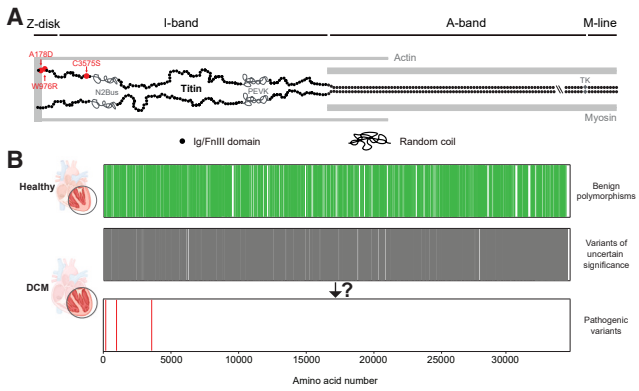


Figure 1. Landscape of titin missense variants in DCM

(A) Representation of one-half a sarcomere (not to scale). N2BA titin molecules are colored in black. Ig and FnIII domains are represented as circles. Positions of PEVK and N2Bus random coil regions are shown. The gap indicated by double backslash corresponds to 100 domains in the A-band that are not represented for simplicity. To date, only three missense variants in titin have been demonstrated to cause DCM (shown in red).

(B) Location of missense variants in the sequence of titin. Benign polymorphisms are represented in green (top), while variants of uncertain significance identified in DCM patients are colored in gray (middle). Pinpointing pathogenic variants among these VUSs has not been possible so far. In red (bottom), location of the three pathogenic variants shown in (A).

been linked to DCM so far based on robust cosegregation with disease[15–19] (Figure 1). It is tempting to hypothesize that the pathogenic features of these DCM-causing variants could also be present in some of the myriad other missense VUSs in titin, potentially enabling their reclassification as pathogenic. In this regard, it is interesting to note that the three aforementioned DCM-causing missense variants destabilize their parent domains.[15,17,19] Hence, we hypothesized that uncovering molecular mechanisms inducing titin domain destabilization could lead to identification of DCM-causing missense variants.

To examine this hypothesis, here we have studied *TTN* p.C3575S (reference Uniprot sequence Q8WZ42-1), the titin missense variant with the strongest association with DCM so far.[19] Using X-ray crystallography, protein biochemistry, molecular simulations, and bioinformatics analyses, we show that the strong destabilization induced by the cysteine to serine substitution results from a substantial decrease in hydrophobicity of the core of the parent I21 domain. Building on these results, we have found that similar variants targeting titin domains are enriched in DCM cohorts, an observation that places these variants as a causative factor of DCM.

RESULTS

Crystal structure of wild-type and C3575S I21 titin domains

We first solved the structure of *E coli*-expressed wild-type (WT) human I21 domain to 1.95-Å resolution (PDB: 8OVU, diffraction data statistics and model refinement parameters are in Data S1). The asymmetric unit of the crystal contains two copies of the domain that show highly similar structures (root-mean-square displacement, RMSD, is 0.50 Å, Figures 2A and S1A). The fold of the I21 domain falls within the intermediate set (type I) of the immunoglobulin superfamily.[20,21] Two antiparallel β sheets formed by four beta strands (A-B-E-D and A′-G-F-C) are packed in a β sandwich. Residue C3575 is located on strand F and is deeply buried in the core of the domain (solvent-accessible surface [SAS] < 0.01 Å^2). In one of the chains, a Mg^{2+} atom is coordinated in octahedral geometry with D3548 and five water molecules (Figure 2B). Mg^{2+} binding is accompanied by minor rearrangement of surrounding side chains, including rotation of D3548 and E3568 (Figure 2C).

To examine whether I21 has the typical titin Ig domain fold, we aligned I21 to 13 available high-resolution structures of Ig domains of human titin covering all relevant titin regions (Table S1). We found that all domains show remarkably high structural similarity. The most dissimilar domain is PDB: 1NCT, located in the M-line (RMSD 1.28 Å, Figure 2D), whereas the I-band domain PDB: 1G1C shows the lowest RMSD with I21 (0.63 Å, Figure 2E). This high structural similarity is not strongly correlated with sequence identity (Figure 2F), as observed in many other members of the Ig superfamily.[22]

Next, we solved the crystal structure of the I21 C3575S mutant domain to 2.2-Å resolution (PDB: 8P35, diffraction data statistics and model refinement parameters are in Data S1). The asymmetric unit of the crystal contains six copies of the domain (RMSD range: 0.25–0.48 Å using chain A as a reference, Figure 3A). Strikingly, despite the highly destabilizing effect of C3575S,[19] the mutation does not induce any noticeable change in the general fold of the domain (RMSD I21 WT vs. I21 C3575S = 0.54 Å, Figure 3B). Consistent with the similar volume of cysteine and serine,[23] the side chains of 3575 and surrounding residues align to a great extent in both structures. There are only slight rotations of I3534 and V3509, which however are unlikely to affect the main interactions of the fold (Figure 3C).

I21 C3575 domain destabilization is not caused by loss of a disulfide bond

The structure of the I21 WT domain shows a second cysteine residue, C3586, at a distance from C3575 that is compatible with disulfide bond formation (Figure 4A). The equivalent positions in many titin Ig domains are typically occupied by cysteines that have been shown to be able to form disulfide bonds.[24,25] The destabilizing effect of the C3575S mutation could stem from the inability of serine to establish a disulfide bond with C3586. However, we found no evidence of disulfide bond formation in the experimental electron density of the WT domain (Figure S1B). To confirm that I21 domain destabilization is not caused by loss of disulfide bond C3575–C3586, we examined the stability of WT and mutant domain preparations whose redox state was evaluated concomitantly using a biochemical NEM-PEGylation assay. In this assay, both domains were incubated with an N-ethylmaleimide-functionalized PEG molecule that binds covalently to reduced cysteines, causing a mobility shift in SDS-PAGE electrophoresis (Figure 4B). As expected, NEM-PEG-treated I21 C3575S domain shows a shift corresponding to the addition of a single PEG molecule (Figure 4C). The WT domain also shows full modification of cysteines, but in this case, the shift in electrophoretic mobility corresponds to two PEG

Please cite this article in press as: Martinez-Martin et al., Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy, Cell Reports (2023), https://doi.org/10.1016/j.celrep.2023.113490

CellPress
OPEN ACCESS

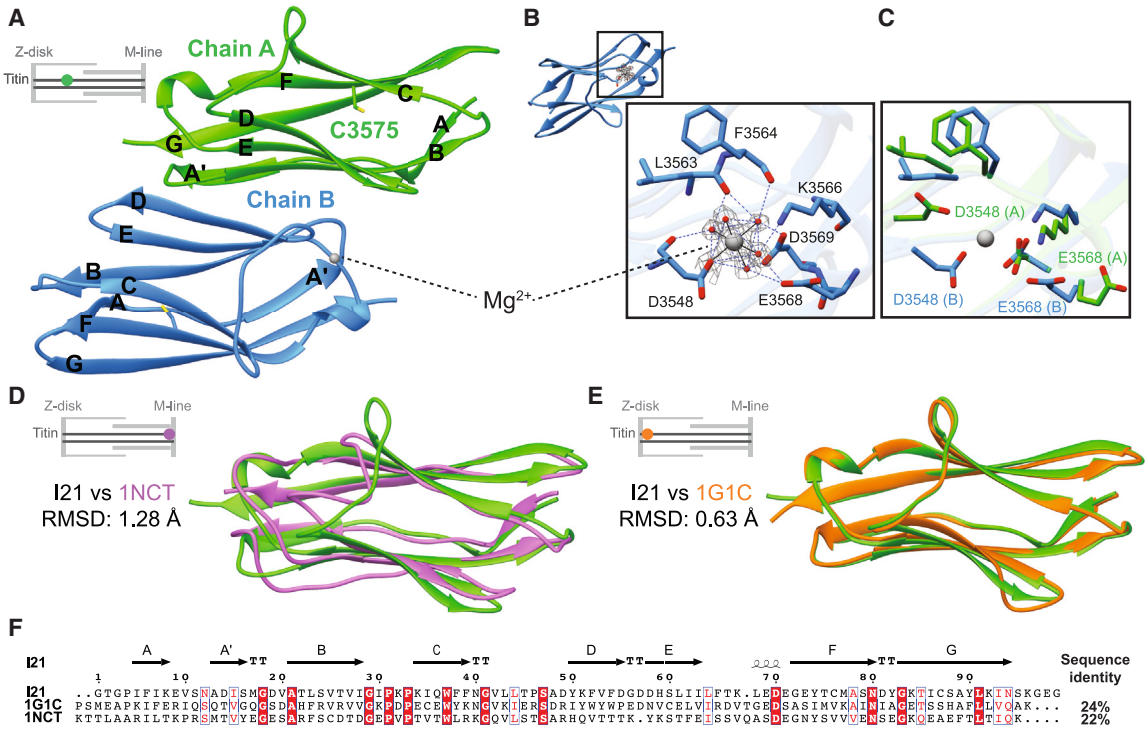


Figure 2. The I21 domain of titin shows a typical Ig fold

(A) Ribbon cartoon of the asymmetric unit of the crystal containing the I21 domain. C3575 is shown as sticks with sulfur atoms in yellow. Chain B (blue) interacts with a Mg^{2+} atom depicted as a gray sphere.

(B) Mg^{2+}-binding site in the crystal structure of I21. Residues and water molecules (red spheres) coordinating the metal are shown (interactions as dashed lines). The electron density surrounding the Mg^{2+} is contoured at 1.2σ (gray mesh).

(C) Structural alignment of the Mg^{2+}-binding site in chains A (green) and B (blue).

(D) Ribbon cartoon of the structural alignment of domains I21 and 1NCT.

(E) Ribbon cartoon of the structural alignment of domains I21 and 1G1C.

(F) Sequence alignment of I21, 1G1C, and 1NCT domains. The secondary structure assignment of I21 is shown on top. Sequence identities between I21 and 1G1C and 1NCT appear on the right. Insets of (A), (D), and (E) indicate the location of the corresponding titin domains in the sarcomere. See also Figure S1 and Table S1.

molecules (Figure 4C). Hence, results of the NEM-PEGylation assay demonstrate that both domain preparations did not contain disulfide bonds nor any other form of cysteine oxidation. Next, using circular dichroism (CD) with these fully reduced protein preparations, we verified that C3575S causes strong destabilization of I21. Specifically, the WT domain has a melting temperature (T_m) of 52°C ± 0.3°C, while the T_m of mutant I21 is 38°C ± 0.5°C (errors are SD from the sigmoidal fit; Figure 4D). Hence, our results confirm that the loss of thermal stability of I21 C3575 is not caused by the loss of a disulfide bond.

C3575S perturbs the I21 equilibrium folding free energy

To characterize the thermodynamic changes induced by the C3575S mutation, we estimated the associated variation in folding free energy (ΔΔG) from molecular dynamics (MD) simulations using an alchemical formalism. This approach, which has been applied to calculate the effects of mutations in protein stability[26] and ligand binding affinity,[27] is based on a thermodynamic cycle that includes both physical (un)folding reactions and unphysical mutation transitions (Figure 5A). According to this cycle, ΔΔG can be obtained from the difference in ΔG of the mutation transitions in the unfolded and folded states ($\Delta G_{m,unfolded}$ and $\Delta G_{m,folded}$, respectively). This approach greatly alleviates the computational expense of the free-energy calculation compared to the simulation of reversible folding and unfolding events for the two forms of the protein. ΔG_m values are calculated from atomistic, explicit-solvent MD trajectories in which the residue of interest is artificially switched from WT to mutant and vice versa (Figures 5A and 5B). Using two different MD force fields, our alchemical calculations indicate that mutation C3575S induces high thermodynamic destabilization (average ΔG = 16.5 ± 0.5 kJ/mol) (Figure 5C), which correlates well with the 14°C decrease in T_m we have measured experimentally.[28]

To look for insights into the molecular mechanisms of destabilization, we analyzed the long equilibrium MD trajectories used to seed the switching simulations. Despite the remarkable destabilization induced by C3575S, we could not detect any substantial effect of the mutation on the alpha carbon RMSD distributions (Figures 5D, S2A, and S2B), on the global conformational mobility of the folded domain (Figure S2C and SD), its radius of gyration (Figures 5E, S2E, and S2F), time evolution of secondary structure (Figures S2G and S2H), or fraction of native contacts (Figures S2I

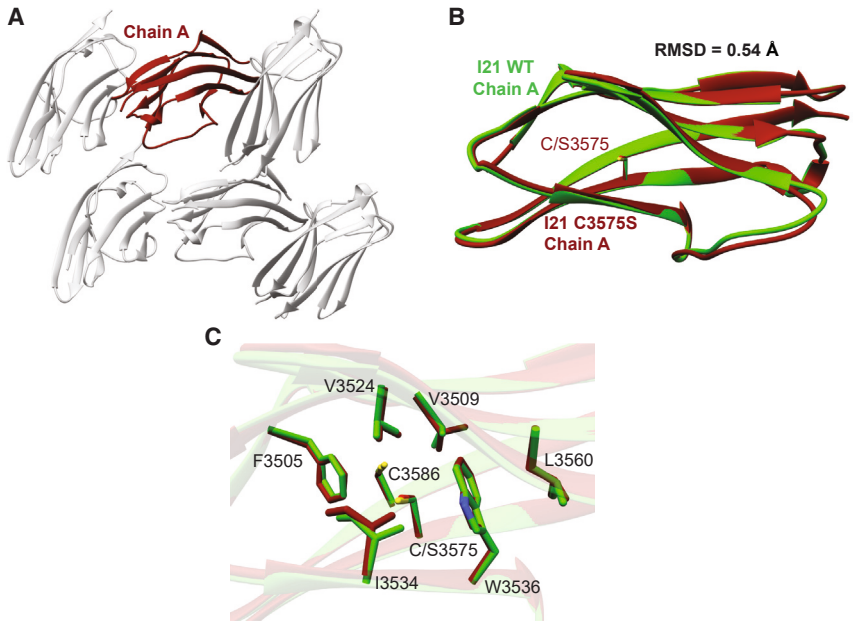


Figure 3. The C3575S mutation does not induce major structural changes in domain I21

(A) Ribbon cartoon of the asymmetric unit of the crystal containing the I21 C3575S domain. Chain A is colored in dark red, and chains B–F in light gray. (B) Structural alignment of I21 WT (green) and C3575S (dark red). The RMSD of the alignment is 0.54 Å. Position 3575 is represented with sticks. (C) Detailed view of the region surrounding position 3575 in the structural alignment of I21 WT (green) and I21 C3575S (dark red).

and S2J). We also observed that the C3575S substitution does not disrupt any of the native hydrogen bonds established by residue 3575 (Figures S3A–S3C). Although residue 3575 remained buried throughout the simulated time for both WT and mutant (Figures S3D and S3E), the serine in the C3575S folded state explored more frequently alternative rotameric states characterized by X_1 angles between $-\pi/2$ and $\pi/2$ (Figures 5F and S3F—S3G). This higher conformational heterogeneity suggests entropic stabilization of the native state of the C3575S domain.[29] Using the Shannon entropy formula, we have estimated that the associated entropic stabilization of the mutant is 0.6–0.8 kJ/mol. Hence, this effect is small and easily compensated by the enthalpic stabilization of the cysteine residue in the core of the folded domain. Indeed, according to statistical inter-residue contact energies for proteins,[30] the cysteine to serine substitution is expected to result in a 4–7 kJ/mol decrease per contact with nearby residues F3505, V3524, I3534, W3536, and C3586 (Figure 3C).

Hydrophobicity loss in DCM-associated titin variants

Our MD simulations confirm the strong thermodynamic destabilization of the C3575S domain, which most likely originates from the loss of hydrophobic contacts induced by the cysteine to serine substitution. Indeed, most hydrophobicity scales, including the commonly used Kyte-Doolittle scale, agree that the hydrophobicity of cysteine is higher than that of serine[31,32] (Figure 6A). Considering also that 3575 is the least solvent-accessible position in I21 (Figure 6B), it appears that C3575 is a fundamental residue to maintain key hydrophobic interactions in the core of the domain.[33] Accordingly, 99% of titin domains contain hydrophobic residues in the position equivalent to 3575 (Figure 6C), while only 55% residues are hydrophobic in the less buried position corresponding to vicinal C3586 (SAS = 10.63 $Å^2$, Figures 6B and 6D).

To investigate if similar hydrophobicity-reducing variants cause DCM as p.C3575S, we examined whether they are specifically found in DCM populations. To this aim, we first calculated the changes in hydrophobicity (ΔHydrophobicity) of thousands of publicly available VUSs in ClinVar potentially associated with DCM that target Ig domains constitutively expressed in the N2B isoform of cardiac titin and, therefore, also present in the longer N2BA isoform. As controls, we used polymorphisms from gnomAD with allelic frequencies that are not compatible with the prevalence of DCM.[34,35] We plotted ΔHydrophobicity values according to the average SAS of the mutated position in titin Ig domains (Figure S4A). As expected, global results do not show noticeable differences in the distribution of DCM-associated VUSs and benign polymorphisms. However, variants causing substantial drops in hydrophobicity (ΔHydrophobicity < −2.2 corresponding to Ala > Gly, the most conservative hydrophobic-to-hydrophilic substitution) and targeting extremely buried positions (SAS < 1.12 $Å^2$) are almost exclusively populated by DCM-associated VUSs (p = 0.008, hypergeometric test) (Figures 6E, S4B, and S4C).

Next, we examined to what extent variants showing those physicochemical features are enriched in DCM patients. To this aim, we performed an independent case-control study using a cohort of 4,440 DCM cases with an unidentified genetic cause following evaluation with a 121-gene next-generation-sequencing DCM panel, and 4,125 control individuals referred for genetic testing due to conditions not related to DCM. Results show that hydrophobicity-reducing titin variants affecting deeply buried residues associate specifically with DCM with an odds ratio of 3.26 (p = 0.04, 95% confidence interval = 1.07–9.9) (Figures 6F–6H; Table S2).

DISCUSSION

The identification of genetic variants causing DCM is revolutionizing the clinical management of patients and their families.[3,5] However, currently only ~30% DCM patients can benefit from improved care, i.e., those carrying well-known pathogenic variants.[36,37] This situation highlights the pressing need to identify DCM-causing variants, which typically relies on genetic screening of very large families in the quest for cosegregating candidates[17,19] or identification of specific enrichment in large

Please cite this article in press as: Martinez-Martin et al., Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy, Cell Reports (2023), https://doi.org/10.1016/j.celrep.2023.113490

CellPress
OPEN ACCESS

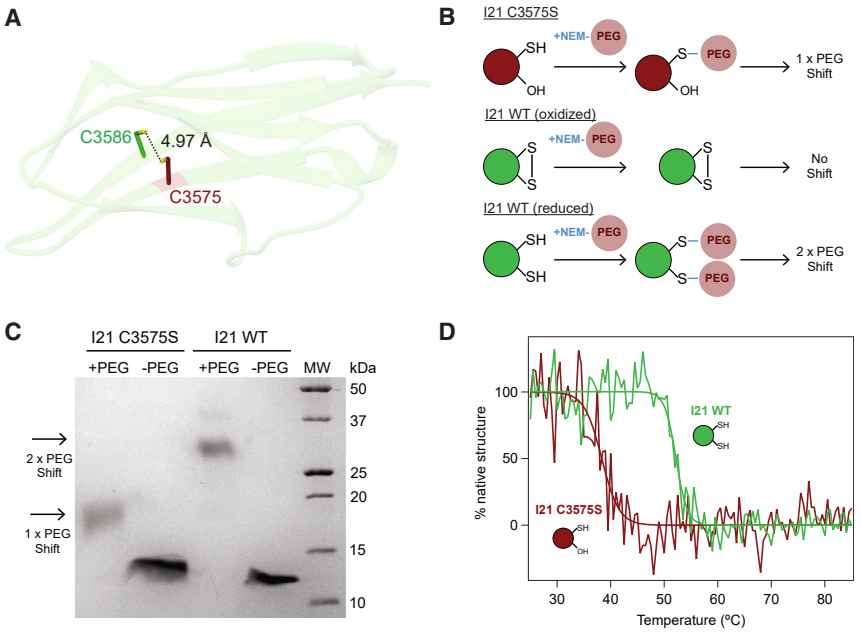


Figure 4. I21 domain destabilization induced by C3575S is not related to the disruption of a disulfide bond

(A) Cartoon representation of I21 WT depicting the two cysteines of the domain as sticks with sulfur atoms in yellow. The dashed line indicates the distance between sulfur atoms (4.97 Å).

(B) Schematic representation of the NEM-PEG assay applied to the I21 C3575S (dark red) and WT (green) domains.

(C) 17% SDS-PAGE showing the results of the NEM-PEG assay. MW, molecular weight markers (Bio-Rad Precision Plus Protein).

(D) Thermal unfolding curves of I21 WT (green) and C3575S (dark red) measured by tracking CD signal at 215 nm. Sigmoidal fits to the data are shown. See also Figure S1.

DCM cohorts.[6] Here, we have followed a third, mixed approach that first identifies pathogenic features in well-established DCM-causing variants, which are then screened for in VUSs found in DCM patients.

Loss-of-function truncations in the titin gene are a common cause of DCM.[6,14] Prompted by this evidence, we and others have speculated that a fraction of titin missense variants could also result in protein loss of function leading to DCM.[12–14] However, to date, only three titin missense variants have been shown to cause DCM (Figure 1). Among these variants, we have studied p.C3575S due to its very high cosegregation with DCM and the fact that it causes strong destabilization of the parent I21 domain.[19] Our results indicate that while mutation C3575S does not affect the structure of I21, it does severely destabilize the hydrophobic core of the domain. We have then screened thousands of VUSs for the presence of this deleterious molecular phenotype and observed that hydrophobicity-reducing variants targeting the three least solvent-accessible positions of titin Ig domains are specifically enriched in DCM cohorts. These three positions together with a strictly conserved tryptophan constitute the hydrophobic core of titin Ig domains and are fundamental for their folding.[22,38] Interestingly, this conserved tryptophan in domain I3 is the target of the pathogenic variant *TTN* p.W976R, which causes ΔHydrophobicity = −5.4.

Our data strongly suggest that the hydrophobicity-reducing nature of variants targeting three core residues of titin domains is an indication of pathogenicity. In Data S2, we provide a list with all such variants affecting the N2B isoform of titin. It is important to consider though that we have also detected this type of potentially damaging variants in a few non-DCM individuals. This is not totally unexpected if we consider that titin-truncating variants, which are typically classified as pathogenic, are also found in 0.5%–1% of the general population, at a much higher frequency than expected from the prevalence of DCM.[6,14,39]

These observations highlight the need to understand downstream pathomechanisms in *TTN*-related DCM.[7,40,41] In the case of the missense variants described here, it is expected that perturbation of the hydrophobic core induces strong thermodynamic destabilization of titin domains. Domain destabilization can result in protein loss of function leading to disease, as it has been observed for cardiac myosin binding protein C, a partner protein of titin in the sarcomere, in the context of hypertrophic cardiomyopathy.[42] However, in the case of a protein the size of titin, how destabilization of a single domain out of hundreds can result in such poisonous consequences is puzzling.[7,16,40] This "Achilles heel" effect could stem from different molecular pathways, like induction of full titin degradation as observed for truncating variants,[7,40] exposure of cryptic cleavage sites altering titin mechanics,[43] or perturbed posttranslational modifications,[7] biomolecular interactions,[18] or sarcomere incorporation.[16] It is important to consider that destabilizing titin missense variants have also been linked to arrhythmogenic cardiomyopathy[44,45] and skeletal myopathy.[46] These observations suggest that domain-specific effects downstream to domain destabilization may be important for phenotype specification.

Our work reinforces the growing evidence that the genetic landscape associating titin with DCM goes beyond truncations and includes also missense variants.[19] Traditionally, titin missense variants have been classified as VUSs because, as a whole, they are not enriched in DCM cohorts.[12] However, this lack of general enrichment probably reflects the large relative weight of highly abundant, non-damaging polymorphisms. Our results illustrate that searching for specific damaging molecular features is useful to identify new DCM-causing variants. We anticipate this approach can contribute to reduce the number of DCM patients with unexplained genetic etiology and to inform development of tailored, mechanism-based preventive and therapeutic strategies.[3] Specifically, we envision that emerging high-throughput *in silico*[47–50] and *in vitro*[51] methods will be able to capture pathogenicity potential related to domain destabilization of all possible titin missense variants. Identification of these pathogenic variants is a prerequisite for the development of therapies able to

Please cite this article in press as: Martinez-Martin et al., Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy, Cell Reports (2023), https://doi.org/10.1016/j.celrep.2023.113490

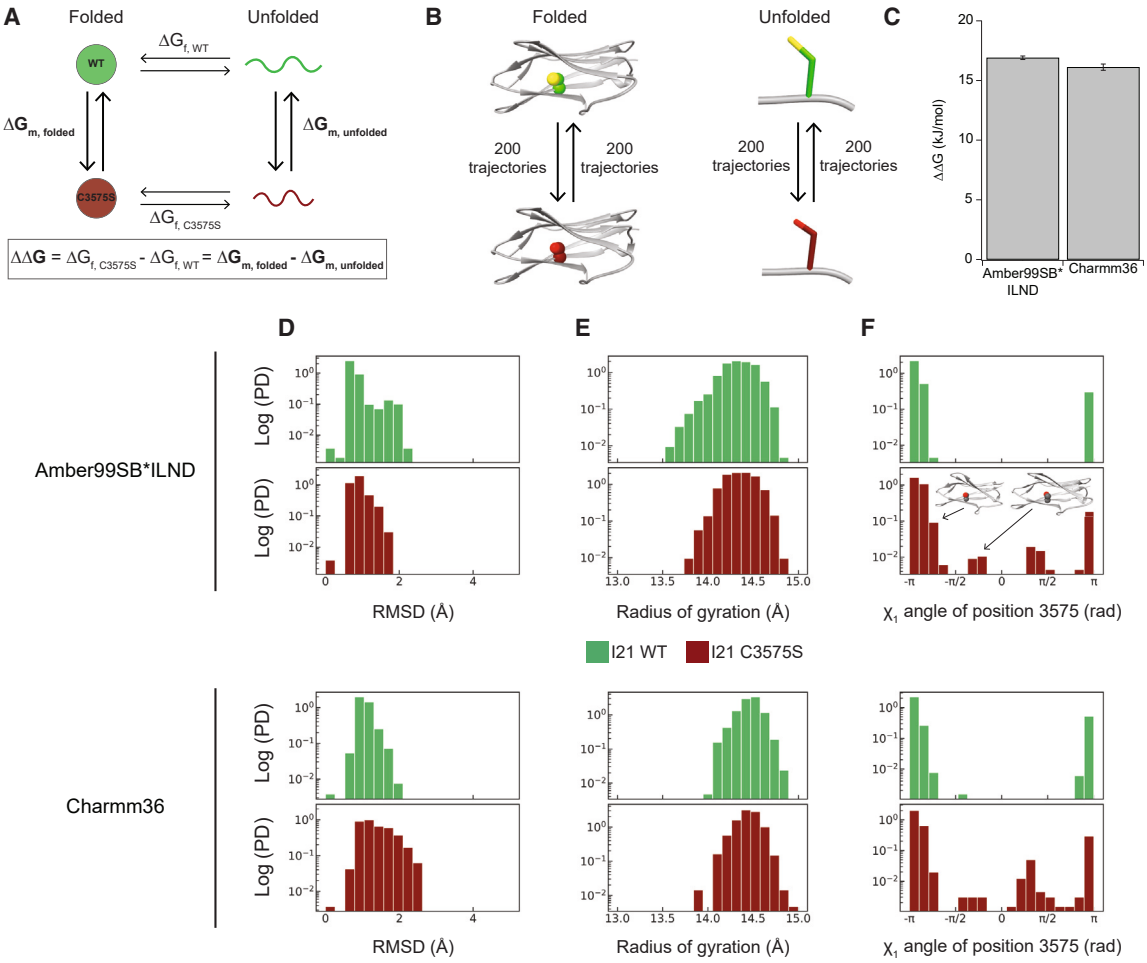


Figure 5. Molecular dynamics simulations capture destabilization of I21 C3575S domain

(A) Thermodynamic cycle used for ΔΔG calculations.

(B) Non-physical transitions used by the alchemical MD-based free-energy calculation.

(C) ΔΔG values using two different MD force fields. Error bars show standard deviation.

(D) Probability density histograms of global RMSD values (alpha carbons) in the folded state of wild-type and mutant domains.

(E) Probability density histograms of radius of gyration in the folded state of wild-type and mutant domains.

(F) Probability density histograms of X1 angle values of position 3575 in the folded state of wild-type and mutant domains. A ribbon cartoon of the I21 domain illustrates the X1 angles seen in the trajectories. Histograms in (D)–(F) were calculated pooling data from two independent trajectories with Amber99SB*ILND (top) or Charmm36 (bottom) force fields. See also Figures S2 and S3.

counterbalance domain destabilization, including strategies based on exon skipping[52] or pharmacological chaperones.[53] Furthermore, considering the broad functions played by Ig proteins,[22] which indeed are the most common family in the human genome,[54] we propose that searching for hydrophobicity-reducing variants as we have done here could benefit management of a wide range of human conditions including other (cardio) myopathies,[55,56] cancer,[57] and neurodegeneration.[58]

Limitations of the study

Assignment of variant pathogenicity in a complex, multifunctional protein like titin is challenging. Our work identifies variants reducing hydrophobicity of titin domains as causative of DCM. The approach we have followed builds on the high-resolution structure determination of bacterially expressed single titin domains, an approach that has limited ability to capture alternative molecular pathomechanisms such as those derived from defective protein-protein interactions or posttranslational modifications. Due to the scarce structural information available for titin (<15% of the >150 titin Ig domains), the accuracy of our estimates of SAS values could be improved as more high-resolution structures of titin domains become available. To limit complexity that stems from RNA splicing of the titin transcript, our analyses have focused on the N2B isoform because its exons are constitutively expressed in cardiac titin. However, sensitivity of variant identification may be increased by studying variants in domains belonging to the longer N2BA isoform, maybe at the expense of specificity. Finally, our analysis has identified a few hydrophobicity-reducing variants in non-cardiomyopathy patients. The reasons for this, which may range from inaccurate SAS predictions

Please cite this article in press as: Martinez-Martin et al., Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy, Cell Reports (2023), https://doi.org/10.1016/j.celrep.2023.113490

CellPress
OPEN ACCESS

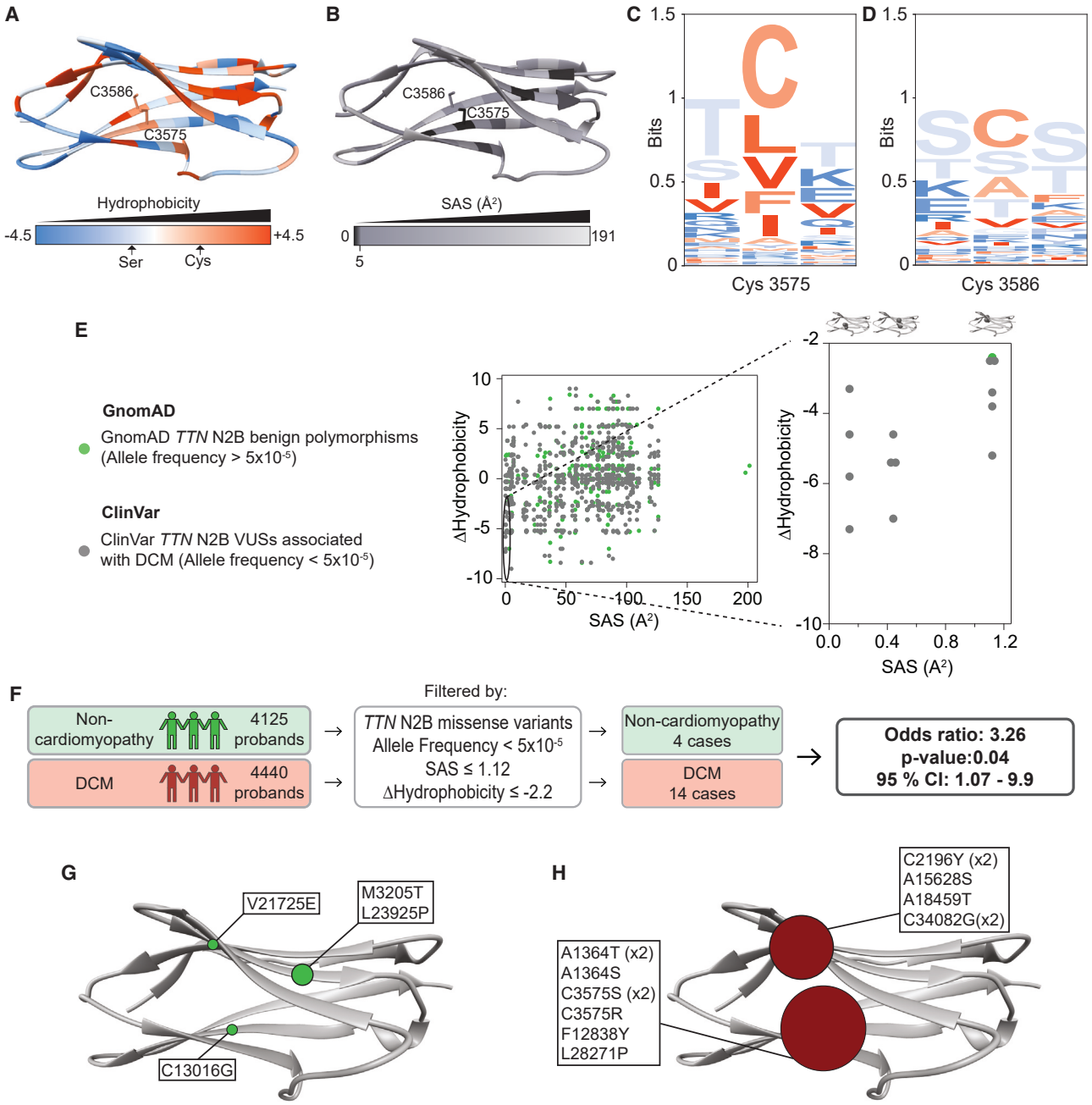


Figure 6. Variants reducing hydrophobicity of the core of titin Ig domains associate specifically with dilated cardiomyopathy

(A) Cartoon ribbon representation of I21 WT colored according to residue hydrophobicity (Kyte-Doolittle scale). C3575 and C3586 are represented as sticks.

(B) Cartoon ribbon representation of I21 WT colored according to residue SAS. C3575 and C3586 are represented as sticks.

(C and D) Sequence logos of positions equivalent to 3574–3576 (C) and 3585–3587 (D) in the alignment of titin Ig domains.[24] Residues are colored according to their hydrophobicity.

(E) SAS and ΔHydrophobicity values of all DCM-associated VUSs (gray) and benign polymorphisms (green). Zoom of variants with SAS <1.12 Å^2 and ΔHydrophobicity <−2.2 is on the right.

(F) Scheme of the case-control study showing specific association of hydrophobicity-reducing variants in the core of titin Ig domains with DCM. Odds ratio, p value, and 95% confidence interval (95% CI) are indicated.

(G and H) Schematic representation of the positions of the hydrophobicity-reducing titin variants affecting deeply buried residues found in the non-cardiomyopathy (G, green) and DCM (H, dark red) cohorts. Variants are placed in the equivalent positions of the I21 structure based on the titin domain sequence alignment also used to assign SAS values.[24] The size of the circles represents the number of variants found for each position. See also Figure S4 and Table S2.

or patient diagnosis to singularities of the targeted domains, remain unknown. As a consequence, identification of variants reducing core hydrophobicity in titin domains can contribute to pathogenicity assessment but always in the context of clinical and genetic evaluation of patients and their families.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Protein expression and purification for X-Ray crystallography
  - Protein crystallization
  - X-Ray data collection and processing
  - Protein expression and purification for biochemistry and circular dichroism
  - NEM-PEGylation assay for the assessment of cysteine redox state
  - Circular dichroism
  - Molecular dynamics (MD) simulations and alchemical free-energy calculations
  - Analysis of *TTN* variants from public databases
  - Case-control study
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113490.

ACKNOWLEDGMENTS

J.A.-C. acknowledges funding from the Ministerio de Ciencia e Innovación (MCIN, MCIN/AEI/10.13039/501100011033) through grants BIO2017-83640-P (AEI/FEDER, EU), PID2020-120426GB-I00, and RED2022-134242-T and the Regional Government of Madrid (grant Tec4Bio S2018/NMT-4443, 50% co-financed by the European Social Fund and the European Regional Development Fund for the programming period 2014–2020). The CNIC is supported by the Instituto de Salud Carlos III (ISCIII), the MCIN, and the Pro CNIC Foundation and is a Severo Ochoa Center of Excellence (grant CEX2020-001041-S funded by MCIN). I.M.-M. holds a fellowship from 'la Caixa' Foundation (ID 100010434, fellowship code LCF/BQ/DR20/11790009) and received support from a Erasmus + Training fellowship. Financial support to D.d.S. comes from Eusko Jaurlaritza (Basque Government) through the project IT1584-22 and from the MCIN through grants PID2021-127907NB-I00 (AEI/FEDER, UE) and RYC-2016-19590 (AEI/FSE, EU). We thank Maria Rosaria Pricolo for analysis support with database analysis, Vytautas Gapsys for his assistance using the pmx software package, and the staff at the DIPC Supercomputing Center and the Sample Preparation and Characterization facility of the EMBL Hamburg Unit for technical support. The synchrotron data were collected at beamline operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We thank the Spectroscopy and Nuclear Magnetic Resonance Core Unit at CNIO for access to CD instrumentation. Biorender.com was used as a source for icons included in some of the figures of this paper.

AUTHOR CONTRIBUTIONS

I.M.-M. and J.A.-C. conceived and led the project. I.M.-M., A.C., S.A.M., and M.W. designed and performed the structural determination of titin domains. I.M.-M., D.V.-C., and E.H.-G. did the biochemical and circular dichroism experiments. I.M.-M., D.d.S., and J.D. did computational analysis and protein stability predictions. I.M.-M., J.P.O., F.D., and P.G.-P. procured and analyzed genetic data. I.M.-M. and J.A.-C. drafted the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 25, 2023
Revised: September 28, 2023
Accepted: November 6, 2023

REFERENCES

1. Schultheiss, H.P., Fairweather, D., Caforio, A.L.P., Escher, F., Hershberger, R.E., Lipshultz, S.E., Liu, P.P., Matsumori, A., Mazzanti, A., McMurray, J., and Priori, S.G. (2019). Dilated cardiomyopathy. Nat. Rev. Dis. Primers *5*. 32–19.
2. Weintraub, R.G., Semsarian, C., and Macdonald, P. (2017). Dilated cardiomyopathy. Lancet *390*, 400–414.
3. Fatkin, D., Huttner, I.G., Kovacic, J.C., Seidman, J.G., and Seidman, C.E. (2019). Precision Medicine in the Management of Dilated Cardiomyopathy: JACC State-of-the-Art Review. J. Am. Coll. Cardiol. *74*, 2921–2938.
4. Japp, A.G., Gulati, A., Cook, S.A., Cowie, M.R., and Prasad, S.K. (2016). The Diagnosis and Evaluation of Dilated Cardiomyopathy. J. Am. Coll. Cardiol. *67*, 2996–3010.
5. Rosenbaum, A.N., Agre, K.E., and Pereira, N.L. (2020). Genetics of dilated cardiomyopathy: practical implications for heart failure management. Nat. Rev. Cardiol. *17*, 286–297.
6. Herman, D.S., Lam, L., Taylor, M.R.G., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S.R., McDonough, B., Sparks, E., et al. (2012). Truncations of Titin Causing Dilated Cardiomyopathy. N. Engl. J. Med. *366*, 619–628.
7. Fomin, A., Gärtner, A., Cyganek, L., Tiburcy, M., Tuleta, I., Wellers, L., Folsche, L., Hobbach, A.J., von Frieling-Salewsky, M., Unger, A., et al. (2021). Truncated titin proteins and titin haploinsufficiency are targets for functional recovery in human cardiomyopathy due to TTN mutations. Sci. Transl. Med. *13*, 1–16.
8. Linke, W.A., and Hamdani, N. (2014). Gigantic Business. Circ. Res. *114*, 1052–1068.
9. LeWinter, M.M., and Granzier, H.L. (2014). Cardiac Titin and Heart Disease. J. Cardiovasc. Pharmacol. *63*, 207–212.
10. Freundt, J.K., and Linke, W.A. (2019). Titin as a force-generating muscle protein under regulatory control. J. Appl. Physiol. *126*, 1474–1482.
11. Neagoe, C., Opitz, C.A., Makarenko, I., and Linke, W.A. (2003). Gigantic variety: expression patterns of titin isoforms in striated muscles and consequences for myofibrillar passive stiffness. J. Muscle Res. Cell Motil. *24*, 175–189.
12. Akinrinade, O., Heliö, T., Lekanne Deprez, R.H., Jongbloed, J.D.H., Boven, L.G., van den Berg, M.P., Pinto, Y.M., Alastalo, T.P., Myllykangas, S., and Koskenvuo, J. (2019). Relevance of Titin Missense and Non-Frameshifting Insertions/Deletions Variants in Dilated Cardiomyopathy. Sci. Rep. *9*, 1–9.
13. Begay, R.L., Graw, S., Sinagra, G., Merlo, M., Slavov, D., Gowan, K., Jones, K.L., Barbati, G., Spezzacatene, A., Brun, F., et al. (2015). Role of

titin missense variants in dilated cardiomyopathy. J. Am. Heart Assoc. *4*, e002645.

14. Ware, J.S., and Cook, S.A. (2018). Role of titin in cardiomyopathy: From DNA variants to patient stratification. Nat. Rev. Cardiol. *15*, 241–252.
15. Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitás, K., Sasse-Klaassen, S., Seidman, J.G., Seidman, C., Granzier, H., Labeit, S., et al. (2002). Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. Nat. Genet. *30*, 201–204.
16. Hinson, J.T., Chopra, A., Nafissi, N., Polacheck, W.J., Benson, C.C., Swist, S., Gorham, J., Yang, L., Schafer, S., Sheng, C.C., et al. (2015). Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. Science *349*, 982–986.
17. Hastings, R., De Villiers, C.P., Hooper, C., Ormondroyd, L., Pagnamenta, A., Lise, S., Salatino, S., Knight, S.J.L., Taylor, J.C., Thomson, K.L., et al. (2016). Combination of Whole Genome Sequencing, Linkage, and Functional Studies Implicates a Missense Mutation in Titin as a Cause of Autosomal Dominant Cardiomyopathy with Features of Left Ventricular Noncompaction. Circ. Cardiovasc. Genet. *9*, 426–435.
18. Jiang, H., Hooper, C., Kelly, M., Steeples, V., Simon, J.N., Beglov, J., Azad, A.J., Leinhos, L., Bennett, P., Ehler, E., et al. (2021). Functional analysis of a gene-edited mouse model to gain insights into the disease mechanisms of a titin missense variant. Basic Res. Cardiol. *116*, 14.
19. Domínguez, F., Lalaguna, L., Martínez-Martín, I., Piqueras-Flores, J., Rasmussen, T.B., Zorio, E., Giovinazzo, G., Prados, B., Ochoa, J.P., García-Pavía, P., et al. (2023). Titin Missense Variants as a Cause of Familial Dilated Cardiomyopathy. Circulation *147*, 1711–1713.
20. Pfuhl, M., and Pastore, A. (1995). Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. Structure *3*, 391–401.
21. Wang, J.-H. (2013). The sequence signature of an Ig-fold. Protein Cell *4*, 569–572.
22. Halaby, D.M., Poupon, A., and Mornon, J. (1999). The immunoglobulin fold family: Sequence analysis and 3D structure comparisons. Protein Eng. *12*, 563–571.
23. Chothia, C. (1975). Structural invariants in protein folding. Nature *254*, 304–308.
24. Giganti, D., Yan, K., Badilla, C.L., Fernandez, J.M., and Alegre-Cebollada, J. (2018). Disulfide isomerization reactions in titin immunoglobulin domains enable a mode of protein elasticity. Nat. Commun. *9*, 185.
25. Herrero-Galán, E., Martínez-Martín, I., Sánchez-González, C., Vicente, N., Bonzón-Kulichenko, E., Calvo, E., Suay-Corredera, C., Pricolo, M.R., Fernández-Trasancos, Á., Velázquez-Carreras, D., et al. (2022). Basal oxidation of conserved cysteines modulates cardiac titin stiffness and dynamics. Redox Biol. *52*, 102306.
26. Seeliger, D., and de Groot, B.L. (2010). Protein thermostability calculations using alchemical free energy simulations. Biophys. J. *98*, 2309–2316.
27. Aldeghi, M., Gapsys, V., and De Groot, B.L. (2018). Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. ACS Cent. Sci. *4*, 1708–1718.
28. Gapsys, V., Michielssens, S., Seeliger, D., and de Groot, B.L. (2016). Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angew. Chem. *128*, 7490–7494.
29. Berezovsky, I.N., Chen, W.W., Choi, P.J., and Shakhnovich, E.I. (2005). Entropic stabilization of proteins and its proteomic consequences. PLoS Comput. Biol. *1*, e47.
30. Miyazawa, S., and Jernigan, R.L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J. Mol. Biol. *256*, 623–644.
31. Tesei, G., Schulze, T.K., Crehuet, R., and Lindorff-Larsen, K. (2021). Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. Proc. Natl. Acad. Sci. USA *118*, e2111696118.
32. Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.
33. Nagano, N., Ota, M., and Nishikawa, K. (1999). Strong hydrophobic nature of cysteine residues in proteins. FEBS Lett. *458*, 69–71.
34. Walsh, R., Thomson, K.L., Ware, J.S., Funke, B.H., Woodley, J., McGuire, K.J., Mazzarotto, F., Blair, E., Seller, A., Taylor, J.C., et al. (2017). Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet. Med. *19*, 192–203.
35. Bourfiss, M., Van Vugt, M., Alasiri, A.I., Ruijsink, B., Van Setten, J., Schmidt, A.F., Dooijes, D., Puyol-Antón, E., Velthuis, B.K., van Tintelen, J.P., et al. (2022). Prevalence and Disease Expression of Pathogenic and Likely Pathogenic Variants Associated with Inherited Cardiomyopathies in the General Population. Circ. Genom. Precis. Med. *15*, E003704.
36. Verdonschot, J.A.J., Hazebroek, M.R., Krapels, I.P.C., Henkens, M.T.H.M., Raafs, A., Wang, P., Merken, J.J., Claes, G.R.F., Vanhoutte, E.K., van den Wijngaard, A., et al. (2020). Implications of Genetic Testing in Dilated Cardiomyopathy. Circ. Genom. Precis. Med. *13*, 476–487.
37. Escobar-Lopez, L., Ochoa, J.P., Mirelis, J.G., Espinosa, M.Á., Navarro, M., Gallego-Delgado, M., Barriales-Villa, R., Robles-Mezcua, A., Basurte-Elorz, M.T., Gutiérrez García-Moreno, L., et al. (2021). Association of Genetic Variants With Outcomes in Patients With Nonischemic Dilated Cardiomyopathy. J. Am. Coll. Cardiol. *78*, 1682–1699.
38. Fowler, S.B., and Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state. Structure *9*, 355–366.
39. Roberts, A.M., Ware, J.S., Herman, D.S., Schafer, S., Baksi, J., Bick, A.G., Buchan, R.J., Walsh, R., John, S., Wilkinson, S., et al. (2015). Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. Sci. Transl. Med. *7*, 270ra6.
40. McAfee, Q., Chen, C.Y., Yang, Y., Caporizzo, M.A., Morley, M., Babu, A., Jeong, S., Brandimarto, J., Bedi, K.C., Flam, E., et al. (2021). Truncated titin proteins in dilated cardiomyopathy. Sci. Transl. Med. *13*, eabd7287.
41. Zou, J., Tran, D., Baalbaki, M., Tang, L.F., Poon, A., Pelonero, A., Titus, E.W., Yuan, C., Shi, C., Patchava, S., et al. (2015). An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of Titin in zebrafish. Elife *4*, e09406.
42. Suay-Corredera, C., Pricolo, M.R., Herrero-Galán, E., Velázquez-Carreras, D., Sánchez-Ortiz, D., García-Giustiniani, D., Delgado, J., Galano-Frutos, J.J., García-Cebollada, H., Vilches, S., et al. (2021). Protein haploinsufficiency drivers identify MYBPC3 variants that cause hypertrophic cardiomyopathy. J. Biol. Chem. *297*, 100854.
43. Rivas-Pardo, J.A., Li, Y., Mártonfalvi, Z., Tapia-Rojo, R., Unger, A., Fernández-Trasancos, Á., Herrero-Galán, E., Velázquez-Carreras, D., Fernández, J.M., Linke, W.A., and Alegre-Cebollada, J. (2020). A HaloTag-TEV genetic cassette for mechanical phenotyping of proteins from tissues. Nat. Commun. *11*, 2060.
44. Bogomolovas, J., Fleming, J.R., Anderson, B.R., Williams, R., Lange, S., Simon, B., Khan, M.M., Rudolf, R., Franke, B., Bullard, B., et al. (2016). Exploration of pathomechanisms triggered by a single-nucleotide polymorphism in titin's I-band: the cardiomyopathy-linked mutation T2580I. Open Biol. *6*, 160114.
45. Anderson, B.R., Bogomolovas, J., Labeit, S., and Granzier, H. (2013). Single molecule force spectroscopy on titin implicates immunoglobulin domain stability as a cardiac disease mechanism. J. Biol. Chem. *288*, 5303–5315.
46. Rees, M., Nikoopour, R., Fukuzawa, A., Kho, A.L., Fernandez-Garcia, M.A., Wraige, E., Bodi, I., Deshpande, C., Özdemir, Ö., Daimagüler, H.S., et al. (2021). Making sense of missense variants in TTN-related congenital myopathies. Acta Neuropathol. *141*, 431–453.
47. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

48. Thompson, A.D., Helms, A.S., Kannan, A., Yob, J., Lakdawala, N.K., Wittekind, S.G., Pereira, A.C., Jacoby, D.L., Colan, S.D., Ashley, E.A., et al. (2021). Computational prediction of protein subdomain stability in MYBPC3 enables clinical risk stratification in hypertrophic cardiomyopathy and enhances variant interpretation. Genet. Med. *23*, 1281–1287.

49. Suay-Corredera, C., and Alegre-Cebollada, J. (2021). Correspondence on "Computational prediction of protein subdomain stability in MYBPC3 enables clinical risk stratification in hypertrophic cardiomyopathy and enhances variant interpretation" by Thompson et al. Genet. Med. *23*, 2009–2010.

50. Buel, G.R., and Walters, K.J. (2022). Can AlphaFold2 Predict the Impact of Missense Mutations on Structure? Nat. Struct. Molecular Bio. *29*, 1–2.

51. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature *562*, 217–222.

52. Romano, R., Ghahremani, S., Zimmerman, T., Legere, N., Thakar, K., Ladha, F.A., Pettinato, A.M., and Hinson, J.T. (2022). Reading Frame Repair of TTN Truncation Variants Restores Titin Quantity and Functions. Circulation *145*, 194–205.

53. Gil-Martínez, J., Bernardo-Seisdedos, G., Mato, J.M., and Millet, O. (2022). The use of pharmacological chaperones in rare diseases caused by reduced protein stability. Proteomics *22*, e2200222.

54. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

55. Scharner, J., Lu, H.C., Fraternali, F., Ellis, J.A., and Zammit, P.S. (2014). Mapping disease-related missense mutations in the immunoglobulin-like fold domain of lamin A/C reveals novel genotype-phenotype associations for laminopathies. Proteins *82*, 904–915.

56. Suay-Corredera, C., Alegre-Cebollada, J., Suay-Corredera, C., and Alegre-Cebollada, J. (2022). The mechanics of the heart: zooming in on hypertrophic cardiomyopathy and cMyBP-C. FEBS Lett. *596*, 703–746.

57. Vattepu, R., Klausmeyer, R.A., Ayella, A., Yadav, R., Dille, J.T., Saiz, S.V., and Beck, M.R. (2020). Conserved tryptophan mutation disrupts structure and function of immunoglobulin domain revealing unusual tyrosine fluorescence. Protein Sci. *29*, 2062–2074.

58. Dean, H.B., Roberson, E.D., and Song, Y. (2019). Neurodegenerative Disease–Associated Variants in TREM2 Destabilize the Apical Ligand-Binding Region of the Immunoglobulin Domain. Front. Neurol. *10*, 1252.

59. Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

60. Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. *43*, W174–W181.

61. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

62. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

63. Zheng, H., Cooper, D.R., Porebski, P.J., Shabalin, I.G., Handing, K.B., and Minor, W. (2017). CheckMyMetal : a macromolecular metal-binding validation tool. Acta Crystallogr. D Struct. Biol. *73*, 223–233.

64. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

65. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics *29*, 845–854.

66. Gapsys, V., Michielssens, S., Seeliger, D., and De Groot, B.L. (2015). pmx: Automated protein structure and topology generation for alchemical perturbations. J. Comput. Chem. *36*, 348–354.

67. McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J., and Pande, V.S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys. J. *109*, 1528–1532.

68. Costa, S., Almeida, A., Castro, A., and Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. Front. Microbiol. *5*, 63.

69. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, pp. 571–607.

70. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

71. Mills, J.E., and Dean, P.M. (1996). Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. J. Comput. Aided Mol. Des. *10*, 607–622.

72. Pimenta-Lopes, C., Suay-Corredera, C., Velázquez-Carreras, D., Sánchez-Ortiz, D., and Alegre-Cebollada, J. (2019). Concurrent atomic force spectroscopy. Commun. Phys. *2*, 1–11.

73. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

74. Gapsys, V., and De Groot, B.L. (2017). pmx Webserver: A User Friendly Interface for Alchemistry. J. Chem. Inf. Model. *57*, 109–114.

75. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

76. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–1958.

77. Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. J. Chem. Theory Comput. *8*, 3257–3273.

78. Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

79. Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

80. Sikosek, T. (ed) Computational Methods in Protein Evolution. Methods in Molecular Biology, *1851*. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8736-8_2.

81. Li, D.-W., and Brüschweiler, R. (2009). In silico Relationship between Configurational Entropy and Soft Degrees of Freedom in Proteins and Peptides. Phys. Rev. Lett. *102*, 118108.

82. Best, R.B., Hummer, G., and Eaton, W.A. (2013). Native contacts determine protein folding mechanisms in atomistic simulations. Proc. Natl. Acad. Sci. USA *110*, 17874–17879.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli: strain BL21	Alegre-Cebollada Lab	N/A
Escherichia coli: strain BLR (DE3)	Alegre-Cebollada Lab	N/A
Chemicals, peptides, and recombinant proteins		
NcoI-HF restriction enzyme	New England Biolabs	Cat# R3193S
BamHI-HF restriction enzyme	New England Biolabs	Cat# R3136S
BamHI restriction enzyme	New England Biolabs	Cat# R0136S
BglII restriction enzyme	New England Biolabs	Cat# R0144S
Methoxypolyethylene glycol maleimide (mPEG)	Sigma-Aldrich	Cat# 63187; CAS: 99126-64-4
N-ethyl maleimide	Sigma-Aldrich	Cat# E3876; CAS: 128-53-0
Critical commercial assays		
Morpheus HT-96 screening plate	Molecular Dimensions	Cat# MD1-47
PEGs I suite screening plate	Qiagen/Nextal	Cat# 130904
Deposited data		
Titin I21 WT crystal structure	This paper	PDB ID: 8OVU
Titin I21 C3575S crystal structure	This paper	PDB ID: 8P35
MD simulations files and scripts	This paper	Zenodo: https://doi.org/10.5281/zenodo.8383009
Recombinant DNA		
Titin I21 WT cDNA	Alegre-Cebollada Lab	N/A
Titin I21 C3575S cDNA	Alegre-Cebollada Lab	N/A
pTEM14 expression vector	Wilmanns Lab	N/A
pETtrx_1a expression vector	Wilmanns Lab	N/A
pQE80-based expression plasmid	Alegre-Cebollada Lab	N/A
Software and algorithms		
XDS Program Package	Kabsch et al.[59]	RRID:SCR_015652; https://xds.mr.mpg.de/html_doc/downloading.html
I-TASSER	Yang et al.[60]	RRID:SCR_014627; http://zhanglab.ccmb.med.umich.edu/I-TASSER/
COOT	Emsley et al.[61]	RRID:SCR_014222; http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Afonine et al.[62]	RRID:SCR_014224; https://www.phenix-online.org/
CheckMyMetal	Zheng et al.[63]	RRID:SCR_016887; https://csgid.org/csgid/metal_sites
Chimera USFC	Pettersen et al.[64]	RRID:SCR_004097; http://plato.cgl.ucsf.edu/chimera/
GROMACS (v 2021.4)	Pronk et al.[65]	RRID:SCR_014565; http://www.gromacs.org
pmx	Gapsys et al.[66]	http://pmx.mpibpc.mpg.de/instructions.html
MDtraj Python library	McGibbon et al.[67]	https://www.mdtraj.org/1.9.8.dev0/index.html
Other		
Benign variants database	Genome Aggregation Database v2.1.1	RRID:SCR_014964 https://gnomad.broadinstitute.org/
VUS variants database	ClinVar	RRID:SCR_006169 http://www.ncbi.nlm.nih.gov/clinvar/
Health in code database	Health in code	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ines Martinez-Martin (email: ines.martinez@cnic.es).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact upon request.

Data and code availability

- Structural data has been deposited at the Protein Data Bank (PDB) database and is publicly available as of the date of publication. Accession numbers are listed in the Key resources table.
- All code related to MD simulations has been deposited at Zenodo and is publicly available as of the date of publication. DOI is indicated in the Key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study uses *E.coli* strains BL21 and BLR (DE3) for production of recombinant proteins. Cells are grown at 37°C in LB Agar plates supplemented with antibiotics for colony selection and at 37°C and 250 rpm in LB medium supplemented with antibiotics for liquid cultures. After expression induction with IPTG the cells are incubated at different temperatures according to expression conditions (see Method details for expression conditions).

METHOD DETAILS

Protein expression and purification for X-Ray crystallography

Human titin I21 WT cDNA was cloned in a pTEM14 expression vector using NcoI-HF and BamHI-HF restriction enzymes (sequence available in Data S3) and expressed in *E. coli* BL21 cells by overnight induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C and 250 rpm agitation. This protocol did not result in enough protein concentration in the case of the I21 C3575S domain. Hence, the mutant domain was cloned using NcoI-HF and BamHI-HF restriction enzymes in a pETtrx_1a plasmid vector, which includes a cleavable thioredoxin tag (TrxA) that works as a solubility enhancer for recombinant proteins expressed in *E. coli*[68] (sequence available in Data S3). The protein expression protocol was similar to the one used for the WT but the temperature was decreased to 16°C to improve protein solubility. In both cases the cells were lysed by sonication in purification buffer (30mM MES buffer pH 5.5, 100 mM NaCl, 5% glycerol for the I21 WT and 30mM sodium phosphate buffer pH 7.2, 100 mM NaCl, 5% glycerol for the I21 C3575S domain) supplemented with 20 mM imidazole, 0.01 mg/mL DNAse, 100 μg/mL lysozyme and a Complete EDTA-free protease inhibitor mixture tablet (Roche Applied Science). Then, proteins were purified using chromatography-based methods. First, proteins were purified by nickel-affinity chromatography and eluted using purification buffer including 300mM imidazole. Then, proteins were dialyzed overnight in purification buffer including 10 mM imidazole in the presence of specific proteases to remove the protein tags at a 1:100 protease:protein ratio (3C protease for the WT domain and TEV protease for the C3575S domain). After dialysis, proteins were separated from their tags by reverse nickel-affinity chromatography, concentrated and further purified by size-exclusion chromatography in purification buffer including 10 mM imidazole using the column HiLoad 16/600 Superdex 75 prep grade (GE healthcare). Finally, proteins were dialyzed in purification buffer to completely remove imidazole, concentrated to saturation, and stored at −80°C. All chromatography steps were done in Fast Protein Liquid Chromatography systems (GE Healthcare).

Protein crystallization

Crystallization conditions were screened using sitting drop vapor diffusion using 18.3 mg/mL I21 WT or 13 mg/mL I21 C3575S. To estimate concentrations, theoretical extinction coefficients were used ($E^{0.1\%}$ = 1.032, $E^{0.1\%}$ = 1.036 for WT and mutant domains, respectively).[69] Screenings were done using 1:1 and 1:1.5 protein:mother liquor mixtures on 96-condition commercial protein crystallization screening plates. I21 WT crystallized in a buffer containing 0.06 M divalent ions ($CaCl_2$, $MgCl_2$), 0.1 M HEPES sodium salt, pH 7.5, 50% v/v Polyethylene glycol monomethyl ether (PEGMME) 550, PEG 20K from Morpheus HT-96 screening plate (Molecular Dimensions). I21 C3575S crystallized in 0.1 M HEPES sodium salt, pH 7.5, 25%(w/v) PEG 2000 MME from PEGs I suite screening plate (Qiagen/Nextal). Then, crystals were soaked in cryo-solution containing the crystallization mother liquor including 26% (v/v) ethylene glycol, mounted on cryo-loops (Hampton Research) and fast-cooled in liquid nitrogen.

X-Ray data collection and processing

Diffraction data were collected on the synchrotron radiation beamlines EMBL P14 (WT domain) and P13 (C3575S domain) at Petra III (EMBL/DESY, Hamburg) and processed using XDS.[59] I21 WT and I21 C3575S structures were solved by molecular replacement using

Phaser.[70] For the I21 WT domain the reference structure was a homology model generated by I-Tasser[60] using the PDB: 1TIT structure and the I21 domain sequence. I21 C3575S structure was calculated using the solved I21 WT as a reference model. Manual modeling and refinement of both structures was carried out using COOT[61] and Phenix[62] (Data collection and refinement statistics in Data S1). The presence of the magnesium ion in the I21 structure was confirmed using CheckMyMetal.[63] Chimera USFC[64] was used for protein analysis and visualization. RMSD calculations refer to the structural alignment of alpha carbons in all cases. For structural comparison between WT and C3575S domains we chose chain A of the WT structure because it does not contain Mg^{2+} and chain A of the mutant, which has the best-defined electron density in the asymmetric unit. Assignment of hydrogen bonds was done according to Mills et al.[71]

Protein expression and purification for biochemistry and circular dichroism

I21 WT and C3575S cDNAs were cloned in a custom made pQE80-based expression plasmid (Qiagen) using BamHI and BglII restriction enzymes (sequences available in Data S3). Proteins were expressed in *E. coli* BLR (DE3) in both cases. I21 WT domain was expressed by incubation with 1 mM IPTG for 3 h at 37°C and 250 rpm agitation. For I21 C3575S IPTG concentration was reduced to 0.4 mM and incubated at 16°C overnight. Cells were lysed by sonication and French Press in purification buffer (50mM sodium phosphate buffer pH 7, 300 mM NaCl) supplemented with 5 μg/mL DNaseI (Roche), 5ug/mL RNAse (Sigma-Aldrich), 100 μg/mL lysozyme (Sigma-Aldrich), 10 mM $MgCl_2$ and protease inhibitor cocktail III (Calbiochem). Purification of His-tagged domains was achieved by nickel-affinity using purification buffer supplemented with 1 mM DTT and gel filtration chromatography in 50mM ammonium bicarbonate pH 7 using a Superdex 200 Increase 10/300 column (GE Healthcare) following published protocols.[72] After purification the proteins were lyophilized and stored at −20°C.

NEM-PEGylation assay for the assessment of cysteine redox state

Lyophilized proteins were reconstituted in 20 mM sodium phosphate pH 6.5, 50 mM NaCl. 1 μg of protein was incubated with 20 mM methoxypolyethylene glycol maleimide (mPEG, Sigma-Aldrich) in the presence of 3% SDS for 10 min at 60°C. Negative control samples were prepared using 10 mM N-ethymaleimide (NEM, Sigma-Aldrich) instead of mPEG. After incubation, results were analyzed by electrophoresis using 17% SDS-PAGE gels following standard protocols.[73]

Circular dichroism

CD spectra were collected using a Jasco J-810 spectropolarimeter. Lyophilized proteins were reconstituted in 20 mM NaPi pH 6.5, 50 mM NaCl and tested at 0.3–0.4 mg/mL protein concentration in 0.1-cm-pathlength quartz cuvettes. To estimate concentrations, theoretical extinction coefficients were used ($E^{0.1\%} = 0.924$, $E^{0.1\%} = 0.925$ for WT and mutant domains, respectively).[69] To study thermal denaturation, CD signal at 215 nm was monitored as temperature increased from 25°C to 85°C at a rate of 30°C/h. Temperature control was achieved using a Peltier thermoelectric system. To estimate T_m, changes in CD signal were fit to a sigmoidal function considering a two-state unfolding process using IGOR Pro (Wavemetrics).

Molecular dynamics (MD) simulations and alchemical free-energy calculations

For molecular dynamics simulations and alchemical free energy calculations we generated a hybrid structure and topology that is based on the I21 domain structure but contains a hybrid atom in position 3575 that changes from cysteine to serine according to the value of λ (λ = 0 for cysteine and λ = 1 for serine). This topology was generated using the pmx software package and webserver[66,74] by introducing the C3575S mutation in the I21 WT pdb file. Protein domains were protonated considering neutral pH and canonical pK_a values. Molecules were then introduced in dodecahedral boxes, solvated with TIP3P water molecules[75] and neutralized with Na^+ and Cl^- ions at 0.1 M concentration. The system was energy minimized using the steepest descent algorithm and equilibrated during 100 ps in the NPT ensemble. 1 μs equilibrium simulations were run for the WT (λ = 0) and mutant (λ = 1) versions of the topology using Amber99sb*ILDN[76] or Charmm36[77] force fields. In all simulations an integration step of 0.002 ps was used. Temperature was maintained at 298 K using a velocity-rescaling thermostat.[78] Pressure was maintained at 1 bar using a Parrinello-Rahman barostat.[79] Changes in folding free energy upon mutation (ΔΔG) were calculated from the equilibrium simulations using the protocol proposed by Aldeghi et al.[80] Briefly, 200 non-equilibrium 200 ps simulations with an integration step of 0.002 ps were run using 200 equidistant frames from the equilibrium trajectories. During these simulations, the λ value changed from 0 to 1 (or vice versa) at a $\Delta\lambda = 1 \times 10^{-5}$ per simulation step. Finally, the ΔG value was calculated from these non-equilibrium trajectories using the pmx analyze Python script, which integrates over the $\partial H/\partial\lambda$ curves obtained from the non-equilibrium trajectories and estimates the ΔG value using Bennet's Acceptance Ratio. This process was repeated for the unfolded state of the protein, which was approximated by a tripeptide with the cysteine (WT) or serine (C3575S) residue surrounded by two glycines.[26] The final ΔΔG (± standard deviation error) value was calculated according to:

$$\Delta\Delta G = \Delta G_{mutation,folded} - \Delta G_{mutation,unfolded}$$

All MD simulations were run using Gromacs version 2021.4.[65] Trajectories were analyzed using Gromacs and the MDtraj Phython library.[67] Changes in entropy associated with the variations in the distribution of rotameric states were calculated by integration of the X_1 angle distribution according to:

$$S = -k_B \int_0^{2\pi} P(\chi_1) \ln(P(\chi_1)) d\chi_1,$$

where $P(\chi_1)$ is the probability density of the X_1 angle and k_B is the Boltzmann constant.[81] Fraction of native contacts along MD trajectories was calculated using the definition from Best, Hummer, and Eaton.[82]

Analysis of *TTN* variants from public databases

A database of titin missense variants was built using information from publicly available databases. The DCM-associated group was obtained from ClinVar database. We included missense variants located in Ig domains of the constitutively expressed N2B isoform of titin and showing an allele frequency (AF) $< 5 \times 10^{-5}$, which is compatible with conservative estimates of DCM prevalence.[34,35] The group of benign variants was populated by N2B titin missense variants from the gnomAD with AF $> 5 \times 10^{-5}$. For all variants, the ΔHydrophobicity value was calculated according to the Kyte-Doolittle hydrophobicity scale[32] (ΔHydrophobicity = $\text{Hydrophobicity}_{\text{mutant}} - \text{Hydrophobicity}_{\text{native}}$). Given the high structural similarity of titin Ig domains, SAS values were estimated according to the average SAS of equivalent positions of 13 available crystal structures of Ig domains of human titin (Table S1; Figure S4).

Case-control study

The variants for the case-control study were obtained from the Health in Code database. The case group was built with a cohort of genotype-negative patients diagnosed with DCM who underwent genetic testing using an NGS library of 251 genes (including 121 DCM associated genes present in the Health in Code-DCM panel). The control group included patients diagnosed with diseases different from cardiomyopathies (mainly channelopathies and aortic diseases) who were referred for genetic testing and were sequenced with the same library as the case group. The odds ratio was calculated by determining the number of variants in each group that are expressed in the constitutive N2B titin, have an AF $< 5 \times 10^{-5}$, a SAS < 1.12 Å^2 and that induce a ΔHydrophobicity < -2.2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details of all quantifications can be found in the figure legend of the corresponding result and/or in the corresponding paragraph of the Method Details section. Statistical significance was considered when p value <0.05.