doi:10.1016/j.jmb.2011.10.029 J. Mol. Biol. (2011) 414, 735–748

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Human Cardiac Myosin Binding Protein C: Structural Flexibility within an Extended Modular Architecture

Cy M. Jeffries, Yanling Lu, Robert M. G. Hynson, James E. Taylor, Mercedes Ballesteros, Ann H. Kwan and Jill Trewhella*

School of Molecular Bioscience, University of Sydney, New South Wales 2006, Australia

Received 6 September 2011; received in revised form 15 October 2011; accepted 18 October 2011
Available online 25 October 2011

Edited by R. Huber

Keywords:
C protein;
actomyosin regulation;
cardiac muscle;
small-angle X-ray scattering;
NMR

New insights into the modular organization and flexibility of the N-terminal half of human cardiac myosin binding protein C (cMyBP-C) and information on the association state of the full-length protein have been deduced from a combined small-angle X-ray scattering (SAXS) and NMR study. SAXS data show that the first five immunoglobulin domains of cMyBP-C, which include those implicated in interactions with both myosin and actin, remain monodisperse and monomeric in solution and have a highly extended yet distinctively 'bent' modular arrangement that is similar to the giant elastic muscle protein titin. Analyses of the NMR and SAXS data indicate that a proline/alanine-rich linker connecting the cardiac-specific N-terminal C0 domain to the C1 domain provides significant structural flexibility at the N-terminus of the human isoform, while the modular arrangement of domains C1–C2–C3–C4 is relatively fixed. Domain fragments from the C-terminal half of the protein have a propensity to self-associate *in vitro*, while full-length bacterially expressed cMyBP-C forms flexible extended dimers at micromolar protein concentrations. In summary, our studies reveal that human cMyBP-C combines a distinctive modular architecture with regions of flexibility and that the N-terminal half of the protein is sufficiently extended to span the range of interfilament distances sampled within the dynamic environment of heart muscle. These structural features of cMyBP-C could facilitate its putative role as a molecular switch between actin and myosin and may contribute to modulating the transverse pliancy of the C-zone of the A-band across muscle sarcomeres.

Introduction

The primary contractile components of muscle are made up of the proteins actin and myosin. Muscle shortens as actin thin filaments slide past myosin thick filaments via the ATP-driven interaction of myosin heads (S1) that extend from the myosin filaments and bind to actin. Calcium signals, transmitted via the thin filament accessory proteins troponin and tropomyosin, are the primary regulators of contraction. Recently, myosin binding protein C (MyBP-C) has been identified as playing a role in modulating these calcium signals.[1–6] Although

*Corresponding author. E-mail address: jill.trewhella@sydney.edu.au.

Abbreviations used: cMyBP-C, cardiac myosin binding protein C; SAXS, small-angle X-ray scattering; MyBP-C, myosin binding protein C; P/A_L, proline/alanine-rich linker; EM, electron microscopy; EOM, ensemble optimization method; NOE, nuclear Overhauser enhancement; MALLS, multiangle laser light scattering; Na_2EDTA, disodium ethylene diaminetetraacetic acid; BME, β-mercaptoethanol; BA, benzamidine; TCEP, tris(2-carboxyethyl)phosphine; PDB, Protein Data Bank.

0022-2836/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.

MyBP-C is located in a very specific region of muscle sarcomeres (at every third level of myosin heads in the C-zone of the A-band[7–9]), mutations in the cardiac protein [cardiac myosin binding protein C (cMyBP-C)] have marked effects on cardiac output, and mutations in the *MYBPC3* gene encoding the protein are a leading cause of familial hypertrophic cardiomyopathy.[10–13]

First identified as a thick filament accessory protein,[14] MyBP-C has been isolated from smooth, skeletal, and cardiac muscles.[15] Belonging to the intracellular immunoglobulin (Ig)/fibronectin superfamily of proteins, the cardiac homologue consists of 11 domains designated C0 through C10 (Fig. 1a). The C-terminal domains (C7 through C10) anchor cMyBP-C to the thick filament by binding to meromyosin of the myosin heavy chain or to titin,[16–18] while there is accumulating evidence for a regulatory role of the N-terminal domains, via interactions with both myosin and actin, that is critical to maintaining regular heart function.[19–27] Of note, the N-terminal Ig domains C0–C1–C2 have the cardiac-specific C0 domain connected to C1 via an ~50-residue proline/alanine-rich linker (P/A_L), the proline/alanine composition of which has been correlated to heart rates,[28] and a module of unknown structure, called the 'motif' (or m-domain), is positioned between C1 and C2. The m-domain undergoes cardiac-specific phosphorylation[29–33] in response to inotropic stimuli[34] that 'untethers' the N-terminal domains of cMyBP-C from myosin ΔS2, a region of myosin proximal to the myosin head. This untethering allows myosin heads to reposition and bind actin and consequently affect the rate and force of heart muscle contraction[35–38] (Fig. 1b). More recently, it has been shown that C0 can bind to the regulatory light chain of myosin.[22] Thus, there have been a number of results showing that the N-terminal domains of cMyBP-C undergo key interactions with myosin in regions proximal to myosin heads to modulate cardiac output.

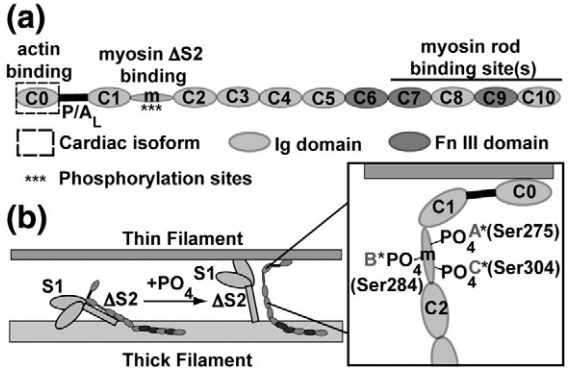

Fig. 1. Schematic representations of cMyBP-C structure and function. (a) An outline of the Ig and fibronectin organization of cMyBP-C and the position of P/A_L and the m-domain (m). (b) The tethering mechanism: The m-domain undergoes phosphorylation in response to inotropic stimuli in (at least) three phosphorylation sites (B, A, and C), resulting in cMyBP-C uncoupling from myosin ΔS2 and myosin head (S1) reorientation. A direct interaction with actin at the C0–C1 end of cMyBP-C may further regulate calcium signaling on the thin filament.

The N-terminal domains of cMyBP-C also interact specifically with actin[19,20,39,40] to affect calcium signals, thin filament activation, and myosin head cross-bridge kinetics.[1,2,4,5,21] Small-angle scattering studies revealed that mouse C0C2 (C0–C1–m–C2) adopts an extended conformation in solution[41] and that the domains stabilize and regularly decorate actin filaments, binding in a position where they are poised to interfere with myosin head attachment or to alter tropomyosin dynamics.[20] The interaction between the N-terminal domains of cMyBP-C and actin has been supported by a number of recent electron microscopy (EM) investigations,[19,42,43] including our recent study, which identifies two distinct cMyBP-C binding regions on actin.[43] Electron tomography investigations of intact muscle sarcomeres show that cMyBP-C spans distances across the interfilament space.[44] Furthermore, Kulikovskaya *et al.* showed that the N-terminal domains can alternate binding between actin and myosin, suggesting that 'cMyBP-C-mediated interplay' occurs between thick and thin filaments.[1] We have recently reported that a common myosin/actin interface could facilitate such a switch exists on C0C1.[45]

The architecture of cMyBP-C should be relatively defined but at the same time sufficiently flexible or elastic to withstand sustained mechanical movement to enable these interactions in the dynamic environment of heart muscle. Here we use small-angle X-ray scattering (SAXS) and NMR spectroscopy to show that the N-terminal region of human cMyBP-C has a modular architecture that adopts a highly extended conformation in solution incorporating a flexible linker between C0 and C1. In addition, the full-length protein forms flexible dimers, at micromolar protein concentrations, that are mediated by interactions involving the C-terminal domains.

Results and Discussion

The N-terminal region of human cMyBP-C is monomeric and elongated in solution

SAXS data (Fig. 2) were obtained from seven human cMyBP-C protein fragments that encompass the first five Ig domains of the protein: C0-P/A_L, C0C1, C0C2, C1C2, C2C4, C3C4, and C0C4. The naming convention for these fragments designates the first domain and the last domain in linear

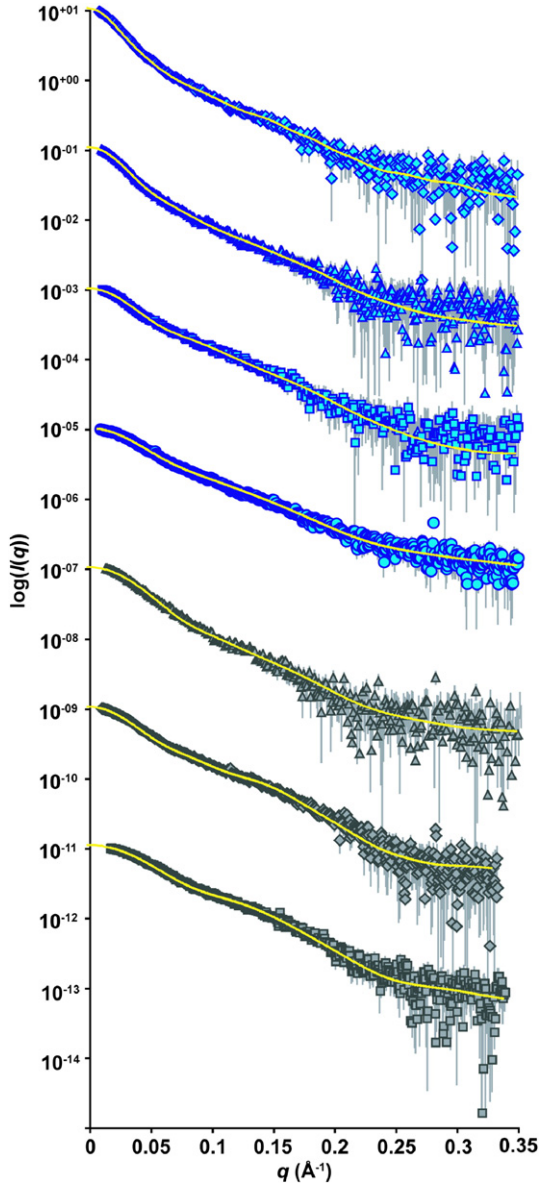

Fig. 2. SAXS data and model fits. SAXS data from C0C4 (blue diamonds), C0C2 (blue triangles), C0C1 (blue squares), C0-P/A_L (blue circles), C1C2 (gray triangles), C2C4 (gray diamonds), and C3C4 (gray squares). The continuous yellow lines show the average fits to the data of the ensemble of C0-P/A_L structures, as shown in Fig. 5, or the fits of the BUNCH-refined models, as shown in Fig. 6. The measured data were all placed on an absolute scale (Table 1) using scattering from water but are shown here multiplied on the $I(q)$ axis for clarity.

sequence only (e.g., C0C2 ≡ C0-P/A_L–C1–m–C2). Guinier plots of the SAXS data (Fig. 3a)[46] indicate a systematic increase in the radius of gyration (R_g) of the fragments as domains are added (e.g., from 32.5 Å for C0-P/A_L to 59.8 Å for C0C4; Table 1). Furthermore, excellent linear correlations within the Guinier regions of the profiles are observed (Fig. 3a, yellow lines; Supplementary Fig. 1), consistent with each cMyBP-C construct being free of aggregation or

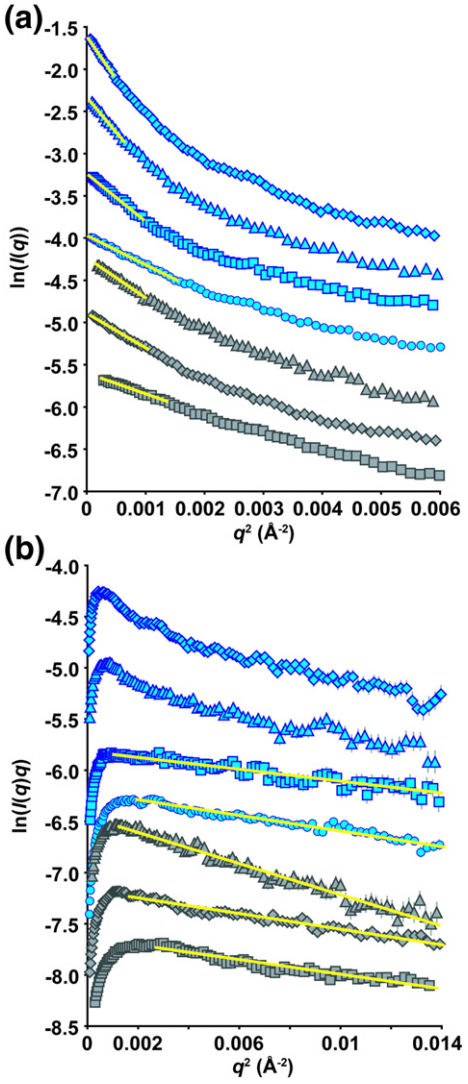

Fig. 3. Guinier analysis of the N-terminal domain fragments of human cMyBP-C. (a) Guinier plot and (b) modified Guinier plot of C0C4 (blue diamonds), C0C2 (blue triangles), C0C1 (blue squares), C0-P/A_L (blue circles), C1C2 (gray triangles), C2C4 (gray diamonds), and C3C4 (gray squares). The Guinier plots for all constructs in (a) show excellent linear correlations in the very-low-q^2 Guinier region (continuous yellow lines; $R^2 = 0.99$; see also Supplementary Fig. 1). The continuous parallel lines on the modified Guinier plots in (b) for C0-P/A_L, C0C1, C2C4, and C3C4 indicate that these constructs have very similar average radii of gyration of cross sections (R_g^c = 8.1–8.5 Å). The corresponding plot for C1C2 indicates that it is also extended but has a larger R_g^c (12.6 Å). The lack of a linear correlation after the 'knee' in the modified Guinier plots for C0C2 and C0C4 indicates that both constructs, although extended, are more irregular and do not have a uniform cross section compared to the other constructs.

Table 1. SAXS parameters for the N-terminal domain constructs of human cMyBP-C

	C0-P/A_L	C0C1	C0C2	C0C4	C1C2	C2C4	C3C4
Structural parameters							
$I(0)$ (cm^{-1})	0.1323	0.1231	0.1413	0.1615	0.1090	0.1782	0.1359
R_g (Å) (from $P(r)$)	33.9	44.1	51.6	59.3	39.7	37.3	28.2
R_g (Å) (from Guinier; $qR_g<1.0$)	32.5±0.3	42.3±1.3	51±1.2	60±1.2	37±1.0	37.2±0.4	27.2±0.3
R_g^c (Å)	8.5±0.05	8.1±0.1	—	—	12.6±0.2	8.5±0.04	8.1±0.04
D_{max} (Å)	120	166	177	209	143	127	95
Molecular mass determination							
Protein concentration (g cm^{-3})	0.00817	0.00555	0.00377	0.00292	0.00376	0.00672	0.00768
Partial specific volume (cm^3 g^{-1})	0.732	0.733	0.733	0.738	0.737	0.740	0.741
Contrast $\Delta\rho\times10^{10}$ cm^{-2})	2.964	2.850	2.852	2.801	2.848	2.854	2.844
Molecular mass (from $I(0)$) (Da)	20,700	30,600	51,600	77,800	39,600	35,700	24,000
Calculated molecular mass for the monomer (from sequence) (Da)	16,900	29,200	50,800	71,300	35,600	31,400	22,500

other factors that might bias derived structural parameters (e.g., interparticle interference). Analysis of the forward scattering intensity at zero angle ($I(0)$) yields molecular mass values similar to those expected for monodisperse and monomeric forms in solution (Table 1)—although for C0-P/A_L, the molecular mass determined from the data is ~20% higher than expected, which we attribute to increased error in protein concentration estimation due to the low extinction coefficient for this construct (0.500 g L^{-1}).

Modified Guinier plots for rod-like particles ($\ln I(q)q$ *versus* q^2; Fig. 3b)[47,48] for the N-terminal fragments generally have shapes indicative of extended particles in solution, consistent with their respective atom-pair distance distributions ($P(r)$ *versus* r; Fig. 4). The mean radii of gyration of cross section (R_g^c) derived from these plots are similar for C0-P/A_L and for the two-domain and three-domain constructs C0C1, C3C4, and C2C4 (8.5 Å, 8.1 Å, 8.1 Å, and 8.5 Å, respectively; Fig. 3b and Table 1), while the larger value (12.6 Å) for C1C2 indicates a significantly thicker average cross section. The plots for the longest constructs, C0C2 and C0C4, do not yield a single well-characterized slope corresponding to a well-defined average R_g^c value, suggesting that the organization of the domains in these larger fragments is more irregular. Indeed, as domains are added to the N-terminal fragments, the $P(r)$ profiles (Fig. 4) show an increasing number of shoulders at mid-range to long-range vector lengths with corresponding increases in the maximum particle dimension (D_{max}) so that, ultimately, C0C4 displays a distribution characteristic of an extended modular particle with kinks and bends.

The 120-Å D_{max} value determined for C0-P/A_L is ~70 Å longer than the maximum dimension of the NMR structure of C0[22] (~50 Å for a 95-residue domain), while the D_{max} value for C0C1 (D_{max}=166 Å) is ~40 Å longer than the three-domain C2C4 fragment (127 Å; Table 1). These results indicate that the ~50-residue P/A_L is highly extended (Fig. 4a). The C1C2 construct has similar R_g and D_{max} values as the larger three-domain C2C4 construct but smaller R_g and D_{max} values compared to C0C1 (Table 1), in spite of the additional 42 amino acids in C1C2. Combined, these results suggest that the m-domain between C1 and C2 is compact, as is the case for the m-domain in the previously studied mouse isoform.[41] Indeed, the human C1C2 fragment has similar domain dispositions and average R_g^c values as the mouse C1C2 (R_g^c=12.6 Å *versus* 12.4 Å[41]), but a larger

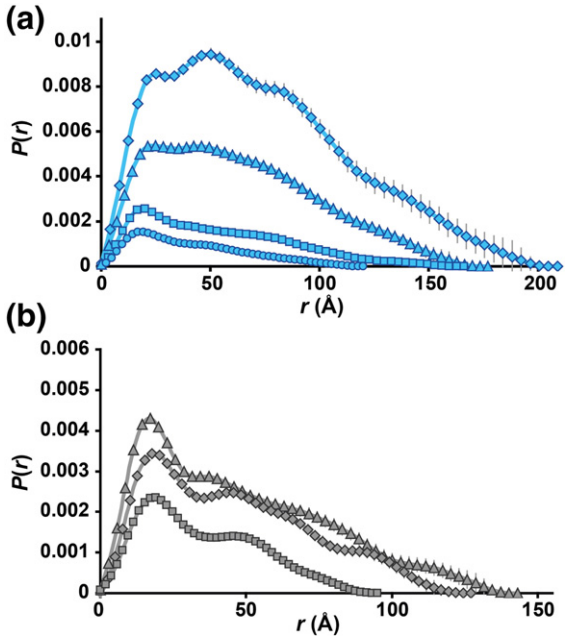

Fig. 4. Atom-pair distributions. The $P(r)$-*versus*-r profiles of each cMyBP-C construct used in this study scaled to the ratio of the squares of the molecular masses of each protein. (a) C0C4 (light blue diamonds), C0C2 (light blue triangles), C0C1 (light blue squares), and C0-P/A_L (light blue circles). (b) C1C2 (gray triangles), C2C4 (gray diamonds), and C3C4 (gray squares).

overall R_g value (39.7 Å *versus* 35.2 Å; Table 1 and Fig. 6). This larger R_g value for human C1C2 could arise from structural flexibility in the human m-domain, as identified by Ababou *et al.*[24] Flexibility in the m-domain might enable human C1C2 to sample more extended states in solution compared to the mouse homologue. The R_g value of the m-domain derived from BUNCH modeling refinements of the human homologue (see the text below) is ~17.5 Å, whereas the mouse equivalent is more compact (R_g ~14.5 Å, derived from Jeffries *et al.*[41]). However, the SAXS data show that the ~105-amino-acid human m-domain is not hyperextended like P/A$_L$ and occupies a relatively compact volume between C1 and C2, with approximate dimensions of ~35 Å × 35 Å × 45 Å (compared to the mouse m-domain of ~25 Å × 25 Å × 45 Å).

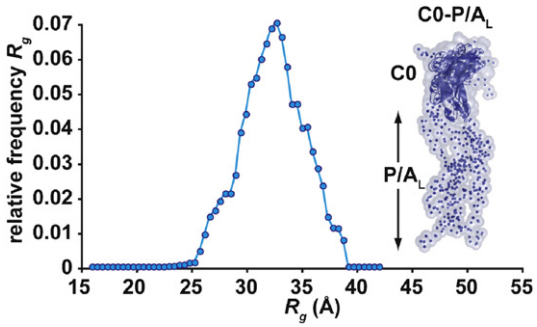

Fig. 5. Ensemble optimization results for C0-P/A$_L$. The R_g distribution of the C0-P/A$_L$ ensemble (blue circles) and a representation of the C0-P/A$_L$ ensemble that fits the scattering data. P/A$_L$ can adopt extended or more compact states in solution (for fit to the SAXS data, see Fig. 2).

The N-terminal half of human cMyBP-C has a modular architecture with built-in flexibility

A global molecular representation of nearly half the human cMyBP-C protein (C0C4) has been derived using simultaneous rigid-body refinement against multiple SAXS data sets from various truncated protein constructs. We used the program BUNCH[49] to refine the domain dispositions within C0C4 against the SAXS data to generate an average rigid-body model. To model the C0-P/A$_L$ construct, an ensemble optimization method was used as implemented in the EOM[50] program. The models were constructed using the previously determined high-resolution structures of the C0,[22] C1,[51] and C2,[25] and Phyre[52] homology models of the C3 and C4 Ig domains while also representing regions of unknown structure (P/A$_L$ and m-domain) as dummy atoms. Although we have previously suggested that the mouse m-domain has the dimensions of an Ig-like domain and may have a labile or reversible Ig-like fold,[41] current evidence suggests that the m-domain may not have an Ig-like structure;[24] therefore, we have used dummy atoms to represent the mass of the m-domain in those human constructs containing this region of cMyBP-C.

Initial BUNCH refinements of C0-P/A$_L$ did not produce good models of C0-P/A$_L$ against scattering data (as assessed by $\chi^2 = 1.80$; data not shown). Significant improvements in the fits to the data ($p < 0.01$) were achieved by modeling C0-P/A$_L$ as an ensemble of structures, using EOM[50] to produce an ensemble of structures in which P/A$_L$ adopts a range of compact and extended conformations, with D_{max} values of ~40–80 Å and with R_g distribution spanning 25–40 Å (χ^2; EOM$_{ensemble}$ = 1.13). This result suggests that P/A$_L$ is flexible in solution (Fig. 5; for fit, see Fig. 2).

To model the longer multidomain constructs, we performed BUNCH calculations multiple times to optimize fits to the SAXS data for C0C1, C1C2, C3C4, and C2C4. For C0C2 and C0C4, multiple BUNCH calculations were performed to optimize fits to multiple SAXS data sets simultaneously, including shorter constructs within C0C2 or C0C4, to improve confidence in the positioning of the domains in these longer constructs. The optimized C0C1, C1C2, C2C4, C3C4, C0C2, and C0C4 models (Fig. 6) fit their respective SAXS data sets well (Fig. 2; C0C1$_{average}$ $\chi^2 = 0.67$; C1C2$_{average}$ $\chi^2 = 1.29$; C2C4$_{average}$ $\chi^2 = 0.95$; C3C4$_{average}$ $\chi^2 = 1.34$; C0C2$_{average}$ $\chi^2 = 0.67$; C0C4$_{average}$ $\chi^2 = 0.85$), and C0C4 has an overall extended modular architecture (Fig. 6c). What is particularly striking about the C0C2 and C0C4 models is that the C0 domain can adopt multiple spatial positions at the end of P/A$_L$, consistent with the C0-P/A$_L$ EOM results and with the idea that P/A$_L$ is flexible. At the same time, the relative spatial positions of the remaining Ig domains and the m-domain in these larger constructs are consistent between multiple calculations.

NMR experiments confirm that P/A$_L$ is flexible in human C0C1

To distinguish whether the alternate spatial positions of the human C0 domain modeled from the SAXS data are the result of a bona fide flexible P/A$_L$, as opposed to modeling limitations against the relatively low information content inherent in the SAXS data, we performed NMR ^{15}N relaxation experiments on ^{15}N-labeled C0C1 using previously reported chemical shift assignments, including those spanning P/A$_L$.[45] ^{15}N relaxation parameters T_1, T_2, and heteronuclear nuclear Overhauser enhancement (NOE) were measured for each assigned residue (Fig. 7a–c). These parameters show that C0 and C1 domains are ordered, while residues within P/A$_L$ and in the N-terminal histidine tag are highly

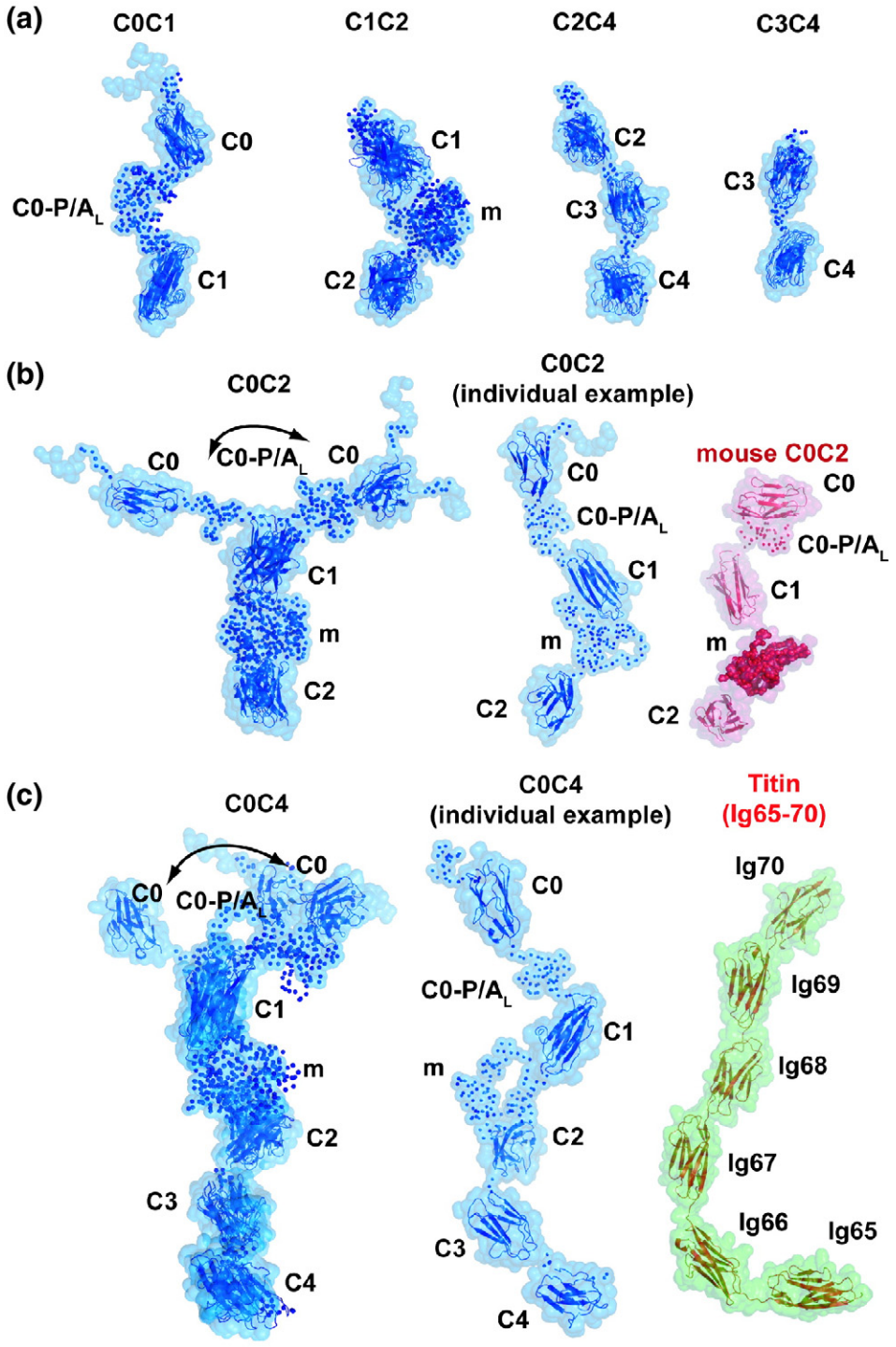

Fig. 6. Rigid-body models of the N-terminal fragments of human cMyBP-C. (a) Multiple BUNCH-refined models of C0C1, C1C2, C2C4, and C3C4. (b) Multiple BUNCH models of C0C2 showing alternate positions of C0 at the end of P/A_L and the comparison of a representative C0C2 model with mouse C0C2[41] (red). (c) The corresponding C0C4 BUNCH models and a comparison of an individual refined C0C4 model with the modular structure found in titin[53] (green; PDB ID: 3B43). P/A_L, m-domain, and other regions of unknown structure are shown as dummy atoms throughout the figure. The corresponding average fits to the data of the BUNCH-refined models are shown in Fig. 2.

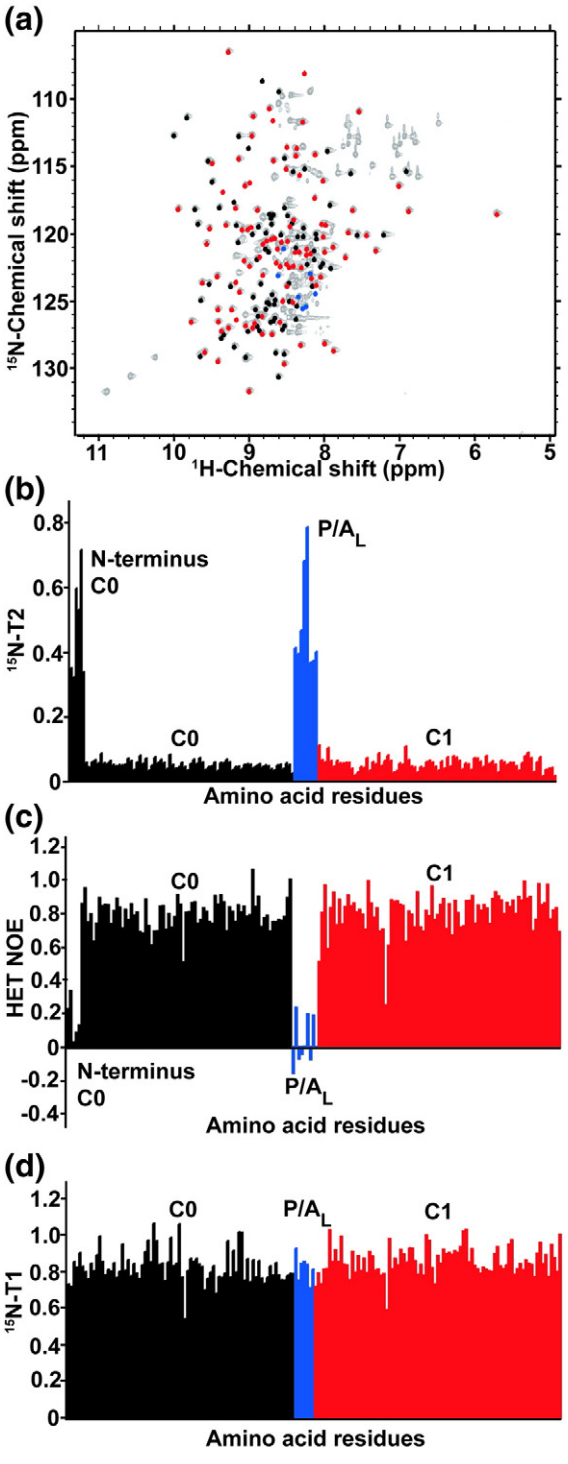

Fig. 7. NMR data from human C0C1. (a) ^{15}N–^{1}H heteronuclear single-quantum coherence spectra of human C0C1. Black, blue, and red dots correspond to the assigned chemical shifts of amino acid residues from C0, P/A$_L$, and C1, respectively. (b–d) ^{15}N T_2 relaxation times, ^{15}N heteronuclear NOE relaxation times, and ^{15}N T_1 relaxation times for assigned residues, with color coding as in (a).

flexible in solution, as indicated by the significant increase in T_2 values and a corresponding decrease in heteronuclear NOE values for these regions of the protein.

Structural features of the N-terminal modules of human cMyBP-C could influence switching between thick and thin filaments

Overall, the C0C4 fragment has a series of tandemly arranged Ig domains organized in a highly extended configuration (D_{max} ~ 210 Å; Table 1) so that the maximum dimension is sufficiently long to accommodate interactions with both thin and thick filament proteins within and/or across the interfilament space (~110–130 Å), in keeping with recent observations from electron tomography reconstructions of MyBP-C within intact muscle sarcomeres.[44] Based on experiments using skinned muscle fiber preparations and their response to exogenously added C0C2 fragments, Kulikovskaya *et al.* suggested that the N-terminal domains of cMyBP-C can also shift between binding to myosin and binding to actin to alter maximal force and calcium sensitivity.[1] We have recently reported that there is a common actin/myosin interface on C0C1 that could facilitate such a switch[45] and that the C0C1, C0C2, and C0C4 fragments reported here all have maximum dimensions sufficient to span the interfilament space to perform this task. However, having the binding interfaces and appropriate dimensions to mediate or regulate multiple interactions may be insufficient in the dynamic and highly coordinated environment of heart muscle. The N-terminal domains of cMyBP-C appear to perform a variety of functions—including the canonical phosphorylation-controlled 'tether' with myosin ΔS2[23,35–37] and the binding actin itself[19,20] to effect thin filament Ca^{2+} sensitivity[2–5,21] and to modulate ATPase rates of actin-bound myosin heads[5,6,21]—in an environment of continuous mechanical movement. Interestingly, SAXS, NMR, and crystallographic investigations of various titin domain fragments reveal that many of the structural features found in titin, which is described as a 'bidirectional molecular spring,'[53,54]are shared with the cMyBP-C models presented here: both proteins have a common extended modular Ig architecture (Fig. 6c) with degrees of conformational flexibility.[53,55] In the regions of titin composed of Ig superrepeats, pliant hinges connecting segments of tandemly arranged Ig domains confer robust elastic-like properties to the protein that enable the extension of the Ig regions under low force. At higher 'stretch,' a significantly more extensible proline-rich PEVK region unravels without compromising the fold of the Ig domains.[53,55–58] Having built-in elastic determinants allows titin to stretch and recoil during the contractile cycle. The similarity between the

modular organization of cMyBP-C and the modular organization of titin, including the incorporation of an extensible proline-rich region, supports the idea that cMyBP-C may also have flexible or elastic-like regions within its structure to enable it to adjust to the highly dynamic environment encountered within heart muscle sarcomeres. A comparison of the Kratky plots ($I(q)q^2$ *versus* q; Supplementary Fig. 3) derived from each N-terminal fragment shows that, in total, C0C4 has both highly flexible and more rigid regions dispersed within its structure. Upturns in $I(q)q^2$ with increasing $q>0.15$, which is a typical feature of the Kratky plots of proteins containing flexible regions, are evident for fragments incorporating C0-P/A_L (C0C1, C0C2, and C0C4), while the C3C4 and C2C4 modules have Kratky plots that are indicative of proteins with a more defined shape and less flexibility (i.e., $I(q)q^2$ decays toward zero at higher q values).

The inclusion of both flexible and defined modular regions within the N-terminal region of cMyBP-C may have consequences for the transverse or lateral stability of the C-zone of the A-band where cMyBP-C is situated in muscle sarcomeres. Of particular interest, the electron tomographic investigation performed by Luther *et al.* suggests that, in view of the arrangement of regularly spaced MyBP-C proteins down the length of the long axis of the A-band of muscle sarcomeres, MyBP-C spanning across the thick/thin interfilament spaces produces a regular transverse network of MyBP-C proteins linking neighboring thick and thin filaments.[44] The potential elasticity of cMyBP-C and the incorporation of regions into proteins that are flexible/extensible, as suggested from our results, may have dramatic consequences on the overall pliancy of 'Luther's net' in this region of the sarcomere.

Govada *et al.* hypothesized that P/A_L between C0 and C1 (Supplementary Fig. 2) is very extended and can consequently span the distances required to bind between myosin and actin.[51] This idea is based on sequence similarities between the proline/alanine extension of the myosin essential light chain and P/A_L,[59,60] as well as a putative actin binding site at the N-terminal end of the linker.[39] Although we have recently shown that P/A_L itself does not interact directly with actin,[45] our combined SAXS and NMR results from C0C1 in this study are consistent with Govada *et al.*'s hypothesis—that is, P/A_L from human cMyBP-C does act as an extensible element that enables the flexion of C0 (relative to the rest of cMyBP-C) to the point where C0C1, by itself, is sufficiently long (~165 Å) to cross interfilament distances.

Proline-rich regions are often disordered and allow for structural extension or bending, or have elastic structural properties.[57,58,61] Shaffer and Harris correlated the percent proline and alanine contents in the linker from different species with average heart rates.[28] They found that P/A_L appears to have evolved to reflect different heart rate requirements; in general, the faster is the heart rate, the lower is the percent proline/alanine content. A comparison of mouse and human cMyBP-C homologues (the mouse heart beats 550 times per minute, while the human heart beats around ~75 times per minute) reveals that P/A_L regions share a 46% sequence identity and that the mouse variant has significantly less proline/alanine content (~30% *versus* ~50% for the human isoform). Interestingly, there are common structural features (tandem modular arrangement and extended configuration) when comparing human C0C2 from this study with the previously reported mouse C0C2 homologue[41] (Fig. 6b). However, the human variant has larger R_g and D_{max} values than the mouse equivalent, and it appears that this difference is primarily due to the more extended P/A_L that allows the human C0 domain to sample diverse spatial positions relative to the remaining modules. Previous scattering investigations from mouse C0C2 in solution[41] or within a C0C2/actin assembly[20] show that the 'C0C1 end' of the mouse protein, especially when bound to actin, has a shorter P/A_L (~25–30 Å) compared to the human P/A_L reported here (~40–80 Å). These differences could indicate that the 'extensibility' of P/A_L has evolved to occupy different distributions of extended or compact states, depending on the different mechanical requirements of human and mouse hearts, and to potentially effect the 'switching rate' of cMyBP-C between myosin and actin of different species.

Full-length cMyBP-C self-associates in a concentration-dependent manner via interactions mediated by C-terminal domains

It has been previously shown, using analytical ultracentrifugation, that rabbit full-length skeletal MyBP-C is monomeric from ~1 mg mL^{-1} to 8 mg mL^{-1} in solutions of high ionic strength (500 mM KCl), while reversible dimerization is observed at lower ionic strength (100 mM KCl).[14] We evaluated the association state of bacterially expressed full-length human cMyBP-C using multi-angle laser light scattering (MALLS) and SAXS. MALLS data assessed at a low protein concentration (~150 μg mL^{-1}; ~1 μM) in buffers containing 200 mM NaCl show that, under these conditions, human cMyBP-C is monomeric (M_r^{MALLS} = 150 kDa; $M_r^{calculated}$ = 143 kDa; Fig. 8a). At the higher protein concentrations required for SAXS (1.2–2.2 mg mL^{-1}; ~8–15 μM), molecular mass estimates from the forward scattering intensities and from the concentration-independent method using the scattering invariant[62] provide evidence that full-length cMyBP-C exists predominantly as a dimer in solution [average $M_r^{experiment}$ = 280 kDa compared

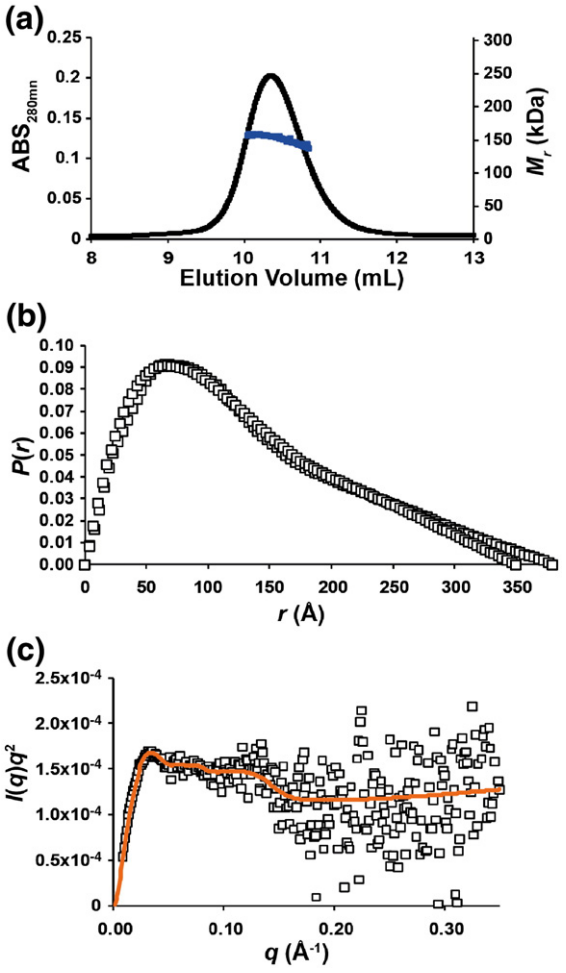

Fig. 8. Full-length human cMyBP-C analysis. (a) MALLS trace of full-length cMyBP-C and corresponding molecular mass estimate (blue line). (b) The P(*r*)-*versus*-*r* profile of the dimerized form of full-length cMyBP-C (scaled as in Fig. 4) with alternate D_{max} estimates (~350–400 Å). (c) A Kratky plot from scattering data obtained from full-length cMyBP-C dimers shows evidence of flexibility. The orange line is derived from the smoothed *I*(*q*)-*versus*-*q* profile fit to the data.

to $M_r^{calculated} = 2 \times 143 = 286$ kDa for a dimer; Table 2]. Therefore, the combined results from the MALLS and SAXS data from full-length human cMyBP-C are in keeping with the previous observations on the association state of the skeletal isoform of the protein[14] (i.e., cMyBP-C can exist as monomers or dimers in solution).

Using *P*(*r*) analysis, we conclude that the maximum dimension of the dimeric form of cMyBP-C is ~350–400 Å (Fig. 8b). The broad range and uncertainty in D_{max} reflect the difficulty of choosing an optimal D_{max} for the *P*(*r*) solution, which can occur when structures have a high degree of flexibility. Indeed, the existence of flexibility in the full-length dimer is supported by a Kratky plot of the SAXS data (Fig. 8c) and is consistent with published EM micrographs from the chicken skeletal isoform. The EM micrographs, taken from solutions prepared at a low protein concentration (~50 μg mL^{-1}; ~0.3 μM), show several open/closed V-shaped or U-shaped particles with maximum dimensions varying between 150 Å and 400 Å and with two extended arms having a 'central or paracentral region of greater flexibility' that facilitates the opening or closing of the arms.[63]

To further identify which domains are responsible for inducing full-length cMyBP-C dimerization, we analyzed the SAXS structural parameters derived from the C-terminal domain fragments of human cMyBP-C. Of note, the expression of the C5C10 fragment was not reproducible, and we therefore were limited to analyzing regions of the protein that lack C9C10 (i.e., C5C6, C5C7, and C5C8; Table 2). A careful analysis of the molecular mass estimates of the fragments indicates that C5C6 is elongated and is a monodisperse monomer in solution ($M_r^{experiment} = 26.7$ kDa; $M_r^{calculated} = 26.6$ kDa; Table 2; Supplementary Fig. 4). Surprisingly, the corresponding Kratky plot shows evidence of significantly greater conformational flexibility (Supplementary Fig. 3) compared to, for example, the analogous two-module C3C4 construct from the N-terminal region. C5C6 also has a larger R_g value (31.1 Å, cf. C3C4 $R_g = 28.2$ Å) and appears to occupy more extended states in solution (see Supplementary Fig. 4 for *P*(*r*) profile and compare D_{max} ~120 Å with 95 Å for C3C4; Tables 1 and 2). It is tempting to suggest that flexibility between the C5 and C6 domains could provide a functionally important flexible central region within the full-length protein; alternatively, the flexibility indicated in the SAXS data may arise from the disordered ~30-amino-acid cardiac-specific loop in domain C5.[64]

The inclusion of domain C7 or C7C8 in the protein constructs yielded molecular masses for scattering particles that are higher than those expected for monomers indicating some sort of self-association for C5C7 and C5C8; mass values are ~25–35% and 40–44% higher than expected for a monomer, respectively (Table 2). Based on yeast two-hybrid and surface plasmon resonance binding studies, it has been suggested that domain C5 can interact with domain C8 of a neighboring cMyBP-C molecule, leading to the idea that the C-terminal domains can trimerize and subsequently wrap around the thick filament within the sarcomere as a 'three-stranded collar[65] as opposed to lying individually and longitudinally along the length of the thick filaments as indicated by electron tomographic reconstructions.[8,44] The SAXS results presented here indicate that the C5C8 fragment does not form a stable oligomer (dimer or

Table 2. SAXS parameters for the C-terminal domain constructs and full-length human cMyBP-C

	C5C6	C5C7 sample 1	C5C7 sample 2	C5C8 sample 1	C5C8 sample 2	cMyBP-C sample 1	cMyBP-C sample 2
Structural parameters							
$I(0)$ (cm^{-1})	0.0964	0.0904	0.1981	0.1094	0.2258	0.290–0.31	0.591–0.615
R_g (Å) (from $P(r)$)	31.1	41.7	42.2	57.6	59.9	~105	~105
R_g (Å) (from Guinier; $qR_g<1$)	29.1±0.3	39.5±0.8	40.4±0.6	56.5±1.3	56.0±1	~104	~104
R_g^c (Å)	10.2±0.1	~13±0.4	~14±0.2	—	—	—	—
D_{max} (Å)	118	~145	~145	~200	~210	~350–400	350–400
Molecular mass determination							
Protein concentration (g cm^{-3})	0.00475	0.0026	0.0053	0.0022	0.00435	0.0012	0.0022
Partial specific volume (cm^3 g^{-1})	0.738	0.739	0.739	0.742	0.742	0.741	0.741
Contrast $\Delta\rho\times10^{10}$ cm^{-2})	2.898	2.873	2.873	2.837	2.837	2.851	2.851
Molecular mass (from $I(0)$) (Da)	26,700	46,500	49,900	68,800	70,400	—	—
Molecular mass from scattering invariant (Da)[a]						264,000	296,000
Calculated molecular mass for the monomer (from sequence) (Da)	26,600	37,400	37,400	48,900	48,900	143,000	143,000

[a] Uncertainty in D_{max} from $P(r)$ and a large R_g value made $I(0)$-based concentration determination potentially error prone, so we used a method developed by Williamson, based on higher-q data and independent of protein concentration, to estimate molecular mass.[61]

trimer) and that, furthermore, additional C-terminal domains (e.g., C9 and/or C10) are required for cMyBP-C to form the dimeric-sized particle observed under solution conditions for the scattering experiments.

Materials and Methods

DNA, protein expression, and purification

N-terminal domains of human cMyBP-C

A synthetic human cMyBP-C gene fragment was designed by incorporating near-optimal *Escherichia coli* codon usage for the expression of the N-terminal amino acid sequence of human cMyBP-C, specifically C0C4 (amino acid residues 1–635 derived from GenBank accession number NG_007667). The synthetic gene was manufactured by Genscript Inc. and provided in a pUC57 cloning vector. BamH1 and EcoR1 restriction endonuclease (unique sites engineered into the 5′ and 3′ ends of the synthetic gene) were used to subclone the C0C4 gene into the protein expression vector pETM, a derivative of pET28a that encodes an in-frame N-terminal histidine tag (HMHHHHHHSSGLVPRGSH). PCR primers were designed, and high-fidelity plaque-forming unit–DNA polymerase PCR was performed to isolate gene fragments encoding human C0-P/A_L (residues 1–152), C0C1 (residues 1–258), C0C2 (residues 1–451), C1C2 (residues 151–451), C2C4 (residues 358–635), and C3C4 (residues 453–635), and the isolated PCR products were subcloned back into pETM using the same BamH1 and EcoR1 restriction sites. The respective pETM-cMyBP-C fragment expression plasmids were individually transformed into *E. coli* Rosetta2(DE3) (pLysS; Invitrogen). All proteins were expressed and purified as previously described,[41] except that the affinity tag of C2C4 was removed using thrombin. The other fragments included the histidine tags due to their propensity to undergo nonspecific cleavage when attempting to remove the tag. Sample purity was assessed by SDS-PAGE (Supplementary Fig. 5).

C-terminal domains and full-length human cMyBP-C

A synthetic gene with optimized *E. coli* codons encoding full-length human cMyBP-C (GenBank accession number NG_007667) was manufactured by Genscript Inc. and provided in the expression plasmid pET15b (Novagen) containing an N-terminal histidine affinity tag. The isolation of gene fragments encoding the C-terminal domain fragments C5C6 (residues 641–863), C5C7 (residues 641–964), and C5C8 (residues 641–1065), as well as the subsequent subcloning into pETM, was performed as described above for the N-terminal domain fragments of the protein. Expression and purification of these fragments were performed as previously described.[41] A modified protein purification protocol was used to isolate full-length cMyBP-C. *E. coli* Rosetta2(DE3)pLysS cells (Invitrogen) containing the pET15b-cMyBP-C plasmid were grown at 37 °C (150 rpm) in Luria–Bertani broth (starting pH 6.2–6.5) supplemented with 33 μg mL^{-1} chloramphenicol and 60 μg mL^{-1} ampicillin to an A_{595} of 0.6, at which point isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. Cultures were immediately transferred to a 15 °C incubator (150 rpm), and the cMyBP-C protein was left to express for 16 h. All subsequent procedures were performed at 4 °C. Cells were harvested via centrifugation (5000*g*, 20 min), and pelleted cells were lysed using osmotic shock. Cell pellets were resuspended in ~10 mL of 25 mM Tris–HCl and 2.4 M sucrose (pH 7.5) containing 5 mM disodium ethylene diaminetetraacetic acid (Na_2EDTA), 1 μL mL^{-1} β-mercaptoethanol (BME), 2 mM benzamidine (BA), and 2 mM phenylmethysulfonyl fluoride (PMSF). To the resuspended high-sucrose cell slurry, we quickly added ~90 mL of no-sucrose buffer [75 mM NaCl, 5 mM Na_2EDTA, 1 μL mL^{-1} BME, 2 mM BA, 2 mM PMSF, and 25 mM Tris (pH 7.5)] to induce osmolysis. After centrifugation at ~30,000*g* for 1 h, the cleared supernatant had ammonium sulfate powder slowly added to a final

concentration of 30% wt/vol saturation, which induced cMyBP-C precipitation from solution. The insoluble cMyBP-C was harvested by centrifugation (~30,000*g* for 1 h), and the drained cMyBP-C pellet was resuspended in buffer A [200 mM NaCl, 1 μL mL^{-1} BME, 2 mM BA, 2 mM PMSF, 20 mM imidazole, and 25 mM Tris–HCl (pH 7.5)], at which point the protein went back into solution. Full-length cMyBP-C underwent Ni-affinity purification via immobilization and subsequent elution from Ni-NTA agarose using 250 mM imidazole (in buffer A). Imidazole was removed, and cMyBP-C was exchanged into buffer HA [300 mM KCl, 0.1 mM Na_2EDTA, 3 mM BME, 2 mM BA, 2 mM PMSF, 5.2 mM K_2HPO_4, and 4.8 mM KH_2PO_4 (pH 7.5)] using a HiTrap desalting column (Pharmacia). The protein then bound to ~5 mL of ceramic hydroxyapatite column (CHA type II; Bio-Rad) in HA buffer plus 1 μL mL^{-1} BME and underwent shallow elution (~20 column volumes) with a phosphate gradient using buffer HB according to the method of Starr and Offer[66] [HB: 300 mM KCl, 0.1 mM Na_2EDTA, 1 μL mL^{-1} BME, 2 mM BA, 2 mM PMSF, 340 mM K_2HPO_4, and 160 mM KH_2PO_4 (pH 7.5)]. The eluted protein (~100–150 mM phosphate) was applied to a S200 size-exclusion column (Pharmacia) equilibrated in 200 mM NaCl, 5 mM Na_2EDTA, 2 mM tris(2-carboxyethyl)phosphine (TCEP)–HCl, and 25 mM Tris (pH 8.0), and fractions containing purified full-length cMyBP-C were pooled and concentrated using a VivaSpin 100-kDa molecular mass cut-off centrifugation device. The identity of full-length cMyBP-C was confirmed using peptide mass fingerprinting from a Coomassie-stained protein band excised from a polyacrylamide gel using a QSTAR XL Hybrid Mass Spectrometer (Applied Biosystems) equipped with an oMALDI source. Peptide fragments covering 24% of the sequence of full-length cMyBP-C, including fragments from the N-terminal and C-terminal regions of the protein (up to amino acid residue 1240 out of 1274 residues in total), were positively identified.

SAXS sample preparation

Purified human C0-P/A_L, C0C1, C0C2, C0C4, C1C2, C2C4, C3C4, C5C6, C5C7, C5C8, and full-length cMyBP-C were exchanged into the following buffers and yielded samples with the following concentrations: (1) C0-P/A_L (8.17 mg mL^{-1}) and C0C1 (5.55 mg mL^{-1}): 100 mM NaCl, 25 mM Na_2SO_4, 14 mM BME, 1 mM ethylenediaminetetraacetic acid, and 15 mM 4-morpholineethanesulfonic acid (pH 6.0); (2) C0C2 (3.77 mg mL^{-1}) and C1C2 (3.76 mg mL^{-1}): 350 mM NaCl, 2 mM TCEP, and 25 mM Tris (pH 7.5); (3) C0C4 (2.92 mg mL^{-1}): 400 mM NaCl, 2 mM TCEP, and 25 mM Tris (pH 7.5); (4) C2C4 (6.72 mg mL^{-1}), C3C4 (7.68 mg mL^{-1}), C5C6 (4.75 mg mL^{-1}), C5C7 (2.6 and 5.3 mg mL^{-1}), C5C8 (2.16 and 4.3 mg mL^{-1}), and full-length cMyBP-C (1.2 mg mL^{-1} and 2.2 mg mL^{-1}): 200 mM NaCl, 2 mM TCEP, and 25 mM Tris (pH 7.5). Concentrations were determined after dialysis by A_{280} measurements and using the following extinction coefficients (expressed as $E_{0.1\%}$; in g L^{-1}) calculated from the primary amino acid sequence with ProtParam: C0-P/A_L, 0.500 g L^{-1}; C0C1, 0.826 g L^{-1}; C1C2, 0.952 g L^{-1}; C0C2, 0.839 g L^{-1}; C0C4, 0.840 g L^{-1}; C2C4, 0.871 g L^{-1}; C3C4, 0.765 g L^{-1}; C5C6, 1.162 g L^{-1}; C5C7, 1.202 g L^{-1}; C5C8, 1.061 g L^{-1}; full-length cMyBP-C, 1.054 g L^{-1}.

SAXS data acquisition

SAXS data were measured as described previously[41] at 10 °C with a SAXSess (Anton Paar) line collimation instrument (10 mm) equipped with a charge-coupled device detector. Data were reduced to $I(q)$ *versus* q ($q=(4\pi\sin\theta)/\lambda$; 2θ is the scattering angle; $\lambda=1.54$ Å CuK_α) using the program SAXSquant 2.0, which includes corrections for sample absorbance and detector sensitivity. Scattering from the matched solvents (from end-point dialysates or size-exclusion chromatography protein-free eluates) was subtracted from the scattering profiles of each sample to obtain scattering from the proteins alone. The molecular mass of the cMyBP-C variants was assessed from the forward (zero-angle) scattering intensity ($I(0)$) using the method of Orthaber *et al.*[67] by placing the data on an absolute scale using scattering from water. Values for contrast and partial specific volumes were calculated using the MULCh suite of analysis tools.[68] Concentration-independent estimates of the molecular mass of full-length cMyBP-C were also calculated using the method of Fischer *et al.*[62] Both Guinier analysis and modified Guinier analysis of the data[46] were performed using PRIMUS[69] to determine the radius of gyration (R_g), the radius of gyration of cross section (R_g^c), and $I(0)$. The program GIFT,[70] which accounts for the line-source geometry of the instrument, was used to calculate the probable distribution of distances between atom pairs ($P(r)$ profiles) within each protein construct from which the D_{max}, R_g, and $I(0)$ values were determined. The experimental data were multiplied (point by point) by the ratio of the model to the smeared model $I(q)$-*versus*-q profiles from the optimal $P(r)$ fit in order to obtain a scattering profile corrected for slit-smearing effects arising from the instrument geometry. All atomic modelings used these 'desmeared' data as input. This desmearing approach is subject to error if D_{max} is not well determined, so desmearing was evaluated by a comparison with desmeared data obtained using the Lake algorithm[71] in SAXSQuant, which does not depend on the estimation of D_{max} but amplifies pixel-to-pixel noise in the data.

SAXS data modeling

The crystal or NMR structures of human C0 [Protein Data Bank (PDB) ID: 2K1M[22]], C1 (PDB ID: 2V6H[51]), and C2 (PDB ID: 1PD6[25]) were used with the program BUNCH[49] to refine the positions of the relevant domains against the scattering data for C0-P/A_L, C0C1, C1C2, and C0C2. In these refinements, BUNCH generates dummy atoms to account for those parts of the sequence of unknown structure, specifically P/A_L and m-domain, and refines their shape/mass distribution in space. The program Phyre[52] was employed to generate homology models of the C3 and C4 domains that were incorporated into C0C4, C2C4, and C3C4 BUNCH rigid-body modeling refinements. The Phyre-generated C3 and C4 homology models have very-high-scoring predictions against the titin Ig69 and Ig70 modules, as well as a high sequence homology to known structures of the third Ig-like domain from slow-type MyBP-C (PDB ID: 2EDK, which has a 64% sequence identity with C3 from human cMyBP-C) or a fast-type MyBP-C Ig-like domain (PDB ID: 2DLT, which

shares a 65% sequence identity with the human C4 domain). BUNCH refinements were run multiple times in order to obtain an ensemble of structures representing the domain organization/average mass distributions within each construct. To improve confidence in the final average positions of the modules within the longer C0C2 and C0C4 constructs, we performed BUNCH calculations that simultaneously fit multiple data sets (for C0C2, SAXS profiles from C0C1, C1C2, and C0C2 were used; for C0C4, the C0C1, C0C2, C0C4, C2C4, and C3C4 data sets were used). The goodness-of-fit of the final models to the SAXS data was assessed using χ^2 values from CRYSOL.[72] An EOM[50] was used to model human C0-P/A_L, as these data could not be fitted using a single average model structure. Unlike C0-P/A_L, which underwent significant improvement in the fit to the SAXS data, using EOM to model the remaining constructs is not warranted, as the low χ^2 values obtained for the fit of the BUNCH models against the remaining data sets indicate that further modeling using EOM would introduce more parameters into the data modeling process than is justifiable within the error estimates in the data.

NMR sample preparation, data acquisition, and analysis

^{15}N-labeled human C0C1 was prepared and purified using the same bacterial overexpression and purification procedures described by Lu *et al.*[45] Protein spectra, sample preparation, NMR data acquisition, analysis, and chemical shift assignments for C0C1 were reported by Lu *et al.*[45] ^{15}N T_1, T_2, and heteronuclear NOE values were measured on [^{15}N]C0C1 using modified Bruker pulse programs obtained from Dr. Paul Gooley (University of Melbourne) on a Bruker AVIII 800-MHz NMR spectrometer equipped with a triple-resonance TCI cryoprobe. The pulse programs are based on hsqct1etf3gpsi3d and hsqct2etf3gpsi3d but give less phase errors and better water suppression. Relaxation delays of 5 ms, 12 ms, 60 ms, 100 ms, 150 ms, 300 ms, 600 ms, 800 ms, 1000 ms, 1400 ms, and 1500 ms were used in the T_1 experiments, while values of 4.8 ms, 9.6 ms, 16 ms, 24 ms, 32 ms, 40 ms, 48 ms, 80 ms, and 320 ms were used to measure T_2 values. Recycle delays of 3 s were used in both experiments. Peaks were integrated, and T_1 and T_2 data were fitted to two-parameter exponentials using a module within the program SPARKY.

MALLS analysis of full-length cMyBP-C

Full-length cMyBP-C at 1.2 mg mL^{-1} [200 μL in 200 mM NaCl, 1 mM TCEP, and 25 mM Tris–HCl (pH 8.0)] was loaded into a 900-μL loop and injected onto a 10/30 Superdex 200 SEC column (GE Healthcare) connected to an AKTA FPLC system feeding into a Wyatt technology miniDAWN light-scattering unit and an Optilab DSP refractometer. To determine the relative molecular masses of the full-length protein, the system was calibrated to an absolute scale using intrinsic Rayleigh scattering in toluene. A uniform refractive index to a concentration gradient (dn/dc) of 0.19 mL g^{-1} was assumed for all proteins. The concentration estimate at the point of analysis was estimated at 100–200 μg mL^{-1} based on the absorbance at 280 nm.

Acknowledgements

We thank Dr. Ben Crossett for performing mass spectrometry using the Australian Proteome Analysis Facility established under the Australian Government's Major National Facilities program.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.10.029

References

1. Kulikovskaya, I., McClellan, G., Flavigny, J., Carrier, L. & Winegrad, S. (2003). Effect of MyBP-C binding to actin on contractility in heart muscle. *J. Gen. Physiol.* **122**, 761–774.
2. Razumova, M. V., Bezold, K. L., Tu, A. Y., Regnier, M. & Harris, S. P. (2008). Contribution of the myosin binding protein C motif to functional effects in permeabilized rat trabeculae. *J. Gen. Physiol.* **132**, 575–585.
3. Shaffer, J. F., Razumova, M. V., Tu, A., Regnier, M. & Harris, S. P. (2008). The cardiac myosin binding protein-C motif and C1 domain activate actomyosin motility independent of Ca^{2+}. *Biophys. J.* **94**, 293b.
4. Shaffer, J. F., Razumova, M. V., Tu, A. Y., Regnier, M. & Harris, S. P. (2007). Myosin S2 is not required for effects of myosin binding protein-C on motility. *FEBS Lett.* **581**, 1501–1504.
5. Shaffer, J. F., Wong, P., Bezold, K. L. & Harris, S. P. (2010). Functional differences between the N-terminal domains of mouse and human myosin binding protein-C. *J. Biomed. Biotechnol.* **2010**, 789–798.
6. Shchepkin, D. V., Kopylova, G. V., Nikitina, L. V., Katsnelson, L. B. & Bershitsky, S. Y. (2010). Effects of cardiac myosin binding protein-C on the regulation of interaction of cardiac myosin with thin filament in an *in vitro* motility assay. *Biochem. Biophys. Res. Commun.* **401**, 159–163.
7. Craig, R. & Offer, G. (1976). The location of C-protein in rabbit skeletal muscle. *Proc. R. Soc. London Ser. B*, **192**, 451–461.
8. Zoghbi, M. E., Woodhead, J. L., Moss, R. L. & Craig, R. (2008). Three-dimensional structure of vertebrate cardiac muscle myosin filaments. *Proc. Natl Acad. Sci. USA*, **105**, 2386–2390.
9. Bennett, P., Craig, R., Starr, R. & Offer, G. (1986). The ultrastructural location of C-protein, X-protein and H-protein in rabbit muscle. *J. Muscle Res. Cell Motil.* **7**, 550–567.
10. Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T. & Bild, D. E. (1995). Prevalence of hypertrophic cardiomyopathy in a general population of young adults. *Circulation*, **92**, 785–789.
11. Richard, P., Charron, P., Carrier, L., Ledeuil, C., Cheav, T., Pichereau, C. *et al.* (2003). Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation*, **107**, 2227–2232.

12. Tsoutsman, T., Bagnall, R. D. & Semsarian, C. (2008). Impact of multiple gene mutations in determining the severity of cardiomyopathy and heart failure. *Clin. Exp. Pharmacol. Physiol.* **35**, 1349–1357.
13. Girolami, F., Ho, C. Y., Semsarian, C., Baldi, M., Will, M. L., Baldini, K. *et al.* (2010). Clinical features and outcome of hypertrophic cardiomyopathy associated with triple sarcomere protein gene mutations. *J. Am. Coll. Cardiol.* **55**, 1444–1453.
14. Offer, G., Moos, C. & Starr, R. (1973). A new protein of the thick filaments of vertebrate skeletal myofibrils. Extractions, purification and characterization. *J. Mol. Biol.* **74**, 653–676.
15. Yamamoto, K. & Moos, C. (1983). The C-proteins of rabbit red, white, and cardiac muscles. *J. Biol. Chem.* **258**, 8395–8401.
16. Freiburg, A. & Gautel, M. (1996). A molecular map of the interactions between titin and myosin-binding protein C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. *Eur. J. Biochem.* **235**, 317–323.
17. Okagaki, T., Weber, F. E., Fischman, D. A., Vaughan, K. T., Mikawa, T. & Reinach, F. C. (1993). The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, immunoglobulin C2 motif. *J. Cell Biol.* **123**, 619–626.
18. Flashman, E., Watkins, H. & Redwood, C. (2007). Localization of the binding site of the C-terminal domain of cardiac myosin-binding protein-C on the myosin rod. *Biochem. J*, **401**, 97–102.
19. Kensler, R. W., Shaffer, J. F. & Harris, S. P. (2011). Binding of the N-terminal fragment C0–C2 of cardiac MyBP-C to cardiac F-actin. *J. Struct. Biol.* **174**, 44–51.
20. Whitten, A. E., Jeffries, C. M., Harris, S. P. & Trewhella, J. (2008). Cardiac myosin-binding protein C decorates F-actin: implications for cardiac function. *Proc. Natl Acad. Sci. USA*, **105**, 18360–18365.
21. Razumova, M. V., Shaffer, J. F., Tu, A. Y., Flint, G. V., Regnier, M. & Harris, S. P. (2006). Effects of the N-terminal domains of myosin binding protein-C in an *in vitro* motility assay: evidence for long-lived cross-bridges. *J. Biol. Chem.* **281**, 35846–35854.
22. Ratti, J., Rostkova, E., Gautel, M. & Pfuhl, M. (2011). Structure and interactions of myosin-binding protein C domain C0: cardiac-specific regulation of myosin at its neck? *J. Biol. Chem.* **286**, 12650–12658.
23. Kunst, G., Kress, K. R., Gruen, M., Uttenweiler, D., Gautel, M. & Fink, R. H. (2000). Myosin binding protein C, a phosphorylation-dependent force regulator in muscle that controls the attachment of myosin heads by its interaction with myosin S2. *Circ. Res.* **86**, 51–58.
24. Ababou, A., Rostkova, E., Mistry, S., Le Masurier, C., Gautel, M. & Pfuhl, M. (2008). Myosin binding protein C positioned to play a key role in regulation of muscle contraction: structure and interactions of domain C1. *J. Mol. Biol.* **384**, 615–630.
25. Ababou, A., Gautel, M. & Pfuhl, M. (2007). Dissecting the N-terminal myosin binding site of human cardiac myosin-binding protein C. Structure and myosin binding of domain C2. *J. Biol. Chem.* **282**, 9204–9215.
26. Oakley, C. E., Hambly, B. D., Curmi, P. M. & Brown, L. J. (2004). Myosin binding protein C: structural abnormalities in familial hypertrophic cardiomyopathy. *Cell Res.* **14**, 95–110.
27. Palmer, B. M., Georgakopoulos, D., Janssen, P. M., Wang, Y., Alpert, N. R., Belardi, D. F. *et al.* (2004). Role of cardiac myosin binding protein C in sustaining left ventricular systolic stiffening. *Circ. Res.* **94**, 1249–1255.
28. Shaffer, J. F. & Harris, S. P. (2009). Species-specific differences in the Pro-Ala rich region of cardiac myosin binding protein-C. *J. Muscle Res. Cell Motil.* **30**, 303–306.
29. Mohamed, A. S., Dignam, J. D. & Schlender, K. K. (1998). Cardiac myosin-binding protein C (MyBP-C): identification of protein kinase A and protein kinase C phosphorylation sites. *Arch. Biochem. Biophys.* **358**, 313–319.
30. Hartzell, H. C. & Glass, D. B. (1984). Phosphorylation of purified cardiac muscle C-protein by purified cAMP-dependent and endogenous Ca^{2+}-calmodulin-dependent protein kinases. *J. Biol. Chem.* **259**, 15587–15596.
31. Schlender, K. K. & Bean, L. J. (1991). Phosphorylation of chicken cardiac C-protein by calcium/calmodulin-dependent protein kinase II. *J. Biol. Chem.* **266**, 2811–2817.
32. Lim, M. S. & Walsh, M. P. (1986). Phosphorylation of skeletal and cardiac muscle C-proteins by the catalytic subunit of cAMP-dependent protein kinase. *Biochem. Cell Biol.* **64**, 622–630.
33. Venema, R. C. & Kuo, J. F. (1993). Protein kinase C-mediated phosphorylation of troponin I and C-protein in isolated myocardial cells is associated with inhibition of myofibrillar actomyosin MgATPase. *J. Biol. Chem.* **268**, 2705–2711.
34. Hartzell, H. C. & Titus, L. (1982). Effects of cholinergic and adrenergic agonists on phosphorylation of a 165,000-dalton myofibrillar protein in intact cardiac muscle. *J. Biol. Chem.* **257**, 2111–2120.
35. Gruen, M., Prinz, H. & Gautel, M. (1999). cAPK-phosphorylation controls the interaction of the regulatory domain of cardiac myosin binding protein C with myosin-S2 in an on–off fashion. *FEBS Lett.* **453**, 254–259.
36. Gruen, M. & Gautel, M. (1999). Mutations in beta-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C. *J. Mol. Biol.* **286**, 933–949.
37. Gautel, M., Zuffardi, O., Freiburg, A. & Labeit, S. (1995). Phosphorylation switches specific for the cardiac isoform of myosin binding protein-C: a modulator of cardiac contraction? *EMBO J.* **14**, 1952–1960.
38. Weisberg, A. & Winegrad, S. (1996). Alteration of myosin cross bridges by phosphorylation of myosin-binding protein C in cardiac muscle. *Proc. Natl Acad. Sci. USA*, **93**, 8999–9003.
39. Squire, J. M., Luther, P. K. & Knupp, C. (2003). Structural evidence for the interaction of C-protein (MyBP-C) with actin and sequence identification of a possible actin-binding domain. *J. Mol. Biol.* **331**, 713–724.
40. Squire, J. M., Roessle, M. & Knupp, C. (2004). New X-ray diffraction observations on vertebrate muscle: organisation of C-protein (MyBP-C) and troponin and evidence for unknown structures in the vertebrate A-band. *J. Mol. Biol.* **343**, 1345–1363.
41. Jeffries, C. M., Whitten, A. E., Harris, S. P. & Trewhella, J. (2008). Small-angle X-ray scattering

reveals the N-terminal domain organization of cardiac myosin binding protein C. *J. Mol. Biol.* **377**, 1186–1199.
42. Mun, J. Y., Gulick, J., Robbins, J., Woodhead, J., Lehman, W. & Craig, R. (2011). Electron microscopy and 3D reconstruction of F-actin decorated with cardiac myosin-binding protein C (cMyBP-C). *J. Mol. Biol.* **410**, 214–225.
43. Orlova, A., Galkin, V. E., Jeffries, C. M., Egelman, E. H. & Trewhella, J. (2011). The N-terminal domains of myosin binding protein C can bind polymorphically to F-actin. *J. Mol. Biol.* **412**, 379–386.
44. Luther, P. K., Winkler, H., Taylor, K., Zoghbi, M. E., Craig, R., Padron, R. *et al.* (2011). Direct visualization of myosin-binding protein C bridging myosin and actin filaments in intact muscle. *Proc. Natl Acad. Sci. USA*, **108**, 11423–11428.
45. Lu, Y., Kwan, A. H., Trewhella, J. & Jeffries, C. M. (2011). The C0C1 fragment of human cardiac myosin binding protein C has common binding determinants for both actin and myosin. *J. Mol. Biol.* doi:10.1016/j.jmb.2011.09.026.
46. Guinier, A. (1938). The diffusion of X-rays under the extremely weak angles applied to the study of fine particles and colloidal suspension. *C. R. Seances Acad. Sci.* **206**, 1374–1376.
47. Hjelm, R. P. (1985). The small-angle approximation of X-ray and neutron scatter from rigid rods of non-uniform cross section and finite length. *J. Appl. Crystallogr.* **18**, 452–460.
48. Porod, G. (1982). General theory. In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), Academic Press, London, UK.
49. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250.
50. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664.
51. Govada, L., Carpenter, L., da Fonseca, P. C., Helliwell, J. R., Rizkallah, P., Flashman, E. *et al.* (2008). Crystal structure of the C1 domain of cardiac myosin binding protein-C: implications for hypertrophic cardiomyopathy. *J. Mol. Biol.* **378**, 387–397.
52. Kelly, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371.
53. von Castelmur, E., Marino, M., Svergun, D. I., Kreplak, L., Ucurum-Fotiadis, Z., Konarev, P. V. *et al.* (2008). A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. *Proc. Natl Acad. Sci. USA*, **105**, 1186–1191.
54. Granzier, H. L. & Labeit, S. (2004). The giant protein titin: a major player in myocardial mechanics, signaling, and disease. *Circ. Res.* **94**, 284–295.
55. Improta, S., Krueger, J. K., Gautel, M., Atkinson, R. A., Lefevre, J. F., Moulton, S. *et al.* (1998). The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. *J. Mol. Biol.* **284**, 761–777.
56. Granzier, H. & Labeit, S. (2002). Cardiac titin: an adjustable multi-functional spring. *J. Physiol.* **541**, 335–342.
57. Trombitas, K., Greaser, M., Labeit, S., Jin, J. P., Kellermayer, M., Helmes, M. & Granzier, H. (1998). Titin extensibility *in situ*: entropic elasticity of permanently folded and permanently unfolded molecular segments. *J. Cell Biol.* **140**, 853–859.
58. Trombitas, K., Greaser, M., French, G. & Granzier, H. (1998). PEVK extension of human soleus muscle titin revealed by immunolabeling with the anti-titin antibody 9D10. *J. Struct. Biol.* **122**, 188–196.
59. Timson, D. J. & Trayer, I. P. (1997). The role of the proline-rich region in A1-type myosin essential light chains: implications for information transmission in the actomyosin complex. *FEBS Lett.* **400**, 31–36.
60. Aydt, E. M., Wolff, G. & Morano, I. (2007). Molecular modeling of the myosin-S1(A1) isoform. *J. Struct. Biol.* **159**, 158–163.
61. Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260.
62. Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109.
63. Swan, R. C. & Fischman, D. A. (1986). Electron microscopy of C-protein molecules from chicken skeletal muscle. *J. Muscle Res. Cell Motil.* **7**, 160–166.
64. Idowu, S. M., Gautel, M., Perkins, S. J. & Pfuhl, M. (2003). Structure, stability and dynamics of the central domain of cardiac myosin binding protein C (MyBP-C): implications for multidomain assembly and causes for cardiomyopathy. *J. Mol. Biol.* **329**, 745–761.
65. Moolman-Smook, J., Flashman, E., de Lange, W., Li, Z., Corfield, V., Redwood, C. & Watkins, H. (2002). Identification of novel interactions between domains of myosin binding protein-C that are modulated by hypertrophic cardiomyopathy missense mutations. *Circ. Res.* **91**, 704–711.
66. Starr, R. & Offer, G. (1982). Preparation of C-protein, H-protein, X-protein, and phosphofructokinase. *Methods Enzymol.* **85**, 130–138.
67. Orthaber, D., Bergmann, A. & Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* **33**, 218–225.
68. Whitten, A. E., Cai, S. Z. & Trewhella, J. (2008). MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J. Appl. Crystallogr.* **41**, 222–226.
69. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
70. Bergmann, A. & Glatter, O. (2000). Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing. *J. Appl. Crystallogr.* **33**, 1212–1216.
71. Lake, J. A. (1967). An iterative method of slit-correcting small angle X-ray data. *Acta Crystallogr.* **23**, 191–194.
72. Svergun, D., Barberato, C. & Koch, M. H. (1995). CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.