Eur Biophys J (2008) 37:683–691
DOI 10.1007/s00249-008-0308-x

ORIGINAL PAPER

# Analyzing pathogenic mutations of C5 domain from cardiac myosin binding protein C through MD simulations

**Fabio Cecconi · Carlo Guardiani · Roberto Livi**

Received: 26 September 2007 / Revised: 4 February 2008 / Accepted: 10 March 2008 / Published online: 1 April 2008


**Abstract** The folding properties of wild type and mutants of domain C5 from cardiac myosin binding protein C have been investigated via molecular dynamics simulations within the framework of a native-centric and coarse-grained model. The relevance of a mutation has been assessed through the shift in the unfolding temperature, the change in the unfolding rate it determines and Phi-values analysis. In a previous paper (Guardiani et al. Biophys J 94:1403–1411, 2008), we performed Kinetic simulations on native contact formation revealing an entropy-driven folding pathway originating near the FG and DE loops. This folding mechanism allowed also a possible interpretation of the molecular impact of the three mutations, Arg14His, Arg28His and Asn115Lys involved in the Familial Hypertrophic Cardiomyopathy. Here we extend that analysis by enriching the mutant pool and we identify a correlation between unfolding rates and the number of native contacts retained in the transition state.

Advanced neutron scattering and complementary techniques to study biological systems. Contributions from the meetings, "Neutrons in Biology", STFC Rutherford Appleton Laboratory, Didcot, UK, 11–13 July and "Proteins At Work 2007", Perugia, Italy, 28–30 May 2007.

F. Cecconi (✉)
INFM-SMC and Istituto dei Sistemi Complessi ISC-CNR,
Via dei Taurini 19, 00185 Rome, Italy
e-mail: cecconif@roma1.infn.it

C. Guardiani
Centro Interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC) Sezione INFN di Firenze, Florence, Italy

R. Livi
Dipartimento di Fisica Università di Firenze and Centro Interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC), Sezione INFN di Firenze and INFM UdR Firenze, Florence, Italy

## Introduction

The C5 domain from the cardiac Myosin Binding Protein C (MyBP-C) is a Ig-like domain with the typical $\beta$-sandwich topology characterized by two $\beta$ sheets BDE and CFGA′ closely packed against each other (Fig. 1). The structural and thermodynamical properties of the molecule have been experimentally studied by Idowu et al. 2003. The importance of MyBP-C relies on the fact that it is a target of mutations related to familial hypertrophic cardiomyopathy (FHC) (Chung et al. 2003; Richard et al. 2003), a disease impairing the physiologic function of this protein as a potential regulator for cardiac muscular contraction (Charron et al. 1998; Winegrad 2000; Flashman et al. 2004). Genetic analysis on FHC patients (Niimura et al. 1998; Moolman-Smook et al. 2002) revealed the mild form of the disease to be determined by nonsense mutations (premature termination of translation) whereas a more severe variant is related to missense mutations (amino-acid replacements). The C5 domain is known to develop three missense FHC mutations Arg14His, Arg28His, Asn115Lys whose role on stability and folding will be investigated through molecular dynamics simulations within the framework of a native centric coarse-grained approach based on a Gō-like model (Go and Scheraga 1976). Gō-models have the valuable property to correctly embody the native state topology that in many cases is the key determinant of the folding process (Chiti et al. 1999; Riddle et al. 1999; Cecconi et al. 2001; Micheletti et al. 2002). The study of Clarke et al. (1999) on the Ig-superfamily

indicates the native-centric approach as the natural framework to investigate the folding properties of the C5 Ig-like domain. Indeed according to Clarke et al. 1999, Ig-superfamily members, having transition and native states both stabilized by the same interactions dictated by protein topology, share similar folding pathways mainly determined by a common geometry.

Since we address the problem of resolving the effects of mutations, we also introduced some amount of heterogeneity in the energetic couplings of the Gō force field. This approach is necessary to discriminate mutations modeled through the removal of the same number of native contacts. The relevance of the energetic heterogeneity to a reliable mutation analysis was suggested in papers (Matysiak and Clementi 2004; Plotkin and Onuchic 2002; Tiana et al. 2006; Gutin et al. 1995).

## Methods

The coarse-grained description of the C5 domain takes into account only the backbone formed by the Cα carbons including side-chain interaction only through effective Lennard–Jones like potentials. The Gō-model force field minimizes the frustration by promoting the native interactions so that, the energy reduction is always accompanied by an increase in the similarity to the native structure. We refer to the force field introduced by Clementi et al. (2000), including also angular terms in the energy function $V_{\rm tot}$ to favour the correct folding of secondary structural elements:

$$
\begin{aligned}
V_{\rm tot} = & \sum_{i=1}^{N-1} \frac{k_h}{2}(r_{i,i+1} - r^0_{i,i+1})^2 + \sum_{i=2}^{N-1} \frac{k_\theta}{2}(\theta_i - \theta^0_i)^2 \\
& + \sum_{i=3}^{N-2} k^{(1)}_\phi [1 - \cos(\phi_i - \phi^0_i)] + k^{(3)}_\phi [1 - \cos 3(\phi_i - \phi^0_i)] \\
& + \sum_{i,j>i+3} \left\{ \Delta_{ij}\varepsilon_{ij} \left[ 5\left(\frac{r^0_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r^0_{ij}}{r_{ij}}\right)^{10} \right] \right. \\
& \left. + (1 - \Delta_{ij})\varepsilon_r \left(\frac{\sigma}{r_{ij}}\right)^{12} \right\}.
\end{aligned}
\tag{1}
$$

The contact matrix entries are $\Delta_{ij} = 1$ if residues $i$ and $j$ have at least a pair of heavy atoms closer than a distance cutoff $R_c = 5$ Å (native contact), otherwise $\Delta_{ij} = 0$. The cutoff choice identifies on the native structure (pdb id 1GXE) 277 native contacts. In formula (1), $r_{ij}$ indicates the distance between residues $i$ and $j$, $\theta_i$ is the bending angle identified by the three consecutive $C_\alpha$'s $i-1$, $i$, $i+1$, $\phi_i$ is the dihedral angle defined by the two planes formed by four consecutive $C_\alpha$'s $i-2$, $i-1$, $i$, $i+1$. The symbols with the superscript 0 are the corresponding quantities in the native conformation. The energy parameters are proportional to the energy scale $\varepsilon_0 = 0.3$ Kcal/mol such that: $k_h = 1{,}000\varepsilon_0/r_0^2$, ($r_0 = 3.8$ Å), $k_\theta = 20\varepsilon_0$, $k^{(1)}_\phi = \varepsilon_0$ and $k^{(3)}_\phi = 0.5\varepsilon_0$.

The strength of the repulsive Lennard–Jones terms between nonnative contacts are chosen as: $\sigma = 5.0$ Å, $\varepsilon_r = 2/3\varepsilon_0$. To incorporate topology and some specific chemical features as well as the effect of the steric hindrance of side chains, an heavy-map Gō model representation was considered. Within this framework, the strength of attractive native interactions is weighted according to the number of atomic contacts: $\varepsilon_{ij} = \varepsilon_0(1 + v_{ij})$ where $v_{ij}$ is the number of atomic contacts between residues $i$ and $j$ in the native state normalized to the maximum. This approach is supported by both experimental findings (Serrano et al. 1992) and theoretical arguments (Kurochkina and Lee 1995; Zhou and Zhou 2004).

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 $\tau = \sigma\sqrt{M/\varepsilon_0} = 4.67$ ps ($M$ is the average mass of an amino-acid residue estimated to 110 Da).

In the present paper we run unfolding rather than folding simulations to avoid kinetic trapping related to the residual energetic and topological frustration of the force field. Under the assumption that, in the same equilibrium conditions, the direct and reverse reaction must follow the

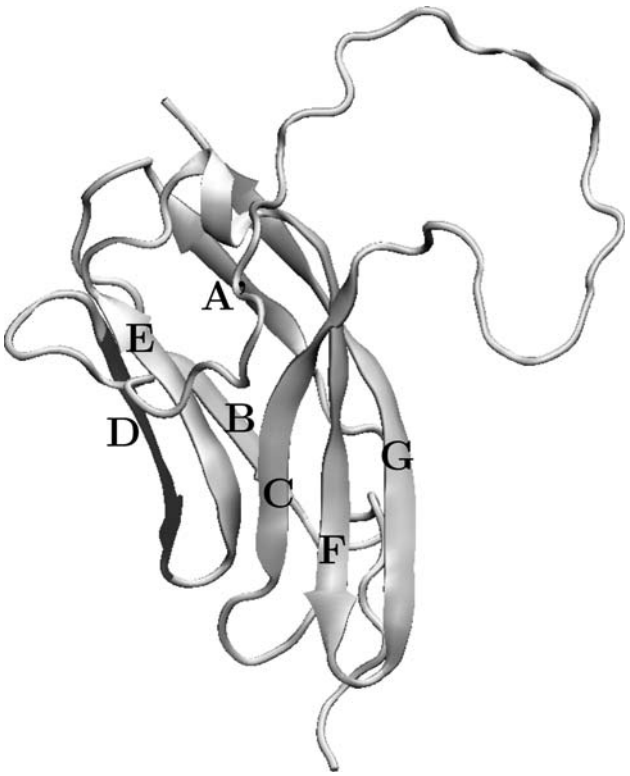


**Fig. 1** Secondary structure elements of the C5-domain (PDB identifier 1GXE) drawn using the VMD software (Humphrey et al. 1996)

same pathway, the use of unfolding simulations is considered a viable strategy to gain information also on the folding process (Cieplak and Sulkowska 2005; Caflish and Karplus 1994). For this reason, thermodynamics was obtained by a very slow heating process of the system in $N_{\text{run}} = 50$ temperature steps from $T = 225$ to $T = 375$ K. Moreover, in order to ensure that Kinetic simulations were not performed in strongly denaturing conditions incompatible with folding, care was also paid in the temperature-jump simulations, by driving the system to a temperature just little above the unfolding point of the WT.

The thermodynamic properties of the unfolding transition were computed using the weighted histogram analysis method (WHAM) (Ferrenberg and Swendsen 1989). According to WHAM, a two-dimensional histogram $h_r(E,Q)$ of energy and overlap $Q$ (fraction of native contacts) has been collected in the $r$th run at temperature $T_r$, with $r = 1,...,N_{\text{run}}$. The histograms have been combined so as to yield the density of states via formula

$$\rho(E,Q) = \frac{\sum_{r=1}^{N_{\text{run}}} h_r(E,Q)}{\sum_{l=1}^{N_{\text{run}}} N_l \mathrm{e}^{-\beta_l(E-F_l)}} \tag{2}$$

where $\beta_l = 1/(RT_l)$ ($R = 1.987 \times 10^{-3}$ Kcal K$^{-1}$ mol$^{-1}$ gas constant), $N_l$ is the number of measurements in the $l$-th histogram and $F_l$ is the free-energy such that

$$\mathrm{e}^{-\beta_l F_l} = \sum_{E,Q} \rho(E,Q)\mathrm{e}^{-\beta_l E} \tag{3}$$

Equations 2 and 3 have to be solved self-consistently in the set $\{F_1,...,F_{\text{run}}\}$ and $\rho(E,Q)$. From the knowledge of $\rho(E,Q)$, we obtain the thermal behavior of any observable that can be expressed as function of $E$ and $Q$, $A = A(E,Q)$

$$\langle A \rangle_\beta = \frac{1}{Z(\beta)} \sum_{E,Q} \rho(E,Q) A(E,Q) \mathrm{e}^{-\beta E}$$

where $Z(\beta)$ is the partition function. In the following, we apply the above formula with $A = E$ to compute the thermal capacity

$$C_v = R\beta^2(\langle E^2 \rangle - \langle E \rangle^2)$$

The density of states can be used to evaluate the probability

$$P_T(Q) = \frac{1}{Z(\beta)} \sum_{E} \rho(E,Q)\mathrm{e}^{-\beta E}$$

for the protein to have a specific value $Q$ at temperature $T$. It is then possible to derive the free energy as a function of $Q$ via the expression

$$G_T(Q) = -RT \log[P_T(Q)]$$

The advantage of the WHAM is twofold: on the one hand, it allows a better sampling of the conformation space than ordinary methods, on the other, it enables interpolation and smoothing of thermodynamic data.

An indicator of the native-likeness of residues in the transition state (TS) is represented by the Φ-value (Fersht et al. 1992): a value $\Phi \sim 1$, characterizes residues establishing native-like interactions already in the TS, whereas a value close to zero is typical of residues involved into a disordered conformation in the TS. We apply the free energy perturbation technique (FEP) (Clementi et al. 2000) to evaluate the Φ-values from our MD simulations. In the framework of FEP, a mutation is seen as a small perturbation of the wild type protein (wt) force field. As a consequence, it is possible to obtain free-energy difference between mutants and wt as

$$G_{\text{mut}} - G_{\text{WT}} = -RT \log \langle \mathrm{e}^{-\beta \delta E} \rangle_{\text{WT}} \tag{4}$$

where $\delta E$ is the amount of perturbation energy induced by a mutation on the wt system and the thermal average is over wt equilibrium conformations only. According to Eq. 4, the FEP Φ-value computation entails a long simulation at the unfolding temperature $T_u$, during which wt conformations are sampled in three windows of the coordinate $Q$, corresponding to the folded ($F$), transition state (TS) and Denaturated ($D$) ensembles. Mutations are then implemented and Φ-values are computed through Eq. 4

$$\Phi = \frac{\log \langle \mathrm{e}^{-\delta E/RT} \rangle_{\text{TS}} - \log \langle \mathrm{e}^{-\delta E/RT} \rangle_D}{\log \langle \mathrm{e}^{-\delta E/RT} \rangle_F - \log \langle \mathrm{e}^{-\delta E/RT} \rangle_D} \tag{5}$$

where the Boltzmann factors depend on the energy difference between the mutant and the wt, and the averages are restricted to wt-conformations of the folded ($F$), transition state (TS) and denaturated ($D$) ensembles.

Equation 5 requires some remarks as it entails the correct sampling of conformations belonging to the TSE, an extremely delicate issue of a long-standing debate still dividing the scientific community. Some authors believe (Shakhnovich 2006) that structural parameters like $Q$ are not sensitive enough to identify TSE structures and that a meaningful TSE sampling could only be accomplished through the $P_{\text{fold}}$ parameter, i.e. the probability that a given structure will reach a decidedly folded state before reaching the unfolded state. The $P_{\text{fold}}$ approach was recently used in a work on simple protein model amenable to exhaustive enumeration (Chang et al. 2004), where a partial disagreement was found between the TS structures identified using $P_{\text{fold}}$ and $Q$. Other researchers (Cho et al. 2006) conversely have argued that $P_{\text{fold}}$ analysis is computationally too demanding, requiring hundreds of simulations for each putative TS structure and thus applicable only to very small sets of conformations. Furthermore, $P_{\text{fold}}$ cannot be measured experimentally, so that the simulation results could not be directly checked against experimental data. On the basis of the energy landscape theory, these authors suggest, that at least in case of

moderate frustration, $P_{\text{fold}}$ may be replaced by a structurally defined reaction coordinate such as the fraction of native contacts $Q$. We believe that in Gō model approaches, the $Q$-based TS-sampling technique, despite its limitations (Chang et al. 2004), can be used to gather coarse-grained information on the regions of the protein more sensitive to mutations. The data produced in this way then serve as a starting point for further analysis and validation.

The C5 domain of MyBP-C is experimentally known to fold/unfold in an all-or-none process without populating partly structured intermediates (Idowu et al. 2003). Therefore, the kinetics of the denaturation reaction should be characterized by an exponential decay of the probability to remain folded at time $t$

$$P_f(t) = \frac{k_f}{k_f + k_u} + \frac{k_u}{k_f + k_u} e^{-(k_f + k_u)t} \quad (6)$$

where $k_f$ and $k_u$ are the folding and unfolding rates respectively. The time behavior of $P_f$ was obtained 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 6 when fitted to simulation data yields estimates of the rates $k_f$, $k_u$ enabling a direct comparison between the kinetics of wt and mutants. During the simulated T-jump experiments, also the probability of any single native contact being intact can be monitored, for details see (Guardiani et al. 2008). 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.

## Simulation results

The thermal denaturation of the C5 domain wt and mutants has been simulated by gradually heating the protein from temperature $T_1 = 225$ K to $T_2 = 375$ K through a sequence of 50 consecutive runs, each performed at constant temperature $T_n = T_1 + n\,\Delta T$, $n = 0,\ldots,49$, with $\Delta T = (T_2 - T_1)/49 = 3$ K. At each temperature, data were collected in a production stage of $5 \times 10^8$ time steps, after an equilibration of $5 \times 10^6$ time steps. The very small temperature increase $\Delta T$, together with the fact that final state (positions and momenta) of the $T_n$ run is used as the input for the next run at $T_{n+1} = T_n + \Delta T$, also ensures an efficient thermalization of the whole denaturation scheduling.

A residue mutation is implemented in the Gō-model by turning all the native contacts in which the site is involved, into non-native ones. The role of the amino-terminal region of the protein was also investigated through unfolding simulations of a deletion mutant where the first 7 residues of the C5 domain were removed. The presence of a single sharp peak in the thermograms of both wt and mutants (Fig. 2) suggests a two-state unfolding mechanism.

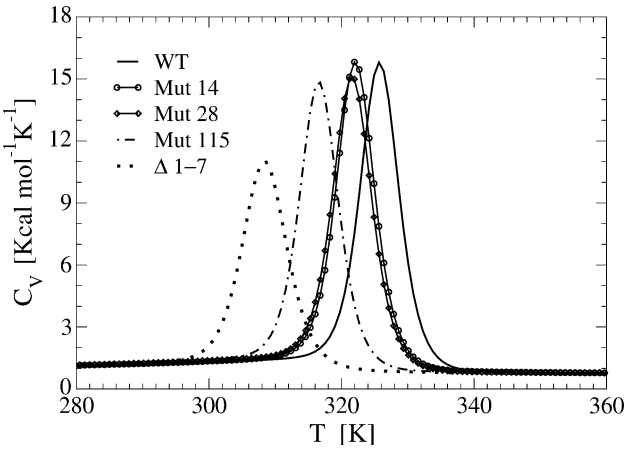


**Fig. 2** Thermal behavior of heat capacity of the WT C5 domain, the missense mutants deprived of the native contacts of Arg14, Arg28 and Asn115, and the deletion mutant lacking the 1–7 subsequence. Computations have been performed processing with the weighted histogram method the data collected during folding simulations

**Table 1** Experimental $T_{\text{exp}}$ and simulated $T_u$ unfolding temperatures of the WT domain C5 of MyBP-C and its mutants

| Species | $T_{\text{exp}}$ (K) | $T_u$ (K) |
|---|---|---|
| WT | 322.06 ± 0.75 | 322 ± 3 |
| Mut14 | 318.46 ± 1.44 | 318 ± 3 |
| Mut28 | NA | 318 ± 3 |
| Mut115 | 309.36 ± 0.41 | 313 ± 3 |
| Δ 1–7 | NA | 305 ± 4 |

*NA* not available data

With reference to Fig. 2 and Table 1, mutations on Arg14 and Arg28 appear to cause just a moderate destabilization of the protein as they determine a little shift of the unfolding temperature $T_u$ to lower values. The mutation of Asn115, 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. 2003, in fact, show that the Asn115Lys mutant is unstable and largely unfolded as compared to the wild-type C5 motif. The NMR spectrum of the Δ 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 thermograms showing that the amino-truncated protein is characterized by the largest shift in $T_u$ thus resulting the most destabilized mutant.

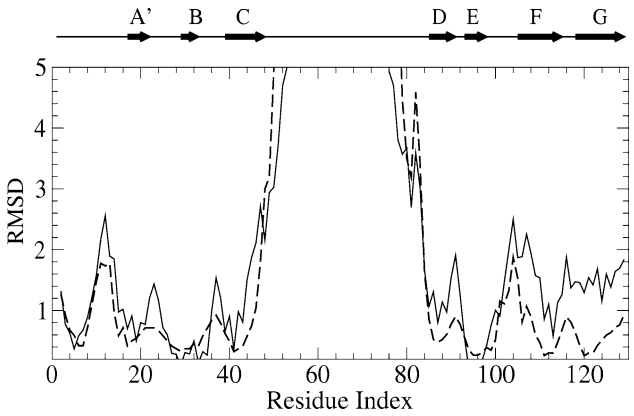


**Fig. 3** Simulated (*solid line*) and experimental (*dashed line*) average RMSD along the 1–52 and 76–130 subsequences outside the CD loop. In order to stress the consistency between the positions of the peaks, the theoretical B-factors have been shifted down by an offset $\delta = 3$ and then rescaled by a factor $\rho = 0.45$

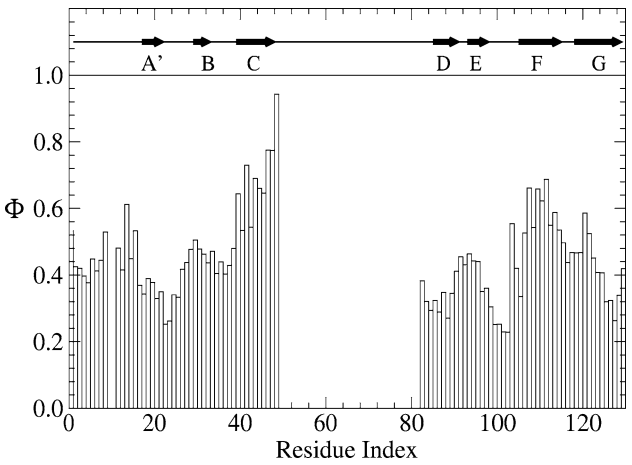


**Fig. 4** Perturbative Φ-values at unfolding temperature $T_u = 322$ K, computed from structures sampled in three windows of overlap corresponding to the unfolded ($0 < Q < 0.15$), Transition State ($0.30 < Q < 0.45$) and folded state ensemble ($0.75 < Q < 0.9$)

To validate the Gō-model approach with the heterogeneous coupling, we checked the consistency of the unfolding temperatures of the mutants with the corresponding experimental temperatures, once the model energy scale $\varepsilon_0$ is tuned such that the wt simulated $T_u$ matches the experimental one. Table 1 indicates that the model reproduces the ranking of the transition temperatures within statistical errors. A further validation of the model comes from computation of the RMSD of the $C_\alpha$-atoms from their positions in the native structure after the structural alignment through Kabsch algorithm (Kabsch 1976). The simulated RMSD profile at 293 K in Fig. 3 reproduces the expected feature of the C5 domain, with loops showing higher mobility than $\beta$-strands. The results are also in qualitative agreement with the average RMSD measurement by Idowu and coworkers (Idowu et al. 2003) for their best ten NMR structures resolved at the same temperature of our simulations.

*Phi*-value analysis allows a characterization of the effect of the three FHC-related mutations. Formula (5) requires a sampling of the folded ($F$), denaturated ($D$) and transition state (TS) conformations during a run at the transition temperature $T_u = 322$ K. The ensembles of structures, $F$, $D$, TS are identified through three windows in the double-well plot of the wt free energy as a function of the overlap $Q$ (fraction of native contacts) at transition temperature $T_u = 322$ K. Domain C5 is characterized by an asymmetric distribution of perturbative Φ-values with the CFGA′ sheet featuring higher values of this parameter as compared to the BDE sheet (Fig. 4).

It is interesting to analyze the role of the three FHC-related mutations with respect to their position in the two $\beta$-sheets of the C5 domain. In fact, Asn115 which lies on the sheet characterized by the highest FEP Φ-values, when mutated to Lys, is experimentally known (Idowu et al. 2003) to completely disrupt the native structure of this protein. On the other hand, Arg14 and Arg28 are located on the sheet with low FEP Φ-values in agreement with the experimental evidence that their mutation does not significantly affect the thermodynamic stability and the folding kinetics.

## Kinetic analysis

The early unfolding kinetics of the wt structure of domain C5 was compared to that of the three missense and Δ 1–7 deletion mutants through $T$-jump simulations. The proteins were first equilibrated at very low temperature $T_{\text{low}} = 225$ K, for $2 \times 10^5$ time steps, and then heated to the temperature $T_{\text{high}} = 322$ K, slightly higher than the unfolding point of the wt. Temperature $T_{\text{high}}$ is sufficiently destabilizing to promote the unfolding of both mutants and wt. The molecules were kept at $T = T_{\text{high}}$ 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 $Q$ still formed. When $Q$ fell below a threshold $Q_u = 0.25$, the protein was regarded as being unfolded. The progress of the denaturation was measured by tracking the fraction of folded trajectories as a function of time, $P_f(t)$, plotted in Fig. 5 for the wt and mutants together with the free energy profiles at temperature $T = T_{\text{high}}$ (inset). Figure 5 shows that all the mutations speed up the unfolding process which however retains a two state character described by Eq. 6. The kinetic behavior is consistent with the free energy profiles (inset of Fig. 5) showing the mutations to reduce the unfolding barrier while simultaneously increasing the folding one. The importance of FHC mutations on the unfolding rate is

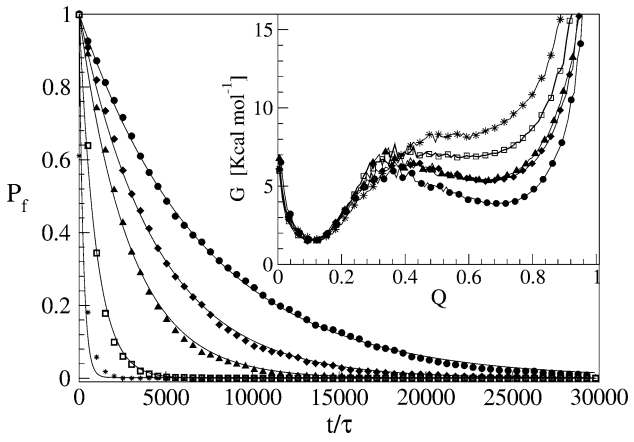


**Fig. 5** Comparison of unfolding kinetics of WT domain C5 with its deletion and missense mutants through the evolution of the fraction of folded trajectories $P_f(t)$. Circles refer to WT, diamonds to Mut14, triangles to Mut28, squares to Mut115, and finally stars to the deletion mutant. The *solid lines* represent the exponential fits according to Eq. 6. *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

quantified by the order: $k_u[\mathrm{WT}] < k_u[14] < k_u[28] < k_u[115] < k_u[\Delta 1{-}7]$.

Further understanding of the unfolding/folding mechanism can be found in Guardiani et al. 2008, where the analysis of the time evolution of native contact probabilities shows the folding to proceed according to the following stages

Stage 1: Formation of FG contacts in proximity of the FG loop.
Stage 2: Appearance of DE hairpin.
Stage 3: Strand C joins the FG hairpin while strand B moves close to the DE hairpin.
Stage 4: Stabilization of C-terminal ends of strands G and A′ and A′B and EF loops in their native conformation.

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 (Fig. 1), in fact, are the first to appear during refolding, as their formation is accompanied with a modest loss in conformational entropy. Of course, that an entropy decrease in the folding pathway of proteins occurs is not surprising. However in the case of C5 domain, following such entropy loss suggests a clear way to rationalize the different impact of the FHC-causing mutations. Residue 115 for example being located in the FG loop, when mutated is likely to hinder the very first events of folding. Residue 28 conversely, is located in the BDE sheet but outside the crucial DE hairpin, and its mutation is therefore expected to have influence only on the structuring of the protein portion near the long A′B, CD and EF loops, thus resulting in a mild phenotype.

In order 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 the contact of mutated residues from native to non-native. The results are summarized on Table 2 showing that the most destabilizing mutations, characterized by $k_u/k_u^{\mathrm{WT}} \geq 5$, are located on strand F (Residues 111, 112, 113), on the FG-loop (Residues 115), on strand E (Residue 93), on the BC-loop (Residue 35) and on strand C (Residue 43). This distribution confirms the conclusions based on 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 caused in FHC by the mutation of residue 115. Our kinetic analysis also underscores the relevance of strands E which is the central element of the BDE sheet and acts as a bridge between strands B and D. Table 2 indicates that mutations of several residues such as 20, 21, 22 (strand A), 44, 46 (strand C) and 127, 128 (strand G) has only a moderate impact on the unfolding rates. This is presumably a consequence of the fact that, as shown in folding/unfolding pathway analysis (Guardiani et al. 2008), these residues all belong to a late folding region of wt, just like residue 28. This may explain the mild clinical phenotype deriving from the mutation of the latter residue.

The data in Table 2 provide the opportunity of a discussion about the way we implemented the mutations to compute Φ-values. We used a simple strategy to introduce mutations within the framework of the Gō-model by turning the native contacts involving the mutated residue into non-native ones. This approach, in which attractive interactions always become repulsive, is clearly an extreme simplification, as real mutations can both stabilize and destabilize a protein. On the other hand, the simulation of the realistic effect of a mutation requires the knowledge of biological information. In the absence of specific indications, the simple replacement of native contacts with non-native ones provides a more unbiased implementation of mutations in the Gō-model context. In this perspective, we have to stress that the method we followed is more similar to a measure of the susceptibility of protein sites to the same kind of perturbations rather than a realistic characterization of mutations. However the information we gain is meaningful to assess the "comparative" physical effects of mutations.

In Gō-like models Φ-values are related to the fraction of native contacts that the *i*th residue is capable of forming in the transition state which can be estimated as $N_{TS}(i) = N_F(i)\,\Phi(i)$. If one or both the factors $N_F(i)$, $\Phi(i)$ gets a low value, the kinetic effect of the mutation is small too. For example, Res42 and Res44, despite their high Φ-values (0.72 and 0.68, respectively), when mutated, result in low unfolding rates ($3.07 \times 10^{-4}$ and $3.01 \times 10^{-4}$, respectively) because

**Table 2** Kinetic simulations: Φ-values, number of native contacts of the mutated residue, number of contacts formed by the residue in the transition state and unfolding rate normalized to the $k_u$(WT) = $1.41 \times 10^{-4}$

| Residue | Strand | Φ | $N_{Nat}$ | $\Phi \cdot N_{Nat}$ | $k_u/k_u^{WT}$ |
|---|---|---|---|---|---|
| Δ1 −7 | H | – | 41 | – | 15.93 |
| 12 | H | 0.48 | 1 | 0.48 | 1.12 |
| 14 | A′ | 0.61 | 3 | 1.83 | 1.60 |
| 20 | A′ | 0.36 | 5 | 1.82 | 2.96 |
| 21 | A′ | 0.32 | 7 | 2.23 | 2.92 |
| 22 | A′B | 0.33 | 3 | 1.00 | 1.43 |
| 28 | A′B | 0.43 | 3 | 1.30 | 2.18 |
| 31 | B | 0.47 | 7 | 3.29 | 4.72 |
| **35** | **BC** | **0.40** | **8** | **3.20** | **6.25** |
| 36 | BC | 0.43 | 4 | 1.71 | 2.50 |
| 42 | C | 0.72 | 4 | 2.90 | 2.18 |
| **43** | **C** | **0.54** | **11** | **5.95** | **8.23** |
| 44 | C | 0.68 | 4 | 2.73 | 2.13 |
| 46 | C | 0.64 | 4 | 2.55 | 1.16 |
| 86 | D | 0.32 | 6 | 1.95 | 1.52 |
| 88 | D | 0.35 | 2 | 0.70 | 1.38 |
| 90 | D | 0.33 | 3 | 0.99 | 1.37 |
| 91 | DE | 0.07 | 3 | 0.22 | 1.95 |
| 93 | E | 0.42 | 6 | 2.52 | 4.82 |
| **96** | **E** | **0.45** | **12** | **5.43** | **11.63** |
| 97 | E | 0.34 | 7 | 2.40 | 3.97 |
| 100 | EF | 0.25 | 4 | 1.00 | 1.36 |
| 104 | EF | 0.55 | 3 | 1.65 | 1.29 |
| 107 | F | 0.53 | 4 | 2.11 | 1.82 |
| 109 | F | 0.23 | 12 | 2.25 | 1.89 |
| **111** | **F** | **0.62** | **10** | **6.18** | **9.93** |
| **112** | **F** | **0.68** | **6** | **4.09** | **5.46** |
| **113** | **F** | **0.54** | **11** | **5.99** | **10.35** |
| **115** | **FG** | **0.53** | **10** | **5.30** | **7.45** |
| 118 | FG | 0.46 | 4 | 1.86 | 2.26 |
| 120 | G | 0.46 | 7 | 3.24 | 3.95 |
| 121 | G | 0.57 | 5 | 2.86 | 2.72 |
| 126 | G | 0.31 | 7 | 2.16 | 3.49 |
| 127 | G | 0.31 | 4 | 1.24 | 2.65 |
| 128 | G | 0.24 | 9 | 2.22 | 2.48 |

The most destabilizing point-like mutations are marked in bold. The unfolding rates of the mutants correlate with the number of contacts that the residue forms in the TS, and that are lost due to the mutation (see Fig. 6)

they form four inter-residue contacts only. On the contrary, Res128 is characterized by a higher contact number $N_{Nat}(128) = 9$ but a low $\Phi\ (128) = 0.24$. This residue thus is important for the protein thermodynamic stability, but has poor influence on kinetics. It seems that a mutation can determine a high unfolding rate when the mutated residue is characterized by both a large number of native contacts

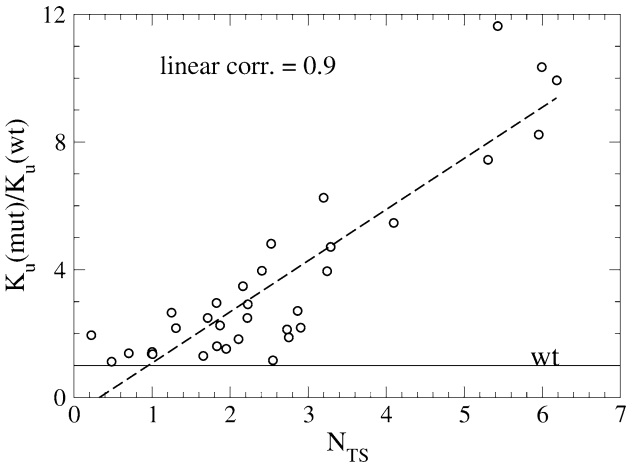


**Fig. 6** Linear correlation plot between the number of native contacts formed by a residue in the transition state and unfolding rate of the corresponding mutant normalized to the WT rate. The horizontal line, $Y = 1$, indicates the WT unitary rate

and a large Φ-value. Therefore, for the domain C5, a suitable parameter for the prediction of the kinetic effect of a mutation on site $i$ is just the product $N_{TS}(i) = N_F(i)\Phi(i)$ which, in fact, shows a linear correlation coefficient $r = 0.89$ with corresponding unfolding rates (Fig. 6).

## Conclusion

The impact of three pathological mutations of domain C5 from MyBP-C related to FHC (Asn115Lys, Arg14His and Arg28His) has been studied through MD simulations. The Ig-like C5 domain was modeled through a Gō force field as, according to Clarke et al. 1999, the members of the Ig-superfamily tend to fold/unfold along a common pathway dictated by the geometry of the common structural core. Thermodynamic and kinetic simulations on the thermal denaturation of the wt and mutants indicate that the most destabilizing mutations are those involving the strand F, and the FG loop which includes the residue 115. In particular Mut115 causes a large decrease in the transition temperature in agreement with the NMR experiments (Idowu et al. 2003). The results of the mutation of residues Arg14 and Arg28, on the other hand, show a more limited shift on the transition temperature suggesting a very little influence of the two residues on protein stability. This potential role of the FHC-related mutations is confirmed by the analysis of Φ-values indicating that the higher Φ's occur on the CFGA′ sheet (including Asn115) and lower values pertain to the BDE sheet (hosting Arg28). The role of the *N*-terminal region of the C5 domain has been also assessed by simulating a deletion mutant where the first 7 residues had been removed. The significant decrease in $T_u$ of the truncated

protein suggests that the *N*-terminal region with its ten-residue long insert typical of the cardiac isoform, is not just a linker between the *C*4 and *C*5 domains, but it gives an important contribution to stability.

The comparison of the fraction of folded trajectories $P_f(t)$ computed through unfolding kinetic (*T*-jump) simulations for the wt and pathologic mutants shows that the mutations determining the largest decrease in the unfolding temperature also induce the highest speed-up 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. In another work (Guardiani et al. 2008) we have shown, that Gō-model depicts the C5-domain folding reaction as mainly entropy-driven process with a specific directionality, starting from the formation of the FG and DE hairpins and concluding to the region corresponding to A′B, CD and EF loops. In that paper, the "folding propagation front" was used to rationalize the impact of pathogenic mutation. That conclusion has been confirmed, here, by the kinetic analysis of a larger number of mutations evenly distributed throughout the molecule, which identified "key regions" overlapping with those found in Ref Guardiani et al. 2008. The kinetic simulation on the C5 domain also revealed an interesting correlation between the unfolding rates and the number of native contacts that a residue forms in the transition state, estimated as the product of the corresponding phi-values and the number of contacts formed in the native states. We suggest that at least in the case of the C5 domain this quantity is a indicator of the kinetic effect of a mutation more reliable than perturbative Φ-values.

## References


Caflisch A, Karplus M (1994) Molecular dynamics studies of protein and peptide folding and unfolding. In: Merz KM Jr, Le Grand SM (eds) The protein folding problem and tertiary structure prediction. Birkhuser, Boston, pp 193–230

Cecconi F, Micheletti C, Carloni P, Maritan A (2001) The structural basis of antiviral drug resistance and role of folding pathways in HIV-1 protease. Proteins Struct Funct Genet 43:365–372

Chang I, Cieplak M, Banavar JR, Maritan A (2004) Thermal unfolding of proteins. Prot Sci 13:2446–2457

Charron P, Dubourg O, Desnos M, Isnard R, Hagege R, Bonne G, Carrier L, Tesón F, Bouhour JB, Buzzi J-C, Feingold J, Schwartz K, Komajda M (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

Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol 6:1005–1009

Cho SS, Levy Y, Wolynes PG (2006) P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. Proc Natl Acad Sci USA 103:586–591

Chung MW, Tsoutsman T, Semsarian C (2003) Hypertrophic cardiomyopathy: from gene defect to clinical disease. Cell Res 13:9–20

Cieplak M, Sulkowska JI (2005) Thermal unfolding of proteins. J Chem Phys 123:194908 1–4

Clarke J, Cota E, Fowler SB, Hamill S (1999) Folding studies of immunoglobulin-like $\beta$-sandwich proteins suggest that they share a common folding pathway. Structure 7:1145–1153

Clementi C, Nymeyer H, Onuchic JN (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

Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. Phys Rev Lett 63:1195–1198

Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzime. 1. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224:771–782

Flashman E, Redwood C, Moolman-Smook J, Watkins H (2004) Cardiac myosin binding protein C: its role in physiology and disease. Circ Res 94:1279–1289

Gō N, Scheraga HA (1976) On the use of classical statistical mechanics in the treatment of polymer chain conformations. Macromolecules 9:535–542

Guardiani C, Cecconi F, Livi R (2008) Stability and kinetic properties of C5-domain of Myosin binding protein C and its mutants. Biophys J 94:1403–1411

Gutin AM, Abkevich VI, Shakhnovich EI (1995) Evolution-like selection of fast-folding model proteins. Proc Natl Acad Sci USA 92:1282–1286

Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. J Mol Graph 14:33–38

Idowu SM, Gautel M, Perkins SJ, 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

Kabsch W (1976) Solution for best rotation to relate two sets of vectors. Acta Crystallogr 32:922–923

Kurochkina N, Lee B (1995) Hydrophobic potential by pairwise surface area sum. Protein Eng 8:437–442

Matysiak S, Clementi C (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

Micheletti C, Cecconi F, Flammini A, Maritan A (2002) Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs. Protein Sci 11:1878–1887

Moolman-Smook J, Flashman E, de Lange W, Li ZL, 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

Niimura H, Bachinski LL, Sangwatanaroj S, Watkins H, Chudley AE, McKenna W, Kristinsson A, Roberts R, Sole M, Maron BJ, Seidman JG, Seidman CE (1998) Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. N Engl J Med 338:1248–1257

Plotkin SS, Onuchic JN (2002) Structural and energetic heterogeneity in protein folding I. Theory J Chem Phys 116:5263–5283

Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, Benaiche A, Isnard R, Dubourg O, Burban M, Gueffet JP, Millaire A, Desnos M, Schwartz K, Hainque B, Komajda, M (2003) Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation 107:2227–2232

Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D (1999) Experiment and theory highlight role of native state topology in sh3 folding. Nat Struct Biol 6:1016–1024

Serrano L, Kellis J, Cann P, Matouschek A, Fersht AR (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

Shakhnovich EI (2006) Protein folding thermodynamics and dynamics: where physics, chemistry and biology meet. Chem Rev 106:1559–1588

Sutto L, Tiana G, Broglia RA (2006) Sequence of events in folding mechanism: beyond the Gō model. Protein Sci 15:1638–1652

Winegrad S (2000) Myosin binding protein C, a potential regulator of cardiac contractility. Circ Res 86:6–7

Zhou H, Zhou Y (2004) Quantifying the effect of burial of amino acid residues on protein stability. Proteins Struct Func Bioinf 54:315–322