

## Stability and Kinetic Properties of C5-Domain from Myosin Binding Protein C and its Mutants

Carlo Guardiani,\* Fabio Cecconi,<sup>†</sup> and Roberto Livi<sup>‡</sup>

\*Centro Interdipartimentale per lo Studio delle Dinamiche Complesse, INFN Sezione di Firenze, Italy; <sup>†</sup>INFM-CNR, Center for Statistical Mechanics and Complexity Istituto dei Sistemi Complessi (ISC-CNR), Rome, Italy; and <sup>‡</sup>Dipartimento di Fisica Università di Firenze, Centro Interdipartimentale per lo Studio delle Dinamiche Complesse, INFN Sezione di Firenze e INFM UdR Firenze, Italy

**ABSTRACT** The impact of three mutations of domain C5 from myosin binding protein C, correlated to Familial Hypertrophic Cardiomyopathy, has been assessed through molecular dynamics simulations based on a native centric protein modeling. The severity of the phenotype correlates with the shift in unfolding temperature determined by the mutations. A contact probability analysis reveals a folding process of the C5 domain originating in the region of DE and FG loops and propagating toward the area proximal to CD and EF loops. This suggests that mutation effects gain relevance in the proximity to the area where folding originates. The scenario is also confirmed by the analysis of the kinetics of 27 test mutations evenly distributed throughout the entire C5 domain.

### INTRODUCTION

Familial Hypertrophic Cardiomyopathy (FHC) is a cardiac disorder characterized by left ventricular hypertrophy. The prevalence of the disease in the population is 0.2% and represents the most common cause of sudden cardiac death in young competitive athletes (1). FHC is clinically heterogeneous with different forms caused by mutations in 10 genes encoding cardiac sarcomeric proteins, the most important of which are the  $\beta$ -myosin heavy chain, the cardiac troponin T, titin, and cardiac myosin binding protein C (MyBP-C) (2). MyBP-C mutations are found in 20–45% of the FHC patients (3), so that alterations on this protein represent the second-most common cause of the disease. Both nonsense (premature termination of translation) and missense (amino-acid substitution) mutations have been found to correlate with the disease. Whereas nonsense mutations on MyBP-C gene result in a mild phenotype (3), a number of missense mutations lead to a severe phenotype and the precise mechanism through which they cause the pathology is still unknown (4). MyBP-C is a linear sequence of 11 IgI-like and fibronectin-like domains, referred to as C0–C10, working as potential regulators of cardiac contractility (1). A structural arrangement has been proposed by Moolman-Smook (5,6) whereby three MyBP-C molecules form a collar around the thick filament (Fig. 1). In particular, domains C5–C10 in each molecule participate in formation of the collar while domains C0–C4 protrude into the interfilament space and interact with subfragment 2 of myosin acting as a brake for muscular contraction (7). The collar is stabilized by specific interactions between domains C5–C8 of a molecule and domains C7–C10 of the neighboring one. Mutations on MyBP-C might either prevent the C0–C4 region

from interacting with the S2 fragment (hypercontractility) or force a carboxy-truncated mutant to permanently interact with S2 (hypocontractility) (5).

In this work we focus on domain C5, taking advantage of the availability of structural and thermodynamic experimental data collected by Idowu et al. (8). This domain, belonging to the Immunoglobulin superfamily, is characterized by a  $\beta$ -sandwich structure with two twisted  $\beta$ -sheets closely packed against each other. The first  $\beta$ -sheet is formed by strands C, F, G, and A', while the second  $\beta$ -sheet comprises strands B, D, and E (see Fig. 2). A specific feature of the cardiac MyBP-C isoform is the presence of two long insertions not present in the fast and slow skeletal isoforms. The first insertion is 10-residues-long and is located in the linker between the C4 and C5 domains. Thermodynamic measurements on the  $\Delta 1-7$  deletion mutant (8), however, showed that this protein segment contributes significantly to the stability of domain C5 so that it cannot be regarded as a simple linking element between two subsequent units of MyBP-C. The second insertion is 28 residues in length and it resides in the CD loop (8) that, owing to its high mobility, destabilizes the protein lowering the folding temperature as compared to other Ig-domains (8). This loop is suspected to form an SH3 domain recognition sequence presumably binding to the CaM-II-like kinase that copurifies with MyBP-C (9).

Three FHC-causing mutations have been identified on C5 domain: Asn<sup>755</sup>Lys, Arg<sup>654</sup>His, and Arg<sup>668</sup>His. The first one lies on the FG-loop and leads to a significant destabilization of the protein yielding a severe phenotype (6,8). Conversely, the Arg<sup>654</sup>His and Arg<sup>668</sup>His mutations, causing a much milder phenotype, are not particularly destabilizing and their pathogenic role is possibly related to the impairment of binding with specific ligands. Residue Arg<sup>654</sup>, in fact, is located on the negatively charged CFGA' surface, and it may regulate the specificity of the binding of positively charged substrates

Submitted February 13, 2007, and accepted for publication September 20, 2007.

Address reprint requests to Fabio Cecconi, Tel.: 39-06-4993-7453; E-mail: cecconif@romal.infn.it.

Editor: Ron Elber.

© 2008 by the Biophysical Society  
0006-3495/08/02/1403/09 \$2.00

doi: 10.1529/biophysj.107.106625

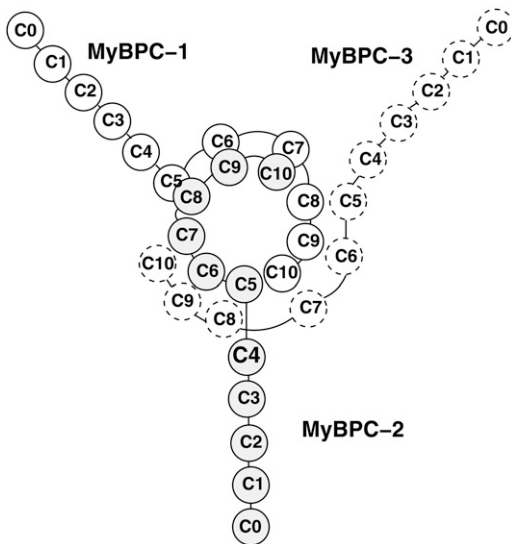


FIGURE 1 Moolman-Smook model of MyBP-C arrangement in the sarcomere: the trimeric collar surrounding the thick filament is stabilized by interactions between domains C5–C8 and C7–C10.

(4,6,8). A role in ligand binding is also postulated for Arg<sup>668</sup> (10).

This work explores through MD simulations the role of the above-mentioned FHC-causing mutations within the framework of a coarse-grained native centric model. Native centric models, or simply Gō-models, minimize the frustration of the free energy hypersurface, thus reducing the risk for the protein to get trapped in a metastable minimum. This approach is particularly meaningful when the folding process is

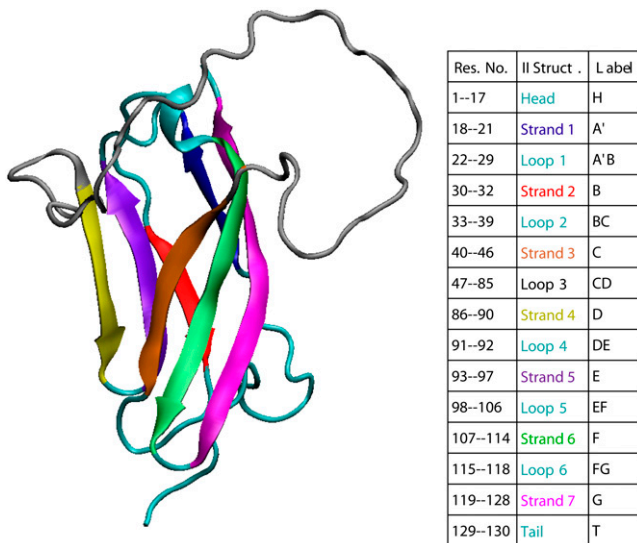


FIGURE 2 The secondary structure elements of domain C5 of MyBP-C. The residues have been renumbered 1–130. The  $\beta$ -strands are highlighted in color and labeled according to the table (figure produced by the VMD software (30)).

strongly driven by the topological properties of the native state (11–17). This is the case of domain C5 of MyBP-C, belonging to the Ig-superfamily where both transition and native states are stabilized by the same contacts dictated by protein topology (18) so that the members of this group share similar folding pathways.

We consider a heavy-map Gō model where native contacts are identified on the basis of the steric hindrance of side chains to complete the topological description with some specific chemical features. Since we address the problem of discriminating the effects of mutations, we also introduced some amount of heterogeneity in the energetic couplings of the Gō force field. This approach is particularly suitable for C5 domain where pathological mutations Arg<sup>654</sup>His and Arg<sup>668</sup>His are modeled through the removal of the same number of native contacts. The relevance of this approach for a reliable mutation analysis was suggested, for example, in Matysiak and Clementi (19), where available experimental data on single mutations of S6 ribosomal protein and its circular permuteds, were reproduced with correlation coefficients larger than 0.9.

## METHODS

In this article, we consider a variant of Gō-like force field proposed by Clementi et al. (13). The polypeptide chain is represented only through  $C_{\alpha}$  carbon positions. The interaction between residues is defined by the potential energy

$$\begin{aligned}
 V_{\text{tot}} = & \sum_{i=1}^{N-1} \frac{k_h}{2} (r_{i,i+1} - R_{i,i+1})^2 + \sum_{i=2}^{N-1} \frac{k_{\theta}}{2} (\theta_i - \theta_i^0)^2 \\
 & + \sum_{i=3}^{N-2} k_{\phi}^{(1)} [1 - \cos(\phi_i - \phi_i^0)] + k_{\phi}^{(3)} [1 - \cos 3(\phi_i - \phi_i^0)] \\
 & + \sum_{i,j>i+3} V_{\text{nb}}(r_{ij}), \quad (1)
 \end{aligned}$$

where the nonbonded interaction potential  $V_{\text{nb}}$  reads

$$V_{\text{nb}}(r_{ij}) = \begin{cases} \epsilon_{ij} \left[ 5 \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 6 \left( \frac{R_{ij}}{r_{ij}} \right)^{10} \right] & \text{if } i - j \text{ is native} \\ 5\epsilon_r \left( \frac{\sigma}{r_{ij}} \right)^{12} & \text{if } i - j \text{ is not native.} \end{cases} \quad (2)$$

In the above equations,  $r_{ij}$  is the distance between residues  $i$  and  $j$ ;  $\theta_i$  is the bending angle identified by the three consecutive  $C_{\alpha}$  values  $i - 1$ ,  $i$ , and  $i + 1$ ; and  $\phi_i$  is the dihedral angle defined by the two planes formed by four consecutive  $C_{\alpha}$  values  $i - 2$ ,  $i - 1$ ,  $i$ , and  $i + 1$ . The symbols with the superscript 0 and  $R_{ij}$  are the corresponding quantities in the native conformation. The force-field parameters are proportional to the energy scale  $\epsilon_0 = 0.3$  Kcal/mol, such that  $k_h = 1000\epsilon_0/r_0^2$  ( $r_0 = 3.8$  Å),  $k_{\theta} = 20\epsilon_0$ ,  $k_{\phi}^{(1)} = \epsilon_0$ , and  $k_{\phi}^{(3)} = 0.5\epsilon_0$ . The parameters of the repulsive Lennard-Jones terms between nonnative contacts are chosen as follows:  $\sigma = 5.0$  Å,  $\epsilon_r = 2/3\epsilon_0$ . Two residues  $i$  and  $j$  interact attractively whenever their side chains have at least a pair of heavy atoms closer than a distance cutoff  $R_c = 5$  Å. The condition selects 227 residue-residue contacts on the PDB structure 1GXE. The strength of attractive native interactions  $\epsilon_{ij}$  is weighted according to the number of atomic contacts,  $\epsilon_{ij} = \epsilon_0(1 + n_{ij}/n_{\text{max}})$ , where  $n_{ij}$  is the number of atomic contacts between residues  $i$  and  $j$  in the PDB structure normalized to

the maximal value of this quantity  $n_{\max} = 43$  found for the couple Lys<sup>45</sup>-Tyr<sup>109</sup>. The heterogeneity in residue contact energies was introduced in this work through their rescaling according to the number of atomic overlaps. This strategy is supported by the experimental and theoretical findings. In particular, the experimental mutation analysis on barnase by Serrano et al. (20) showed a nontrivial correlation between the destabilization induced by the mutation and the number of methyl or methylene side groups surrounding the deleted group. Moreover, theoretical arguments in agreement with our simple approach can be found in the literature (21–23) where linear correlation is found between pairwise sum of the buried surface area and contact potential of Miyazawa and Jernigan (24).

We performed Langevin MD simulations within Leap-Frog integration scheme with time step  $h = 5 \times 10^{-3} \tau$  and a friction coefficient  $\gamma = 0.05 \tau^{-1}$ , where the time unit  $\tau = \sigma \sqrt{M/\epsilon_0} = 4.67$  ps ( $M$  is the average mass of an amino-acid residue estimated to 110 Da). In this work, we considered unfolding rather than folding simulations. Simulations of protein unfolding under denaturing conditions are commonly employed (25,26) to gain insight on the folding mechanism in physiological state. This approach was originally developed within the framework of all-atom MD (see, for example, (26)), because high temperature computations drastically increase the reaction rate, thus significantly reducing the computational time. When coarse-grained models are used, however, as is our case, the key issue is not the computational load but rather the possibility to avoid kinetic traps due to the residual energetic and topological frustration of the force field. It is important to stress that, when using the unfolding strategy, extreme caution must be paid, because an increase in temperature may actually change the reaction mechanism rather than simply accelerating the same process (27). As a result, the folding pathway determined under strong folding conditions may not be the reverse of the unfolding pathway derived in strong denaturing conditions. To overcome this difficulty, we do not follow the unfolding process at a single high temperature; instead, we gradually heat the system in 50 temperature steps from  $T = 1.5 \bar{T}$  to  $T = 2.5 \bar{T}$ , where  $\bar{T} = 151$  K sets the temperature scale. Care is also paid in the temperature-jump simulations, by driving the system to a temperature just slightly above the unfolding point of the wild-type (WT), so that the unfolding mechanism can be safely assumed as being the reverse of folding. Moreover, the experimental evidence (8) that domain C5 of MyBP-C acts as a two-state folder with no intermediates along the folding route, lends further support to the suitability of unfolding simulations for the study of this protein. In fact, in a recent article (28), Karplus showed that when a protein can fold along alternative pathways, some of which involved misfolded intermediates, the unfolding simulation then preferentially follows the path with the lowest free energy barriers neglecting the alternative routes. As a consequence, unfolding studies are better suitable to proteins lacking off-pathway intermediates as is the case of domain C5 of MyBP-C.

The C5 domain of MyBP-C, as many small single-domain proteins, folds and unfolds in an all-or-none process without populating partly structured intermediates (8). The kinetics of the reaction  $U \leftrightarrow F$ , starting from an initial condition with proteins initially in the native state, can thus be modeled as

$$\frac{dF}{dt} = -(k_f + k_u)F + k_f F_0, \quad F(0) = F_0$$

where  $F$  represents the folded protein concentration, and  $k_u$  and  $k_f$  are the rates of unfolding and folding, respectively. The value  $F_0$  is the total concentration of the protein that remains constant throughout the process. Normalization of  $F(t)$  to  $F_0$  yields the probability that the protein be folded at time  $t$ :

$$P_f(t) = \frac{k_u}{k_f + k_u} \exp\{-(k_f + k_u)t\} + \frac{k_f}{k_f + k_u}. \quad (3)$$

The time course of  $P_f$  was computed by averaging in a series of  $10^3$  independent temperature-jump simulations starting from low temperature conformations equilibrated at  $T = 225$  K for a period of  $10^3 \tau$  and then heated at temperature  $T = 332$  K just above the transition point of WT and mutants.

Equation 3, when fitted to simulation data, yields estimates of the rates  $k_f$  and  $k_u$ , enabling a direct comparison between the kinetics of WT and mutants.

During the simulated T-jump experiments, we also computed the probability of any single native contact being intact. As the simulations were performed at a temperature just little higher than the transition point, it is reasonable to assume that the folding pathway is well described by the reverse of the sequence of contact breakdown in the T-jump simulations.

The dynamics of dissolution of interactions between couples of secondary structure elements  $X, Y$  was also tracked through the evolution of local native contacts (29) (also referred to as ‘‘partial overlaps’’). This quantity was computed by averaging over all trajectories ending up in an unfolded conformation according to the equations

$$Q_{XY}(t) = \frac{n_{X,Y}(t)}{n_{X,Y}^{\max}}, \quad (4)$$

$$\langle Q_{XY} \rangle(t) = \frac{1}{N_{\text{ut}}} \sum_{N_{\text{ut}}} Q_{XY}^{\text{Nut}}(t). \quad (5)$$

In Eq. 4,  $n_{X,Y}$  and  $n_{X,Y}^{\max}$  indicate the number of contacts between  $X$  and  $Y$  at time  $t$  and in the native structure, respectively, and  $N_{\text{ut}}$  is the number of trajectories leading to the denaturation of the protein.  $X$  and  $Y$  are any of the 15 secondary structure elements shown in Fig. 2. In this protein, 120 unique ( $X, Y$ ) pairs are possible, but only 13 of them are particularly relevant to monitor the reaction. Pairs CF, FG, and A'G are indicators of the status of  $\beta$ -sheet CFGA'; pairs BE and DE probe the condition of sheet BDE. The interactions among the loops on the same side of the molecule as the long CD loop are studied by analyzing the couples A'-A'B, A'B-EF, and CD-EF. In a similar fashion, the interactions among the loops on the opposite side of the molecule with respect to the CD loop, are monitored through the status of the pairs HG, H-BC; H-FG, HH, and BC-FG (with  $H$  being the N-terminal 1–17 segment). Two parameters were also introduced to follow the interactions between the two  $\beta$ -sheets and between any  $\beta$ -strand and any loop.

## RESULTS

An unfolding simulation of the wild-type (WT) sequence was performed by gradually heating the protein from temperature  $T = 225$  K to  $T = 375$  K in 50 temperature intervals. For each temperature,  $5 \times 10^7$  time steps were spent for the equilibration, followed by a production stage of  $5 \times 10^8$  time steps. A similar schedule was employed to simulate the unfolding of the three missense mutants Asn<sup>115</sup>Lys, Arg<sup>14</sup>His, and Arg<sup>28</sup>His (notice that protein residues have been renumbered 1–130 as a restriction to the C5 domain only). A residue mutation is implemented in the G $\bar{o}$ -model by turning all the native contacts in which the site is involved, into nonnative ones (31). The pathological mutations that we analyze, are strongly nonconservative, causing a significant change of the physical-chemical properties of the residue, justifying their modeling through the removal of all the contacts. We draw the attention on the fact that contacts between residues are weighted with the number of atomic overlaps. This implies that the removal of a single residue-residue contact corresponds to the cut of a different number of atomic interactions. As an example, we consider the mutations on sites 14 and 28. In both cases we delete three residue-residue contacts but in the former case we cut 16 atomic contacts while in the latter 28 are cleaved. In our opinion, this difference is relevant, and it is neglected by all G $\bar{o}$  models with homogeneous energy couplings. The role of the amino-terminal region of the protein was also investigated

through unfolding simulations of a deletion mutant where the first seven residues of the C5 domain were removed.

The presence of a single sharp peak in the thermograms of both WT and all the mutants (Fig. 3), suggests for these proteins a cooperative, two-state unfolding mechanism.

Mutations on Arg<sup>14</sup> and Arg<sup>28</sup> appear to cause just a moderate destabilization of the protein as they determine a little shift of the unfolding temperature to lower values. The mutation of Asn<sup>115</sup>, on the other hand, results in a more pronounced decrease of the transition temperature, suggesting a more destabilizing effect in agreement with experimental data. The NMR spectra recorded by Idowu et al. (8), in fact, show that the Asn<sup>115</sup>Lys mutant is unstable and largely unfolded as compared to the wild-type C5 motif. The NMR spectrum of the  $\Delta 1-7$  deletion mutant also lacks the peak dispersion typical of folded proteins and the molecule appears to be unstructured. This finding is consistent with our result that the amino-truncated protein is characterized by the largest shift in the unfolding temperature  $T_u$ , thus resulting in the most destabilized mutant.

The extent of the transition temperature decrease is clearly related to the number of contacts removed to implement the mutations. The removal of only three contacts to simulate the mutations on Arg<sup>14</sup> and Arg<sup>28</sup>, in fact, results in small and almost identical shifts of the positions of the peaks of the corresponding  $C_V$  plots. The mutation of Asn<sup>115</sup> produces a more relevant temperature shift because it entails the cleavage of 10 residue contacts. Finally, the removal of the N-terminal 1-7 fragment results in the loss of 41 native contacts determining the largest displacement of the thermogram. Not only the number, but also the position, of the removed contacts plays a relevant role. The  $C_V$  plot displacements of the different mutants, in fact, do not reflect the ratio of the lost contacts. As an example, the deletion mutant features a temperature shift twice larger than that of Mut<sup>115</sup>, in contrast with a 4:1

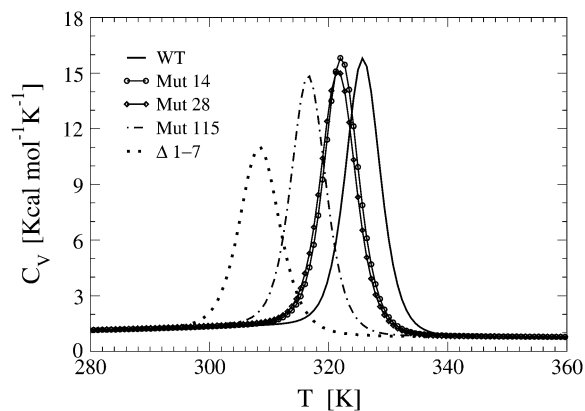


FIGURE 3 Thermal behavior of heat capacity of the WT C5 domain, the missense mutants deprived of the native contacts of Arg<sup>14</sup>, Arg<sup>28</sup>, and Asn<sup>115</sup>, and the deletion mutant lacking the 1-7 subsequence. Using the weighted histogram method, computations have been performed with the data collected during unfolding simulations.

ratio of removed native contacts. A possible explanation of this mismatch is that the contacts involving Asn<sup>115</sup> occupy a central position in the graph of interactions while those involving the first seven N-terminal residues are more peripheral.

The early unfolding behavior of the WT sequence of domain C5 was compared to that of the three missense and  $\Delta 1-7$  deletion mutants through T-jump simulations. The proteins were first equilibrated at very low temperature  $T_{low} = 225$  K for  $2 \times 10^5$  time steps and then heated to a temperature slightly higher than the unfolding point of the WT  $T_{high} = 332$  K. We stress that the choice of  $T_{high} = 332$  K provides a common denaturing temperature for both WT and mutants, representing also a reference point for the comparison of the kinetic rates of all the molecular species under examination. Finally, a temperature slightly higher than the unfolding temperature of mutants and WT, allows the denaturation to occur in reasonably short simulation times. The molecules were kept at  $T = T_{high}$  for  $6 \times 10^6$  time steps and during this period the nativeness of the proteins was checked every  $10^3$  time steps by monitoring the fraction of contacts still intact. When the overlap fell below a threshold  $Q_u = 0.25$  the protein was regarded as being unfolded. The progress of the denaturation process was measured by tracking the fraction of folded trajectories as a function of time,  $P_f(t)$ , plotted in Fig. 4 for the WT and mutants together with the free energy profiles at temperature  $T = T_{high}$  (inset). Fig. 4 shows that all the mutations speed up the unfolding process. As an example, Mut<sup>115</sup> and the deletion mutant become denatured in all trajectories no later than  $t = 5000 \tau$ . Mut<sup>28</sup> conversely, appears to be more stable and it resists until  $t = 15,000 \tau$  before unfolding in all runs. The behavior of Mut<sup>14</sup>, finally, is similar to that of the WT sequence, and even if the  $P_f^{Mut14}(t) < P_f^{WT}(t)$  for  $t > 2500 \tau$ , the difference never exceeds 15%. The kinetic behavior is

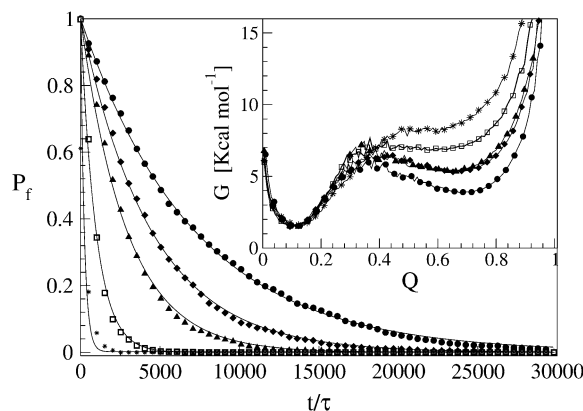


FIGURE 4 Comparison of unfolding kinetics of WT domain C5 with its deletion and missense mutants through the evolution of the fraction of folded trajectories  $P_f(t)$ . (Circles refer to WT, diamonds to Mut<sup>14</sup>, triangles to Mut<sup>28</sup>, squares to Mut<sup>115</sup>, and stars, to the deletion mutant.) The solid lines represent the exponential fits according to Eq. 3. (Inset) Free energy profiles of WT and mutants at the temperature of the T-jump simulations  $T = T_{high} = 332$  K as a function of the fraction of native contacts  $Q$ . Time is in units  $\tau = 4.67$  ps.

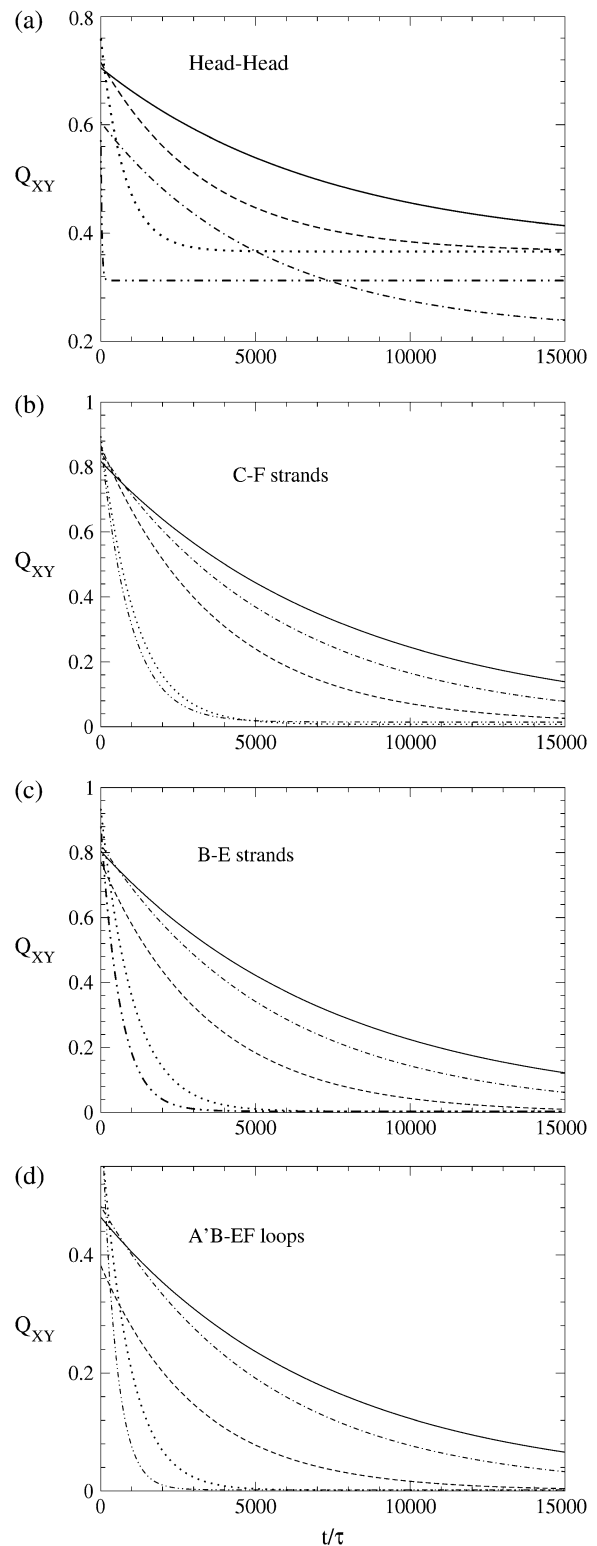
consistent with the free energy profiles (*inset*, Fig. 4) showing the mutations to reduce the unfolding barrier while simultaneously increasing the folding one. Equation 3 was fitted to the  $P_f(t)$  plots so as to estimate the kinetic rates. The excellent fit testified by correlation coefficients, always higher than 0.997, shows that both WT and mutants behave as two-state folders. Due to the high simulation temperature  $T = T_{\text{high}} = 332$  K, the unfolding rates dominate the process being, on average, eight orders-of-magnitude larger than folding ones. Table 1 summarizes the impact of mutations on the unfolding rate by the following ranking:  $k_u^{\text{WT}} < k_u^{\text{Mut}^{14}} < k_u^{\text{Mut}^{28}} < k_u^{\text{Mut}^{115}} < k_u^{\Delta 1-7}$ .

The analysis of the kinetics of local native contacts enables us to explore the local folding/unfolding mechanism occurring in different regions of the protein. Fig. 5 shows the partial overlaps of four different structural elements of domain C5 to assess the impact of the mutations on each of them. Attention is focused on four regions of the protein. First of all,  $\beta$ -sheet CFGA' is monitored in Fig. 5 *b*, through the interactions between strands C and F, while  $\beta$ -sheet BDE is analyzed in Fig. 5 *c* through the breakdown of contacts between strands B and E. The network of interactions among Head, BC loop, and FG loop (referred to as "lower loops") is followed in Fig. 5 *a* by monitoring the contacts within the Head region; finally in Fig. 5 *d*, we consider the evolution of the contacts between A'B and EF loops that, together with the CD loop, we name the "upper loops". The deletion of the 1–7 fragment and the removal of contacts of residue 115 (located in the FG loop) are chemical transformations taking place in the region of the lower loops, and it is therefore not surprising that they significantly speed up the breakdown of Head-Head contacts as shown in Fig. 5 *a*. Panels *b–d* of Fig. 5, however, indicate that Mut<sup>115</sup> and  $\Delta 1-7$  have a nontrivial long-range effect, strongly favoring the dissolution of both  $\beta$ -sheets and destabilizing the interactions of the upper loops. The importance of mutations Mut<sup>115</sup> and  $\Delta 1-7$  can be better appreciated by comparison with Mut<sup>14</sup>. Residue 14, in fact, is located in the Head region, and the effect of its mutation is only local. As it increases the velocity of breakdown of Head-Head contacts (Fig. 5 *a*), its impact on the other protein regions is negligible, as seen by  $Q_{XY}(t)$  plots almost identical to that of the WT (Fig. 5, *b–d*). The behavior of Mut<sup>28</sup> finally appears to be intermediate between the two extremes. The study of partial overlaps also provides useful information about the unfolding and folding mechanisms. Fig. 6 illustrates

**TABLE 1** Unfolding rates of WT and mutants determined from the fitting of the  $P_f(t)$  plots shown in Fig. 4 using Eq. 3

Species	$k_u[\tau^{-1}]$	$k_u/k_u^{\text{WT}}$
WT	$1.41 \times 10^{-4}$	1
Mut <sup>14</sup>	$2.26 \times 10^{-4}$	1.61
Mut <sup>28</sup>	$3.46 \times 10^{-4}$	2.47
Mut <sup>115</sup>	$1.04 \times 10^{-3}$	7.43
$\Delta 1-7$	$1.46 \times 10^{-3}$	15.93

The unfolding rates can be ranked as:  $k_u^{\text{WT}} < k_u^{\text{Mut}^{14}} < k_u^{\text{Mut}^{28}} < k_u^{\text{Mut}^{115}} < k_u^{\Delta 1-7}$ .



**FIGURE 5** Local native contacts computed according to Eqs. 4 and 5. The curves are the fits with an exponential to simulation data. (Curves: *solid*, WT; *dotted-dashed*, Mut<sup>14</sup>; *dashed*, Mut<sup>28</sup>; *dotted*, Mut<sup>115</sup>; and *double-dotted-dashed*,  $\Delta 1-7$  deletion mutant.) The quantity  $\langle Q_{XY} \rangle(t)$  represents the fraction of native contacts formed between the structural elements  $X$  and  $Y$  at time  $t$ . Mut<sup>115</sup> and  $\Delta 1-7$  exert long-range influence on the molecule, whereas Mut<sup>14</sup> and Mut<sup>28</sup> operate on a shorter length scale. Time is in units  $\tau = 4.67$  ps.

the time evolution of the  $Q_{XY}$  parameters of the WT, revealing the existence of two families of curves characterized by different slopes. Fig. 6, *a* and *b*, in particular, show that short-range contacts such as those linking the consecutive strands F-G and D-E and those connecting residues of the Head region with each other, decay more slowly than long-range contacts and their partial overlap reach a stable value at  $\sim 40\%$ . Since we work in a moderately denaturing condition, the unfolding can be considered reversible, and the features of the  $Q_{XY}$  plots suggest that short-range contacts are the first to appear during folding. Fig. 6 *a* indicates that the F-G contacts act as a nucleus for the formation of the CFGA' sheet that is completed much later when the fraction of F-G and A'-G contacts reaches a comparable level. In a similar way, the D-E contacts may represent a nucleus for the formation of the BDE sheet that gains its final conformation when the increase of the fraction of B-E contacts allows strand B to join the structure. A comparison of panels *a* and *c* in Fig. 6 shows that the dissolution/formation of the hydrophobic core follows the same dynamics as the destruction/formation of the two  $\beta$ -sheets. It follows that, in domain C5, the two sheets come close to each other during the process of assembling their structures. A final remark refers to the interactions among the higher loops (Fig. 6 *d*), whose  $Q_{XY}$  parameter does not exceed 50% even at the beginning of the unfolding process, as a consequence of the very high mobility of these poorly structured regions.

Further understanding of the unfolding/folding mechanism can be attained by the analysis of the time evolution of native contact probabilities that, for a T-jump simulation, show a monotonic decay. We assume a contact to be broken when the probability  $P_{ij}$  reaches a cutoff value  $\pi = 0.5$ . We then compute the average time of threshold crossing for all contacts involving the same bead, as an estimate of the unfolding time of that residue. The folding mechanism, i.e., the order according to which residues assume their native position, can be determined by sorting their average unfolding times in a decreasing order. The discontinuous blocks observable in the histogram of unfolding times identify the key stages of the folding process, when groups of residues take their native conformation in a coordinated manner. The folding process is expected to proceed according to the following stages:

- Stage 1: Formation of FG contacts in proximity of the FG loop (Fig. 7, *a* and *b*).
- Stage 2: Appearance of DE hairpin (Fig. 7 *c*).
- Stage 3: Strand C joins the FG hairpin while strand B moves close to the DE hairpin (Fig. 7 *d*).
- Stage 4: Stabilization of C-terminal ends of strands G and A' and A'B and EF loops in their native conformation (Fig. 7, *e* and *f*).

The folding mechanism of the C5 domain thus appears to be mainly driven by conformational entropy. The short-range contacts of the FG and DE hairpins, in fact, are the first to appear during refolding, as their formation is accompanied

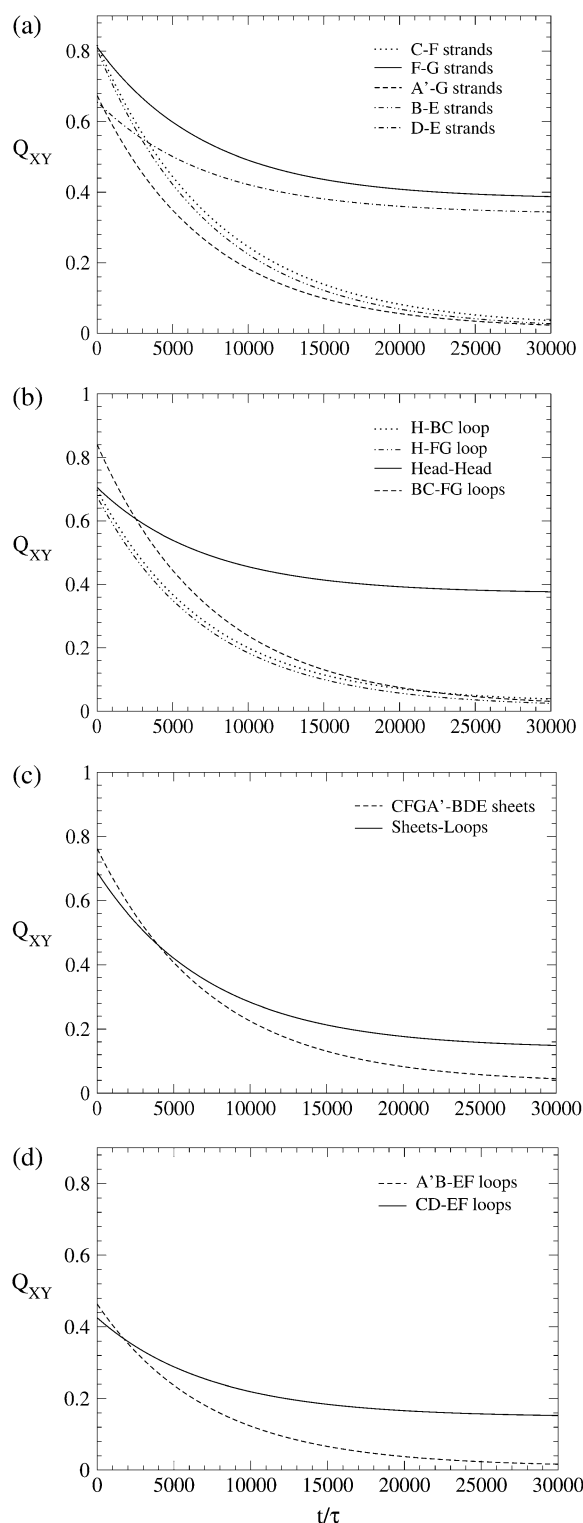


FIGURE 6 Time course of local native contacts (Eqs. 4 and 5) of WT domain C5 in four different regions of the molecule. (*a*)  $\beta$ -sheets; (*b*) lower loops; (*c*) sheet-sheet and loop-sheet interactions; and (*d*) upper loops. The data have been fitted with an exponential to improve readability. Time is in units  $\tau = 4.67$  ps.

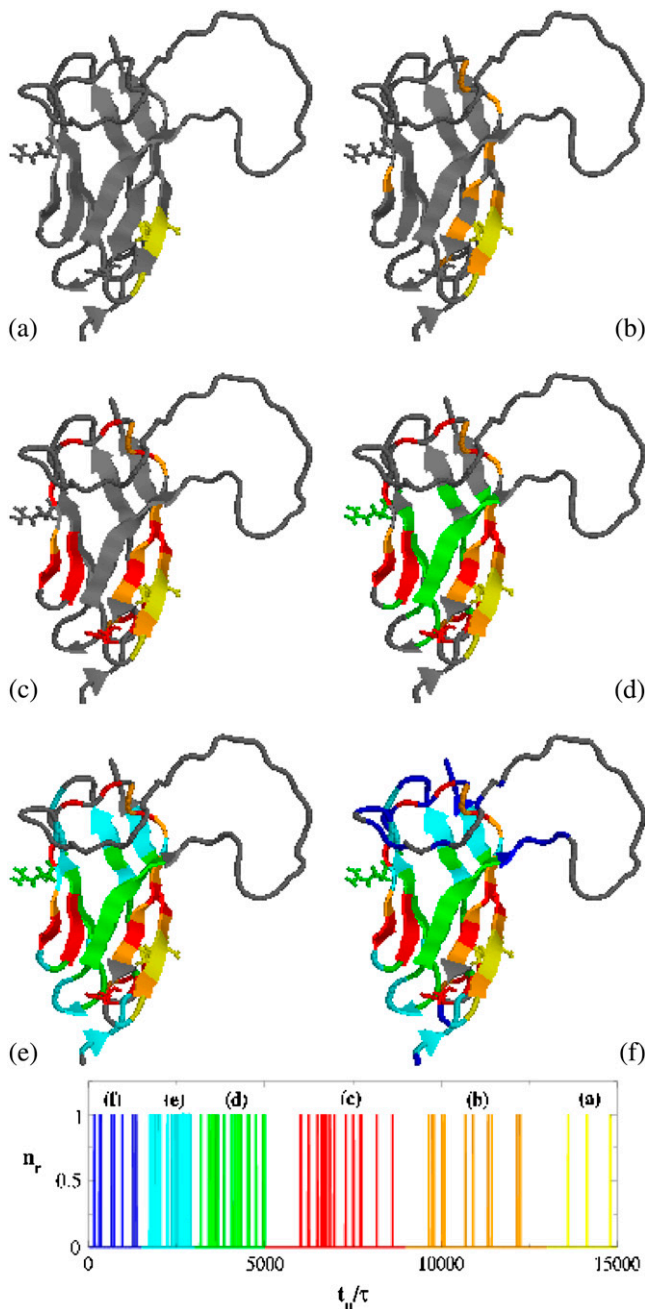


FIGURE 7 Representation of the folding pathway of domain C5 from MyBP-C. (a–f) The regions in color indicate residues in natively like conformation; gray segments are still coil. (Blue,  $0 < t_u < 1.5 \times 10^3$ ; cyan,  $1.5 \times 10^3 < t_u < 3 \times 10^3$ ; green,  $3 \times 10^3 < t_u < 5 \times 10^3$ ; red,  $5 \times 10^3 < t_u < 9 \times 10^3$ ; orange,  $9 \times 10^3 < t_u < 1.3 \times 10^4$ ; and yellow,  $t_u > 1.3 \times 10^4$ .) The bars represent the distribution of the unfolding times. Time is in units  $\tau = 4.67$  ps.

with a modest loss in conformational entropy. In particular, the DE and FG loops, comprising two and four residues, respectively, are the shortest loops of the C5 domain and it is therefore not surprising that the refolding of the two hairpins starts in this region. On the contrary, the major length of the A'B, CD, and EF loops determines their late refolding. This

pathway suggests a way to rationalize the different impact of the FHC-causing mutations on C5 domain. For example, residue 115, being located in the FG loop, when mutated is likely to hinder the very first events of folding upon mutation. Residue 28, conversely, is located in the BDE sheet, but outside the crucial DE hairpin. Therefore its mutation is expected to influence only the structuring of the protein portion near the long A'B, CD, and EF loops, thus resulting in a mild phenotype.

To test our explanation of the pathogenic role of the mutations on residues 14, 28, and 115, we performed temperature-jump simulations of 27 other mutants obtained by changing residues evenly distributed in all districts of domain C5. The results of our simulations are summarized on Table 2. Tables 1 and 2 show that the most destabilizing mutations, characterized by  $k_u/k_u^{\text{WT}} \geq 5$ , are located on strand F (residues 111, 112, 113), on the FG-loop (residue 115), on strand E (residue 93), on the BC-loop (residue 35), and on strand C (residue 43). This distribution confirms the conclusions drawn after analyzing the folding mechanism of the wild-type protein: the strand F and the FG-loop are the most sensitive regions to mutations, explaining the severe phenotype from FHC mutation of residue 115. Our kinetic analysis also underscores the importance of strand E, which is the central element of the BDE sheet and acts as a bridge between strands B and D. Tables 1 and 2 show

TABLE 2 Kinetic parameters of test-mutants determined from temperature-jump simulations: unfolding rates and unfolding rates normalized to the  $k_u$  of the WT

Residue	Sec. Struct.	$k_u[\tau^{-1}]$	$k_u/k_u^{\text{WT}}$
Res 12	Head	$1.43 \times 10^{-4}$	1.01
Res 20	A'	$4.18 \times 10^{-4}$	2.96
Res 21	A'	$4.12 \times 10^{-4}$	2.92
Res 22	A'B	$2.10 \times 10^{-4}$	1.43
Res 31	B	$6.65 \times 10^{-4}$	4.72
<b>Res 35</b>	<b>BC</b>	<b><math>8.81 \times 10^{-4}</math></b>	<b>6.25</b>
Res 36	BC	$3.52 \times 10^{-4}$	2.50
Res 42	C	$3.07 \times 10^{-4}$	2.18
<b>Res 43</b>	<b>C</b>	<b><math>1.16 \times 10^{-3}</math></b>	<b>8.23</b>
Res 44	C	$3.01 \times 10^{-4}$	2.13
Res 46	C	$1.64 \times 10^{-4}$	1.16
Res 86	D	$2.14 \times 10^{-4}$	1.52
Res 90	D	$1.93 \times 10^{-4}$	1.37
Res 91	DE	$2.75 \times 10^{-4}$	1.95
Res 93	E	$6.80 \times 10^{-4}$	4.82
<b>Res 96</b>	<b>E</b>	<b><math>1.64 \times 10^{-3}</math></b>	<b>11.63</b>
Res 97	E	$5.60 \times 10^{-4}$	3.98
Res 109	F	$2.66 \times 10^{-4}$	1.89
<b>Res 111</b>	<b>F</b>	<b><math>1.40 \times 10^{-3}</math></b>	<b>9.93</b>
<b>Res 112</b>	<b>F</b>	<b><math>7.70 \times 10^{-4}</math></b>	<b>5.46</b>
<b>Res 113</b>	<b>F</b>	<b><math>1.46 \times 10^{-3}</math></b>	<b>10.35</b>
Res 118	FG	$3.18 \times 10^{-4}$	2.26
Res 120	G	$5.57 \times 10^{-4}$	3.95
Res 121	G	$3.83 \times 10^{-4}$	2.72
Res 126	G	$4.92 \times 10^{-4}$	3.49
Res 127	G	$3.74 \times 10^{-4}$	2.65
Res 128	G	$3.50 \times 10^{-4}$	2.48

The most destabilizing mutations are marked in bold.



that residue 28, due to its position, is surrounded by a set of residues such as 20, 21, 22 (strand A), 42, 44, 46 (strand C), 127, and 128 (strand G). A mutation on this set has a moderate impact on the unfolding rates, thus explaining the mild clinical phenotype deriving from the mutation of residue 28.

## CONCLUSION

The molecular impact of three mutations, Asn<sup>755</sup>Lys, Arg<sup>654</sup>His, and Arg<sup>668</sup>His, on the domain C5 from MyBP-C related to FHC has been studied through MD simulations. As Ig-like proteins tend to fold and unfold along a common pathway dictated by the geometry of the common structural core (18), C5 domain was modeled through a heavy map Gō-like force field with heterogeneous energy couplings. We preferred to run unfolding rather than folding simulations to avoid kinetic traps related to residual energetic and topological frustration. To determine the kinetics of the folding process by inverting to the unfolding pathway, we performed temperature jump simulations at a temperature just little above the transition point of both WT and mutants. The destabilizing effect of mutations can be quantified by the corresponding shift in the unfolding temperature. The simulations show that the mutation of Asn<sup>755</sup> is particularly destabilizing, causing a large decrease in the transition temperature. This result is consistent with the NMR spectra recorded by Idowu et al. (8), signaling the absence of structure for the mutant even at low temperature. Residues Arg<sup>654</sup> and Arg<sup>668</sup>, on the other hand, when mutated determine similar and more limited shifts of the peaks of the  $C_V$  plots, suggesting a very small influence of these two residues on protein stability.

The role of the N-terminal region was also investigated, studying a deletion mutant where the first seven residues had been removed. The significant decrease in  $T_u$  of the truncated protein suggests that the N-terminal region with its 10-residue-long insert, typical of the cardiac isoform, is not just a linker between the C4 and C5 domains, but it gives a nontrivial contribution to the stability. The low “contact betweenness” (32), however, hints at the involvement of the N-terminal residues in contacts forming a subgraph only weakly connected to the main network. It is thus possible that the N-terminal contacts do form only when the C5 domain is dissected from the rest of the protein and that the natural role of this region is more related to the binding with domain C8, complementing the negatively charged CFGA' surface (8).

The kinetic unfolding simulations show that the mutations determining the largest decrease in the unfolding temperature also induce the highest speedup of the unfolding process. In particular, the mutants retain the two-state unfolding mechanism but they tend to reduce the unfolding barrier while simultaneously increasing the folding one. The analysis of the local native contacts shows that the mutations on Arg<sup>654</sup> and Arg<sup>668</sup> only have a local effect, whereas the mutation of Asn<sup>755</sup> and the  $\Delta 1-7$  deletion have a strong influence even on regions at the opposite end of the molecule. An even stronger

evidence is provided by the contact probabilities  $P_{ij}(t)$ , indicating the folding reaction to start from the region of the DE and FG loops and to proceed toward the region of the CD and EF loops. This may explain the importance of Asn<sup>755</sup>, which is located in the zone where the folding originates. By contrast, Arg<sup>668</sup> lies near the opposite end of the protein where the folding is almost concluded, explaining the modest kinetic and thermodynamic effect of mutating this residue. This conclusion was supported by performing temperature-jump simulations on 27 test mutants, proving the importance of the FG-loop and of the central strands of the two  $\beta$ -sheets (strands E and F). The simulations also showed the existence of a strip of residues (including Arg<sup>668</sup>) on the side of the  $\beta$ -strands close to the CD and EF loops, whose mutation only marginally affects the folding/unfolding kinetics.

Another interesting feature of the cardiac C5 domain is the destabilizing effect of the long CD loop, characterized by a 28-residue-long insert and lowering the folding temperature with respect to other Ig domains, which we investigated in a different work (32). There we found a charge unbalance and a low hydrophobicity of the CD loop, accompanied by a high concentration of Pro, Glu, and Asp. This supports the idea that the C5 domain of MyBP-C can be considered a natively unfolded protein, i.e., a protein lacking a compact, globular structure under physiological conditions (33,34). The high mobility of the unstructured CD loop leads to a reduced number of stabilizing contacts with the EF loop, which is possibly compensated for by a higher relevance of lower loops providing an interpretation for the different pathologic role of the three FHC-related mutations.

This work has been partially supported by EU-FP6 contract 012835 (EMBio).

## REFERENCES

1. Winegrad, S. 2000. Myosin binding protein C, a potential regulator of cardiac contractility. *Circ. Res.* 86:6–7.
2. Mogensen, J., I. C. Klausen, A. K. Pedersen, H. Egeblad, P. Bross, T. A. Kruse, N. Gregersen, P. S. Hansen, U. Baandrup, and A. D. Børglum. 1999.  $\alpha$ -Cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy. *J. Clin. Invest.* 103:R39–R43.
3. Charron, P., O. Dubourg, M. Desnos, R. Isnard, R. Hagege, G. Bonne, L. Carrier, F. Tesón, J. B. Bouhour, J.-C. Buzzi, J. Feingold, K. Schwartz, and M. Komajda. 1998. Genotype-phenotype correlations in familial hypertrophic cardiomyopathy. A comparison between mutations in the cardiac protein-C and the  $\beta$ -myosin heavy chain genes. *Eur. Heart J.* 19:139–145.
4. Daehmlow, S., J. Erdmann, T. Knueppel, C. Gille, C. Froemmel, M. Hummel, R. Hetzer, and V. Regitz-Zagrosek. 2002. Novel mutations in sarcomeric protein genes in dilated cardiomyopathy. *Biophys. Res. Comm.* 298:116–120.
5. Flashman, E., C. Redwood, J. Moolman-Smoock, and H. Watkins. 2004. Cardiac myosin binding protein C: its role in physiology and disease. *Circ. Res.* 94:1279–1289.
6. Moolman-Smoock, J., E. Flashman, W. de Lange, Z. L. Li, V. Corfield, C. Redwood, and H. Watkins. 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.



7. Kunst, G., K. R. Kress, M. Gruen, D. Uttenweiler, M. Gautel, and R. H. A. Fink. 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.
8. Idowu, S. M., M. Gautel, S. J. Perkins, and M. Pfuhl. 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.
9. Hartzell, H. C., and D. B. Glass. 1984. Phosphorylation of purified cardiac muscle C-protein by purified cAMP-dependent and endogenous  $\text{Ca}^{2+}$ -calmodulin-dependent protein kinases. *J. Biol. Chem.* 259:15587–15596.
10. Morner, S., P. Richard, E. Kazzam, U. Hellman, B. Hainque, K. Schwartz, and A. Waldenstrom. 2003. Identification of the genotypes causing hypertrophic cardiomyopathy in northern Sweden. *J. Mol. Cell. Cardiol.* 35:841–849.
11. Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski, and D. Baker. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6:1016–1024.
12. Chiti, F., N. Taddei, P. M. White, M. Bucciantini, F. Magherini, M. Stefani, and C. M. Dobson. 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* 6:1005–1009.
13. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? *J. Mol. Biol.* 298:937–953.
14. Ceconi, F., C. Micheletti, P. Carloni, and A. Maritan. 2001. The structural basis of antiviral drug resistance and role of folding pathways in HIV-1 protease. *Proteins Struct. Funct. Genet.* 43:365–372.
15. Micheletti, C., F. Ceconi, A. Flammini, and A. Maritan. 2002. Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs. *Protein Sci.* 11:1878–1887.
16. Hoang, T. X., and M. Cieplak. 2000. Sequencing of folding events in Gō-type proteins. *J. Chem. Phys.* 113:8319–8328.
17. Ceconi, F., C. Guardiani, and R. Livi. 2006. Testing simplified protein models of the hPin1 WW domain. *Biophys. J.* 91:694–704.
18. Clarke, J., E. Cota, S. B. Fowler, and S. Hamill. 1999. Folding studies of immunoglobulin-like  $\beta$ -sandwich proteins suggest that they share a common folding pathway. *Structure.* 7:1145–1153.
19. Matysiak, S., and C. Clementi. 2004. Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: how far can a minimalist model go? *J. Mol. Biol.* 343:235–248.
20. Serrano, L., J. Kellis, P. Cann, A. Matouschek, and A. R. Fersht. 1992. The folding of an enzyme. II. Substructure of Barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* 224:783–804.
21. Kurochkina, N., and B. Lee. 1995. Hydrophobic potential by pairwise surface area sum. *Protein Eng.* 8:437–442.
22. Sung, S. 1999. Monte Carlo simulations of  $\beta$ -hairpin folding at constant temperature. *Biophys. J.* 76:164–175.
23. Zhou, H., and Y. Zhou. 2004. Quantifying the effect of burial of amino acid residues on protein stability. *Proteins Struct. Funct. Bioinf.* 54:315–322.
24. Miyazawa, S., and R. L. Jernigan. 1985. Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.
25. Cieplak, M., and J. I. Sulkowska. 2005. Thermal unfolding of proteins. *J. Chem. Phys.* 123:194908-1–4.
26. Caffisch, A., and M. Karplus. 1994. Molecular dynamics studies of protein and peptide folding and unfolding. In *The Protein Folding Problem and Tertiary Structure Prediction*. K. M. Merz Jr. and S. M. Le Grand, editors. Birkhauser, Boston.
27. Finkelstein, A. V. 1997. Can protein unfolding simulate protein folding? *Protein Eng.* 10:843–845.
28. Dinner, A. R., and M. Karplus. 1999. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* 292:403–419.
29. Stoycheva, A. D., C. L. Brooks III, and J. N. Onuchic. 2004. Gatekeepers in the ribosomal protein S6: thermodynamics, kinetics, and folding pathways revealed by a minimalist protein model. *J. Mol. Biol.* 340:571–585.
30. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.
31. Sutto, L., G. Tiana, and R. A. Brogna. 2006. Sequence of events in folding mechanism: beyond the Gō model. *Protein Sci.* 15:1638–1652.
32. Guardiani, C., F. Ceconi, and R. Livi. 2007. Computational analysis of folding and mutation properties of C5 domain of myosin binding protein C. *Proteins Struct. Funct. Bioinf.* doi 10.1002/prot.21621.
33. Romero, P., Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guillot, and A. K. Dunker. 1998. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* 3:437–448.
34. Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* 41:415–427.