Hydrophobic confinement modulates thermal stability and assists knotting in the folding of tangled proteins

João Especial[1], Ana Nunes[1,2], Antonio Rey*[3] & Patrícia FN Faísca*[1,2]
[1]Departamento de Física and [2]BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, 1749-016 Lisboa, Portugal
[3]Departamento de Química Física, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain

*Corresponding author:
PFNF, email: pffaisca@fc.ul.pt
AR, email: areygayo@ucm.es

Abstract
There is growing support for the idea that the in vivo folding process of knotted proteins is assisted by chaperonins, but the mechanism of chaperonin-assisted folding remains elusive. Here, we conduct extensive Monte Carlo simulations of lattice and off-lattice models to explore the effects of confinement and hydrophobic intermolecular interactions with the chaperonin cage in the folding and knotting processes. We find that moderate to high protein-cavity interactions (which are likely to be established in the beginning of the chaperonin working cycle) cause an energetic destabilization of the protein that overcomes the entropic stabilization driven by excluded volume, and leads to a decrease of the melting temperature relative to bulk conditions. Moreover, mild-to-moderate hydrophobic interactions with the cavity (which would be established later in the cycle) lead to a significant enhancement of knotting probability in relation to bulk conditions while simultaneously moderating the effect of steric confinement in the enhancement of thermal stability. Our results thus indicate that the chaperonin may be able to assist knotting without simultaneously thermally stabilizing potential misfolded states to a point that would hamper productive folding thus compromising its functional role.

Keywords: Monte Carlo, folding, knotting, structure-based model, hydrophobicity, confinement

Introduction

Knotted proteins are proteins that embed a physical (i.e. open) knot in their native structure. Although the first knotted protein was reported in 1977[1], it was only after the development of knot detection methods[2, 3], and loop closure procedures[4], that these intricate systems came into the spotlight. Two recent surveys of the Protein Data Bank (PDB) indicate that the fraction of knotted proteins stands between 0.71%[5] and 0.77%[6]. The most frequent knot type found in the PDB is the 3_1 (or trefoil) (92.1%), followed by the 4_1 (4.8%), 5_2 (or twisted-three) (2.8%) and 6_1 (0.3%)[6].

The first studies addressing the challenge of understanding how knotted proteins fold date back from 2005[7]. Since then, a plethora of contributions, both experimental and computational (reviewed in [8-10]), have significantly advanced our understanding of this intricate folding mechanism. Although a few studies have been carried out for model systems embedding the twisted-three[11-14] and 6_1 knot[15, 16], the majority of contributions have focused on proteins carrying the most frequent trefoil knot. The current picture emerging from both simulations[17-21] and experiments[7, 22-25] points to a complex folding process for trefoil knots, whereby part of the chain firstly forms a knotting loop, which is later protruded by the reminder of the polypeptide chain via a slipknotted conformation[18, 26-30] or via a direct threading motion of the terminus[17, 19, 20, 31-33], in the entropically costly[34], and rate-limiting[13, 25] knotting step.

Given the complexity of their folding process, knotted proteins typically exhibit slow folding rates both *in vitro* and in simulations (reviewed in[8]). Because the process of knot formation is ordered (in the sense that it should follow a specific sequence of events), it is prone to backtracking (i.e. to the breaking and re-establishment of specific native contacts)[26, 32], and to the formation of trapped intermediates. Some of these may lead to aggregation: there is strong evidence that intermediate states populated by protein UCH-L1, which embeds a 5_2 knot in its native structure, can be aggregation prone[14]..

In vivo, efficient folding of many newly synthesized proteins is achieved through the utilization of a special class of molecular machines called chaperones. The so-called chaperonins comprise an important and universal class of chaperones, which have the unique ability to fold some proteins that cannot be folded by simpler chaperone systems. The most well studied chaperonin is the bacterial GroEL-GroES complex from *E. coli*. GroEL is a cylindrical protein complex comprising two heptameric rings (of ~57 kDa units) stacked back-to-back forming a large cage-like structure. GroES is a smaller heptameric ring with seven identical (10 kDa) subunits that caps the ends of the GroEL cylinder[35]. During the GroEl-GroES operating cycle a non-native substrate protein (SP) binds to the entrance of the GroEL cavity (i.e. the so-called apical domain) within a region of concentrated hydrophobic residues. Subsequently, the binding of seven ATP molecules (one per GroEL subunit), and of GroES, leads to a large conformational change whereby the cavity's volume nearly doubles (~175.000Å^3), being able to encapsulate a SP with up ~60KDa (~545 residues long). Concomitantly, the cage's inner wall changes its physical properties becoming hydrophilic and net-negatively charged[36].

The encapsulated SP is free to fold in this environment for ~6 seconds, which is the time needed for ATP hydrolysis[36]. The exact mechanism by which the chaperonin assists folding is still a matter of debate, but two main models have been proposed. The passive (or Anfinsen's) cage model predicts that the role of chaperonin is simply that of increasing folding yield by avoiding aggregation. One the other hand, the active cage model envisages an active role for the chaperonin, according to which the latter is able to accelerate folding process by modulating the free energy landscape [36].

By using a cell-free translation system, Mallam and Jackson were the first to explore the role of GroEL-GroES on the folding and knotting mechanisms of trefoil proteins YibK and YbeA. In the presence of GroEL-GroES, folding rate starting from unfolded (and presumably unknotted) conformations becomes at least 20-fold higher. Furthermore, GroEL-GroES has no effect on the folding rate of knotted denatured conformations. Since none of those proteins populates misfolded species during or after translation[37], these results indicate that the chaperonin is likely to play a *specific* and active role in assisting the knotting step. These observations led Mallam and Jackson to propose that the folding *in vivo* of knotted proteins YibK and YbeA is assisted by chaperonins[37]. If this is so, it is important to understand its mechanistic underpinnings.

Previous accounts, including seminal contributions by Chan and Dill[38] and Betancourt and Thirumalai[39] (both with on-lattice models), and more recent studies framed in off-lattice simulations and different force fields, addressed the chaperonin assisted folding of topologically regular proteins[40-46]. A general conclusion of these studies is that steric confinement accelerates folding by stabilizing the transition state relative to the unfolded state at the transition temperature. Piana and Shaw [47] performed the very first study of chaperoning-assisted folding framed on unbiased, atomistic molecular dynamics simulations based on a state-of-the-art force field that more accurately reproduces the aggregation propensity of unfolded states. The authors focused on the C-terminal fragment of the villin headpiece (HP35) featuring 35 residues. They observed a decrease of folding rate due to intermolecular interactions between the protein and disordered residues in the GroEL's C-terminal tails. A monotomic decrease in folding rate was also observed for a triple point mutant of the small (120 residues), single domain protein FynSH3 in the context of a coarse-grained C-alpha model and Langevin dynamics, when intermolecular attractive interactions were considered between the protein and GroEL-GroES cavity [48]. Similarly, Sirur and Best [45] reported a strong decrease in folding rate of protein Rhodanase (296 residues long) when specific residue-residue interactions between the protein and the GroEL cavity were modeled by means of a carefully parameterized, sequence-based potential combined with a C-alpha protein representation. These results indicate that intermolecular interactions with the chaperonin cage can offset the effects of steric confinement and reduce the folding rate. Therefore, interactions other than excluded volume should be in order to more accurately capture chaperonin-assisted folding

By using structured-based models and Monte Carlo simulations, Soler *et al.* gave first steps towards understanding the folding mechanism of knotted proteins inside the chaperonin cage[49]. In particular, these authors explored the role of steric confinement in

the folding transition of the smallest (82-residue) knotted trefoil protein in the PDB, MJ0366 (PDB ID: 2efv), from Methanocaldococcus jannaschi. They found that by reducing the entropic barrier to knotting, steric confinement alone is able to significantly enhance knotting frequency, even under temperature conditions in which that native state becomes thermodynamically destabilized, most likely because confinement reduces the frequency of local interactions, thereby decreasing local rigidity of the chain, in line with results reported in [50].

In a subsequent study, Sulkowska *et al.* reported analogous findings for VirC2, a small (121-residue) protein similar to MJ0366, and the (109-residue) DndE protein, with an unclassified fold[51]. In a more recent follow-up, Sulkowska *et al.* investigated in detail the folding under steric confinement of protein Ubiquitin C-terminal Hydrolase 1 (UCH-L1), which embeds a 5_2 knot in its native structure. The authors reported that depending on the degree of steric confinement two different folding pathways can be populated, leading to a fast (weak confinement) or to a slower (strong confinement) folding transition [52].

While steric confinement certainly plays a role in chaperonin-assisted folding, by decreasing the entropic barrier to knotting, Lim and Jackson recently proposed that the chaperonin may play an additional active role by 1) further stabilizing key interactions that promote knotting, and, 2) by facilitating the unfolding of topologically trapped or misfolded states[25]. Motivated by these premises, here we explore the mechanism of chaperonin-assisted folding beyond excluded volume interactions by focusing also on the effects of hydrophobic confinement. As before[49], we start by performing extensive and systematic Monte Carlo simulations of two lattice systems embedding a shallow trefoil and a shallow twisted-three knot, respectively. Subsequently we challenge the lattice predictions in the context of a C-alpha, off-lattice model of protein MJ0366, which also embeds a shallow trefoil knot

Despite their simplicity, structured-based models like the ones employed in the present work have been successful in the study of protein folding (reviewed in [53-56]) and, more recently in exploring other phenomena involving proteins such as aggregation[57-59], protein fold switches[60], and phase separation[61], just to mention a few examples. By using coarse-grained protein representations and coarse-grained models for protein energetics the present study seeks to reflect the semi-quantitative behavior, and offer mechanistic insights into the effects of hydrophobic interactions in chaperonin-assisted folding

Our analysis predicts that intermolecular hydrophobic interactions establishing between the folding protein and the chaperonin cage modulates the folding transition in different ways: 1) It decreases the transition temperature, 2) It decreases the activation energy of folding by creating a rugged free energy landscape and a rugged transition state ensemble[62], and 3) For some model sequences, moderate to high degree of confinement may lead to the formation of structurally consolidated, yet unknotted, intermediate states. Most important, however, is the observation that mild to moderate hydrophobic confinement significantly increases knotting probability in relation to the bulk without enhancing thermal stability to a point that would hamper productive folding. This is a significantly important result because it means that, by taking into account hydrophobic

interactions, the chaperonin is be able to assist knotting without compromising its functional role.

Models and Methods

Lattice model

We use a simple lattice model in which amino acids are reduced to beads of uniform size placed on the vertices of a cubic lattice. To satisfy excluded volume constraints only one bead is allowed per lattice site, and two consecutive beads along the chain are linked by sticks (one lattice spacing long) that represent the peptide bond.

Intramolecular interactions are modelled with the native centric Gō potential[63], i.e., the total energy (in reduced units) of a conformation with bead coordinates $\left\{\vec{\mathbf{r}}_i\right\}$ is given by

$$E\left(\left\{\vec{\mathrm{r}}_i\right\}\right) = \varepsilon \sum_{i,j>i+2}^{N} \Delta\left(\vec{\mathrm{r}}_i - \vec{\mathrm{r}}_j\right), \qquad (1)$$

where N is the chain length measured in number of beads, ε is the uniform interaction energy parameter (equal to -1 in this study), and the contact function Δ is unity, only if beads i and j form a native contact, i.e., a contact that is present in the native structure.

Confinement is modelled by placing the lattice protein inside an elementary geometry. For symmetry reasons we choose a rigid cubic box of linear size L (measured in lattice units), which restricts the conformation of the lattice system and its movements in three dimensions.

To model the intermolecular interactions between the protein and the confining box two scenarios are considered. In the first case, there are only excluded volume (i.e. steric) interactions. In the second case, which is inspired by hydrophobic-polar (HP) model[64], stabilizing hydrophobic interactions may be established between the protein's hydrophobic beads and the box's walls.

The hydrophobic residues in natural proteins can be distinguished on the basis of their hydropathy index (e.g., isoleucine, valine and leucine are clearly the most hydrophobic)[65]. However, in the context of the present work, and in line with the HP model, beads representing hydrophobic residues and the box's inner lining will be taken uniformly hydrophobic. Different interaction strengths for the protein-box intermolecular interaction will be explored. When considering hydrophobic confinement, the total energy of the protein-box system, E_T, is thus given by

$$E_T\left(\left\{\vec{\mathbf{r}}_i\right\},\left\{\sigma_i\right\}\right) = \varepsilon \sum_{i,j>i+2}^{N} \Delta\left(\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_j\right) - \varepsilon_{HP} \sum_{i=1}^{N} \delta\left(\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_L\right), \qquad (2)$$

where $\{\sigma_i\}$ represents the ensemble of hydrophobic residues in the protein, $0 < \varepsilon_{HP} \leq 1$ is the hydrophobic interaction parameter that captures intermolecular hydrophobic interactions, and δ is unity if bead i is one lattice spacing away from the box's walls located at $\vec{\mathbf{r}}_L$.

Off-lattice model

In the off-lattice model an amino acid is reduced to a single bead located in the position of its corresponding C-alpha atom. Neighbour residues along the sequence are connected by rigid bonds of distance equal to 3.8 Å (the length of a trans peptide bond in real polypeptide chains). All residues have the same radius $b = 2.1\text{Å}$, which controls a hard sphere potential to avoid overlapping of beads among themselves, or with the walls of the confining cavity. The details of the model have been previously described [66, 67], and only its main features will be highlighted here.

As in previous work, a native interaction between two residues i and j is introduced through an attractive harmonic potential:

$$u_{ij}(r_{ij}) = \begin{cases} \varepsilon\left[\dfrac{(r_{ij} - d_{ij}^{nat})^2}{a^2} - 1\right] & \text{if} \quad d_{ij}^{nat} - a < r_{ij} < d_{ij}^{nat} + a \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where r_{ij} is the distance between the residues in a given conformation, d_{ij}^{nat} is the corresponding distance in the native structure, ε is the uniform interaction energy parameter (equal to -1 in this study), and provides the depth of the attractive well, and a controls its width. The use of a Monte Carlo sampling procedure (as outlined below) allows for the use of a simple continuous but not differentiable potential as that described by eq. (3). It has the advantage over standardly used Lennard-Jones potentials that we can easily control the potential width without affecting the repulsive term, something which has an important influence in the thermodynamic characteristics of the folding transition[68] . From our previous experience[59, 68, 69] we used $a = 0.6$ Å in this work. Local interactions (with $|i - j| = 3$ or 4) use the same energy term, considering that all the native distances corresponding to virtual bond angles and virtual torsional angles are always native, irrespective of their actual values. The latter term (torsions) uses a chirality check to prevent the model folding into the mirror conformation to the native state.

In addition, the hydrophobic residues, defined according to the protein sequence (see below), interact with the wall of the spherical cavity, whose radius is R_c. To keep the model as simple as possible, we have used for these interactions the same type of truncated harmonic well described in equation (3). However, instead of an intramolecular distance we use here the radial coordinate r_i of each bead i with respect to the cavity centre, resulting:

$$u_{HP}(r_i) = \begin{cases} \varepsilon_{HP}\left[\dfrac{(r_i - r_{max})^2}{a^2} - 1\right] & if \quad r^{max} - a < r_i \\ 0 & otherwise \end{cases} \qquad (4)$$

and $r_{max} = R_c - b$, the radial coordinate for a model bead in direct contact with the wall. The resulting attractive potential is rather narrow, as it is for the intramolecular interactions. The values of ε_{HP} used in this work, in units of ε, range from 0 (plain repulsive walls) to 2.5. The range is larger than that for the lattice calculations, but we must recall that the potential is continuous off-lattice, with values of every individual interactions going from ε_{HP} to zero as a function of the radial coordinate.

We do not consider stronger intermolecular interactions with the cavity in the lattice model and in the off-lattice one because, as we will see below, the folding transition breaks down at ε_{HP}=0.6 on lattice and ε_{HP}=2.35 off-lattice. Stronger intermolecular interactions with the wall will hamper folding even more, and for strong enough interactions the native structure will never be populated because the chain will spend most of its time adhered to the cavity in unfolded conformations.

Monte Carlo folding simulations

The computational methods and techniques adopted in the present work have been extensively used before and are described in detail in previous publications[19, 49, 70-72]. Here we provide a summary and highlight a few details.

The conformational space is explored with the Monte Carlo method with Metropolis sampling[73]. For the lattice model the move set comprises end-, corner-flips and crankshaft moves, while for the off-lattice exploration of the conformational space proceeds via end- and spike moves, and displacement moves which shift a part of the chain (from a randomly selected bead to the chain end) a maximum distance of a virtual bond.

Simulations start from an unfolded conformation and evolution towards the native conformation is followed by registering a series of properties that include, total energy

(E), intra (E_I) and intermolecular (E_{HP}) molecular energy, fraction of native interactions (Q), and gyration radius (R_g), and knotting state. In order to accurately sample equilibrium distributions at different temperatures we conducted long replica-exchange (RE) simulations[74]. The temperature grid and swapping time were both chosen to ensure fast and good convergence of the data towards equilibrium. In particular, the adopted swapping time between neighboring replicas was larger than the largest auto-correlation time. The raw data from RE was used to compute the reduced heat capacity ($C_v = (< E_T^2 > - < E_T >^2) / T^2$) as a function of temperature (T), and the weighted-histogram-analysis method (WHAM)[75] was used to project the free energy on selected reaction coordinates (e.g. Q). In the case of the off-lattice model, instead of using Q we use the energy normalized to the native energy (E/E_{NAT}) because the definition of a native contact is somewhat arbitrary off-lattice. The use of Q as a reaction coordinate to study folding in the presence of confinement was reported in previous simulation studies [49, 51, 52], which included models where interactions with the chaperonin cage extend beyond excluded volume[45]. Indeed, Q (or E/E_{NAT}) should be suitable reaction coordinates as long as confinement conditions do not severely hamper or actually impede the folding transition

The melting temperature, T_m, is the temperature at which the C_v peaks, and coincides with the folding transition temperature, T_f, when the system's two-state transition is exclusively driven by intramolecular interactions. When intermolecular interactions with the wall are considered, T_f is the temperature at which the free energy displays two equal minima in the unfolded and native states [42]. T_m is generally considered a measure of protein's thermal stability.

To investigate the knotted state of a sampled conformation we used an adopted version of the Koniaris–Muthukumar–Taylor (KMT) algorithm[3] for lattice models[19], and results were compared with those obtained with that from Tubiana et al[76, 77]. Folding kinetics was measured from 1000-2000 MC simulations at fixed temperature. The corresponding folding times (i.e. first passage times) allow evaluating the distribution of proteins which remain unfolded as a function of MC 'time' (measured in number of MC steps). The folding rate is given by the slope of the linear fitting of this distribution to a single-exponential decay[71].

Results

Model systems and design/decoration of the lattice sequences

In this work we consider a lattice protein with chain length $N = 41$ (knot 3_1), which was designed by hand to embed a trefoil knot in its native structure (Fig. 1A), and a similarly designed lattice protein with chain length $N = 52$ (knot 5_2), whose native structure embeds a twisted-three knot (Fig. 1B). These models have been explored in previous studies. For the off-lattice model we consider protein MJ0366 (PDB ID: 2efv), from the thermophilic methanogen Methanocaldococcus jannaschii (Fig. 1C), which is 82 amino acids long, being the smallest knotted protein found so far. Its native state embeds a

trefoil knot. The knotted core (KC), i.e., the minimal segment of the chain that contains the knot, is highlighted (in green) in the three-dimensional representation of the three model proteins. All knots are classified as shallow because it suffices to remove a few beads from one of the termini (three in the case of knot 3_1, two in the case of knot 5_2, and 11 in the case of MJ0366) to untangle the protein.

Following the hydropathy scale of Kyte and Doolitle [65] we consider seven hydrophobic amino acids, namely, Ile, Val, Leu, Phe, Cys, Met, and Ala.

Since lattice models do not represent specific proteins, it is necessary to decide how many beads will be tagged hydrophobic, and how to distribute them along the chain (i.e. for the knotted lattices it is necessary to design/decorate the sequences following some physically motivated criterion). In order to do so, we started by analysing the ensembles of 3_1 (606 entries) and 5_2 (19 entries) knotted proteins in the PDB[6] to determine the fraction of hydrophobic residues (i.e. the number of hydrophobic residues normalized by the chain length) for each knot type. We found this number to be 34% in the case of knotted trefoils, and 39% in the case of twisted-three knots. Therefore, 14 beads of the knot 3_1 are considered hydrophobic, and 20 beads of knot 5_2 are hydrophobic as well. To decide which particular beads along the chain are hydrophobic, we reason that hydrophobic residues in a natural protein should establish a large number of intra-molecular interactions in the native state, to avoid interactions with the solvent. Therefore, hydrophobic beads should have many neighbours in the native state.

We thus determined the number of intra-molecular native contacts per bead to identify the so-called contact core (CtC) beads, i.e., those beads that establish the higher number (i.e. three or four) of native contacts. In knot 3_1 there are 14 (34%) CtC beads (SI: Fig. 1A), while this number increases to 19 (36.5%) in knot 5_2 (SI: Fig. 1B). Perhaps not surprisingly, and in line with our premise, these numbers are consistent with the fraction of hydrophobic residues found in the PDB populations of trefoil and twisted-three knots. Sequence CtC01 of knot 3_1 features 14 hydrophobic beads (Fig. 1D and SI Fig. 1A), while sequence CtC01 of knot 5_2 contains 19 hydrophobic beads (Fig. 1E and SI Fig. 1B), coinciding in both cases with the CtC beads. Taking these original sequences as templates, we then prepared a new sequence for each lattice knot by making neutral the two most connected hydrophobic residues of the original sequence, and making hydrophobic two neutral beads (which have less than three contacts) of the original sequence. The newly produced sequences are used as templates, and the procedure is iterated, generating sequences where the hydrophobic residues are progressively removed from the Ctc beads. Eventually we obtained 13 additional sequences for knot 3_1 and 16 additional sequences for knot 5_2. However, as we will discuss later, not all these sequences are able to fold under hydrophobic confinement.

In the case of natural occurring protein MJ0366, 31 (out of the 82) residues are hydrophobic (Fig. 1F).

Steric confinement increases thermal stability

In a previous work, we explored the effect of steric confinement in the folding transition of knot 3_1 and protein MJ0366[49]. Here, we extend this analysis to knot 5_2 and compare the results of both lattice model systems. The largest dimension of the lattice system limits the smallest size of the box to $L = 6$ (knot 3_1), and $L = 7$ (knot 5_2) lattice units (Fig. 1G); in the case of MJ0366, the radius of the confining sphere is limited to $d=2R_c=2.94R_g$ (Fig. 1C, H), R_g being the radius of gyration of the native protein.

We start by noticing an increase of the melting temperature T_m (Fig. 2) with confinement for all model systems, in tandem with the broadening of the C_v *vs.* T curve (SI: Fig. 2A-C). These effects are not specific of knotted proteins and have been reported for topologically trivial model systems[40, 43, 46, 78]. They result from the fact that under confinement the system is no longer allowed to populate spatially extended conformations (i.e. conformations with Q~0). Consequently, the unfolded ensemble becomes thermodynamically destabilized relative to the folded state (leading to an increase in T_m), and energy fluctuations become smaller (broadening of the C_v curve).

Knotted trefoil folds faster than twisted-three knot under steric confinement

The analysis of the free energy profiles shows that the folding transition keeps its two-state character as box size decreases for all model systems (SI: Fig. 2D-F). However, a shift of the unfolded state towards higher Q (or E/E_{NAT} in the case of the off-lattice model) is observed because confinement forces the population of unfolded conformations with residual native structure; this, in turn, leads to a less cooperative folding transition and also contributes to broadening the C_v curve. The activation energy of folding decreases (SI Fig. 2D-F), and as expected from transition state theory (TST), the folding rate should increase. Since it is computationally expensive to accurately evaluate the folding rate for the off-lattice model[79], here, we restrict the kinetic analysis to the two lattice knots. We find that in line with TST, steric confinement indeed accelerates the folding kinetics of both knot 31 (76%) (Fig. 3A) and knot 5_2 (48%) (Fig. 3B) relative to bulk conditions. The folding rate enhancement, however, is highest for the steric box of linear size $L = 8$, presumably because smaller boxes lead to backtracking of compact malformed conformations[32, 49, 80]. Since knot 3_1 is smaller than knot 5_2, and chain length is a well-known determinant of folding rate[70, 81], it is not strictly correct to directly compare the rate enhancement upon confinement for both knotted topologies. We therefore considered two unknotted model systems, constructed from the knotted ones by appropriately manipulating their backbone connectivity. We can see (Figs. 3C and 3D) that upon confinement the folding rate enhancement of unknot 5_2 (117%) is more than twice the one registered for its knotted counterpart, while in the case of unknot 3_1 the

enhancement (79%) is similar. This indicates that while the two model systems exhibit similar qualitative kinetic responses upon steric confinement, the existence of a complex knot in the native state strongly impairs the accelerating effect of confinement. In other words, the rate enhancement is indeed larger for knot 3_1.

Hydrophobic confinement decreases thermal stability

Here we proceed by looking into the effect(s) of hydrophobic interactions established between the chaperonin's confining cage and the folding chain. We recall that during the operating cycle the environment inside the chaperonin's cage changes from hydrophobic (in the beginning of the cycle) to hydrophilic. Therefore, proteins entering the cage are more likely to establish intermolecular hydrophobic interactions with the chaperonin, which may lead to tethered, or even adhered conformations. However, as the cycle progresses, the strength of hydrophobic interactions decreases, so that the number of such intermolecular interactions decreases as well and the protein becomes free to fold. In the context of the adopted model system, a decrease of the intermolecular parameter ε_{HP} will be taken as an indication of the progression of the chaperonin cycle, with $\varepsilon_{HP} = 1.0$ corresponding to the initial (hydrophobic) phase and $\varepsilon_{HP} = 0.0$ to the final (hydrophilic) phase.

We started by evaluating T_m as a function of ε_{HP} for sequence Ctc01 of knot 3_1 (Fig. 4A), knot 5_2 (Fig. 4B), and protein MJ0366 (Fig. 4C) under several confinement conditions. We observe that T_m decreases with ε_{HP}, even for large box sizes (i.e. of linear dimension larger than that of extended chains), and that it drops below T_m observed under bulk conditions (T_{BULK}) when $\varepsilon_{HP} \sim 0.5$ for the two lattice models, and $\varepsilon_{HP} \sim 2.2$ for the off-lattice system. It can be seen from the curves corresponding to L=100, i.e., a box of linear dimension largely exceeding the size of the fully extended (i.e. denatured) chain (Figs. 4A and 4B) that $\varepsilon_{HP} \sim 0.5$ is also the value for which interactions with the walls start to become relevant, destabilizing the native state as adhesion to the walls becomes dominant in the unfolded state. This effect is more pronounced in the smaller boxes, where binding to the wall is more likely. For smaller values of ε_{HP}, the energy of the unfolded state remains unchanged, and we see only the effect of steric confinement in the entropic stabilization of the native state. Qualitatively similar results were reported for an all beta-sheet protein composed of 46 residues and a smaller, 27 residue peptide protein with an alpha/beta fold[78] in off-lattice simulations. We notice that for the smallest box size (L=8) tested on-lattice the decrease in T_m is particularly sharp. This indicates that for these specific model sequences moderate stabilizing energetic interactions with the chaperonin outcompete intramolecular interactions, which causes an energetic destabilization of the native state. Apparently, this destabilization is large enough to counterbalance the thermodynamically driven increase in T_m as a result of strict steric confinement (Fig 2). This behavior is physically sound because if the role of the chaperonin is to assist folding,

then it should provide a physical environment where misfolded states become thermally destabilized and easily unfold.

In order to investigate the universality of this behavior we extended our analysis to all other sequences of the CtC family. We notice (SI: Fig. 3) that for several sequences (e.g. CtC14 of knot 3_1 (SI Fig. 1A) and CtC17 of knot 5_2 (SI Fig. 1B), T_m increases systematically as the interactions with the wall become progressively more stable, even for a moderately large confining box (L=33). These sequences display a significant number of their hydrophobic beads on the lattice surface, and the corresponding intermolecular interactions with the confining box may contribute to energetically stabilize folded (or nearly folded) conformations causing T_m to increase. Such sequences are not good models to study folding under hydrophobic confinement within the context of our model as they are likely to keep a significant number of intermolecular interactions, even for small values of the intermolecular parameter, ε_{HP}, i.e., as the cycle approaches the hydrophilic phase. This implies that the protein does not become progressively detached from the chaperonin along its working cycle. Furthermore, the existence of many HP residues on the protein's surface would make these chains prone to aggregation. In order to explore this hypothesis, we determined the distribution of the number of intermolecular contacts as a function of ε_{HP} for a 'good' (e.g. Ctc0) and a 'bad' (e.g. Ctc14 of knot 3_1 and Ctc17 of knot 5_2) sequence in a box of size L=8 and compare the results with those exhibited by protein MJ0366. In our calculations the temperature is held fixed at an arbitrary temperature T=0.6 (i.e. it is the same for all model sequences and strength of ε_{HP}), which means that the system may be effectively above or below T_m depending on ε_{HP} (Fig. 4). Our results show that sequences Ctc01 of knot 3_1 (SI Fig. 4A) and knot 5_2 (SI Fig. 4B) indeed display a folding-like behavior populating conformational states with many intermolecular interactions (and few intramolecular ones) when $\varepsilon_{HP} \geq 0.5$ (i.e. in conditions where the native state is destabilized), and few intermolecular interactions (and many intramolecular ones) when $\varepsilon_{HP} < 0.5$ (i.e. in conditions where the native state is stabilized). On the other hand, sequence Ctc14 of knot 3_1 (SI Fig. 1A and SI Fig. 4C), and sequence Ctc17 of knot 5_2 (SI Fig. 4D), populate conformations with many (>10) intermolecular contacts even when $\varepsilon_{HP} = 0.1$. For these sequences the simulation temperature is always below their T_m (ranging between 0.73 and 0.65), i.e., in conditions that favor the population of the native state. These sequences would not be able to leave the chaperonin towards the end of its cycle. The results obtained for protein MJ0366 (at fixed temperature T=0.58) are qualitatively similar to those of sequence Ctc01, with the protein populating conformations with progressively less intermolecular interactions as ε_{HP} decreases (SI Fig. 4E). Based on the above, in what follows, we will restrict the lattice simulations to sequence Ctc01.

Hydrophobic confinement enhances knotting probability

A major result from our previous study[49] is the prediction that steric confinement significantly enhances the knotting probability of protein MJ0366 for small sizes of the confining sphere (i.e. R_c=2.0R_g). A similar behavior could not be recapitulated on-lattice because the discrete nature of the representation does not allow accessing equivalently small confinement sizes. Thus, we focus our attention on protein MJ0366 confined to a sphere of radius R_c=2.0R_g. We evaluate the knotting probability, P_{knot}, in ensembles of (~5000) conformations collected from the RE simulations for several values of ε_{HP}, firstly as a function of temperature (Fig. 5A), and then as a function of the reaction coordinate E/E_{nat} (i.e. a proxy of fraction of native contacts off-lattice) at T_m (Fig. 5B). We notice that hydrophobic confinement neither facilitates or hampers knotting below T_m (i.e. in conditions that thermodynamically stabilize the native state) for $\varepsilon_{HP} \leq 2.1$. However, for higher values of ε_{HP} knotting becomes less likely under the same temperature conditions, because the threading terminus becomes tethered to the cage by the terminal bead that is hydrophobic (SI Fig. 5).

More interestingly, however, is the dependence of P_{knot} on the reaction coordinate. If the interaction strength is high ($\varepsilon_{HP} = 2.35$), knotting probability keeps low (and is lower than in the bulk) until very late in folding, in line with lattice results (SI: Fig. 6). In this case the number of established intermolecular contacts/interactions can become rather large (SI Fig. 4), and the conformational space is dominated by states that are significantly stabilized by intermolecular interactions. The latter hamper folding to the native structure, and, accordingly, the native state becomes highly thermodynamically destabilized (Fig. 6 B and C). These conformations are unknotted because the interaction of the termini with the cavity does not allow for threading events, which forces knotting to occur only in native conformations.

However, the knotting probability increases substantially relative to the bulk scenario in pre-TS (i.e. E/E_{nat}<0.5) conformations under moderate hydrophobic conditions (i.e. $1.9 \leq \varepsilon_{HP} \leq 2.1$), and across the whole folding process for mildly hydrophobic conditions (i.e. $\varepsilon_{HP} = 1.75$). The corresponding melting temperature is always smaller than that observed under steric confinement alone, and, for $\varepsilon_{HP} = 2.1$, it is indeed identical to that observed in the bulk, which means that mild to moderate hydrophobic confinement increases knotting probability in relation to the bulk while at the same time moderating the effect of steric confinement in the enhancement of thermal stability This is a remarkably interesting result because it means that the chaperonin will be able to assist knotting without simultaneously thermally stabilizing potential misfolded states to a point that would hamper productive folding.

Hydrophobic confinement enhances the population of unknotted intermediate states

If the chaperonin should play an active role in the folding process, one would expect to observe changes in the folding pathways. Here we investigate if intermolecular hydrophobic interactions modulate the free energy profiles (at T_m) of the studied model systems for different strengths of the intermolecular energy parameter, ε_{HP}. As before, we keep the box size L=8 for the lattice simulations and the radius of the confining sphere $R_c=2R_g$.

In the case of knot 3_1, the folding transition of sequence Ctc01 keeps its two-state character across all tested values of ε_{HP}, with the activation energy of folding decreasing slightly for $\varepsilon_{HP}=0.7$, and the TS shifting also slightly towards higher Q (Fig. 6A). However, in the case of knot 5_2, there is a complete breakdown of the two-state folding transition when $\varepsilon_{HP}\geq 0.6$ with highly stable intermediate states being populated towards late folding (Fig. 6B). This behaviour is qualitatively recapitulated by protein MJ0366, whose free energy landscape becomes dominated by intermediate states when intermolecular interactions with the chaperonin are distinctively strong (i.e. in the beginning of the chaperonin cycle) and the folding transition breaks down (Fig. 6C). We note that as long as the protein is able to fold into the native structure (i.e. $0\leq\varepsilon_{HP}<2.25$) the off-lattice model exhibits a shift of the TS towards high E/E_{NAT}, i.e., it becomes more nativelike, in line with the lattice results. The probability maps for native contacts evaluated over the ensemble of conformations representative of the TS (SI. Fig. 7) show that as ε_{HP} increases helices 1 and 4 get progressively more consolidated, but helices 2 and 3 keep significantly distorted. The native interactions between the C-terminal helix 4 and its surroundings also get more likely indicating the occurrence of threading events in the TS, in line with the knotting probability curves (Fig. 5B).

The formation of intermediate states under hydrophobic confinement is not specific of knotted proteins and has been reported for topologically regular folds[42, 44]. However, it is known that under bulk conditions knotted proteins are prone to populate on- and off-pathway intermediate states, giving rise to complex energy landscapes [13, 14, 16, 17, 22, 26, 82]. The mechanistic origin of these intermediates is, however, different from that of the ones reported here, whose onset is driven by strong intermolecular interactions with the cavity. The intermediate conformations detected here do not have an enhanced knotting frequency despite being structurally consolidated because the intermolecular interactions with the cavity block threading movements of the terminus. One may, however, speculate that these conformations could serve as starting conformations for the hydrophilic phase of the chaperonin cycle since their higher number of native interactions would contribute to decrease the free energy barrier to knotting.

Conclusions

The last decade witnessed an increased interest in knotted proteins and determining their structure, function, folding and knotting mechanisms. Experimental studies focusing on the α/β MTases, YibK from Haemophilus influenzae and YbeA from Escherichia coli, whose native structure embeds a trefoil knot, support the view that their knotting process *in vivo* may be assisted by the chaperonin GroEL-GroES [25]. However, the knotting mechanism inside the "chamber of secrets"[83] remains to be established. Computer simulations have shown that steric confinement provided by the chaperonin accelerates folding by decreasing backtracking, can alter the folding pathway, and facilitates knotting by lowering down the corresponding entropic barrier[49, 51]. However, it has been suggested[25], that the chaperonin is likely to have additional roles in the folding of knotted proteins by actively facilitating the unfolding of misfolded intermediates, and by stabilizing key interactions that promote knotting.

Here, we explored the folding of knotted proteins beyond steric confinement by considering the effects of intermolecular hydrophobic interactions in folding and knotting. We use a combination of lattice and off-lattice simulations, and model protein energetics with a modified structured-based model, where an additional term is added to a native centric Gō potential to capture intermolecular protein-cage interactions. Within the context of our model, strong intermolecular interactions would be established in the beginning of the cycle, causing the protein to tether and even adhere to the cavity's surface, and, as the cycle progresses, these interactions should become sufficiently weaker to allow for the release of the protein in a purely hydrophilic environment. We study several folding properties as a function of the strength of the intermolecular interaction parameter.

We started by comparing the effects of steric confinement of the thermodynamics and kinetics of folding of a lattice trefoil knot and a lattice twisted-three knot. We recapitulate the finding that steric confinement enhances thermal stability, and assists folding by decreasing the activation energy, in line with what has been reported for topologically regular proteins. The corresponding enhancement of folding rate is, however, considerably larger for the knotted trefoil, an observation that could contribute to rationalize the higher frequency of proteins with 3_1 knots in the PDB [5]. We also recapitulate the result that steric confining enhances knotting for protein MJ0366, by decreasing the corresponding entropic barrier. This, however, comes at the expense of an enhanced thermal stability, which could hamper unfolding of potentially trapped states.

Interestingly, moderate to high hydrophobic interactions with the cavity have an opposite effect on protein thermal stability. By competing with intramolecular interactions, protein-cavity interactions cause an energetic destabilization of the protein (when the interaction strength is moderate to high), which overcomes the entropic stabilization driven by a smaller available volume inside the cavity. Therefore, a primary role of transient hydrophobic interactions with the chaperonin, in the initial stage of its working cycle, may be that of destabilizing misfolded states, and facilitate productive refolding to the native structure in line with Lim and Jackson's hypothesis[25]. Most importantly

however, is the finding that mild-to-moderate hydrophobic interactions with the cavity (which would establish later in the cycle) lead to a significant enhancement of knotting probability in relation to bulk conditions while simultaneously moderating the effect of steric confinement in the enhancement of thermal stability It should be noted, however, that these results are based on the analysis of model systems embedding shallow knots and it remains to be investigated if the same conclusions are conserved for proteins with deep knots.

Author contributions

PFNF and AR designed the research. JE, AR and PFNF ran the simulations. All authors analysed the data. PFNF wrote the paper.

Acknowledgments

Work supported by UID/MULTI/04046/2019 Research Unit grant from FCT, Portugal (to BioISI). Part of this work was performed on INCD funded by FCT and EU under the project LISBOA-01-0145-FEDER-022153. AR acknowledges the financial support from Spanish "Ministerio de Economía y Competitividad" (under grant CTQ2016-78895-R). PFNF would like to acknowledge the contribution of the COST Action CA17139.

References

1. Richardson, J. S., [beta]-Sheet topology and the relatedness of proteins. *Nature* **1977**, 268, 495-500.
2. Koniaris, K.; Muthukumar, M., Knottedness in ring polymers. *Physical Review Letters* **1991**, 66, 2211-2214.
3. Taylor, W. R., A deeply knotted protein structure and how it might fold. *Nature* **2000**, 406, 916-919.
4. Millett, Kenneth C.; Rawdon, Eric J.; Stasiak, A.; Sułkowska, Joanna I., Identifying knots in proteins. *Biochemical Society Transactions* **2013**, 41, 533-537.
5. Jamroz, M.; Niemyska, W.; Rawdon, E. J.; Stasiak, A.; Millett, K. C.; Sułkowski, P.; Sulkowska, J. I., KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Research* **2015**, 43, D306-D314.
6. Alexander, K.; Taylor, A. J.; Dennis, M. R., Proteins analysed as virtual knots. *Scientific Reports* **2017**, 7, 42300.
7. Mallam, A. L.; Jackson, S. E., Folding Studies on a Knotted Protein. *Journal of Molecular Biology* **2005**, 346, 1409-1421.
8. Faísca, P. F. N., Knotted proteins: A tangled tale of Structural Biology. *Computational and Structural Biotechnology Journal* **2015**, 13, 459-468.
9. Jackson, S. E.; Suma, A.; Micheletti, C., How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current Opinion in Structural Biology* **2017**, 42, 6-14.
10. Dabrowski-Tumanski, P.; Sulkowska, J., To Tie or Not to Tie? That Is the Question. *Polymers* **2017**, 9, 454.

11. I., A. F.; G., P. D.; L., M. A.; Georg, B.; E., J. S., Untangling the folding mechanism of the 52 – knotted protein UCH – L3. *The FEBS Journal* **2009**, 276, 2625-2635.
12. Soler, M. A.; Nunes, A.; Faísca, P. F. N., Effects of knot type in the folding of topologically complex lattice proteins. *The Journal of Chemical Physics* **2014**, 141, 025101.
13. Ziegler, F.; Lim, N. C. H.; Mandal, S. S.; Pelz, B.; Ng, W.-P.; Schlierf, M.; Jackson, S. E.; Rief, M., Knotting and unknotting of a protein in single molecule experiments. *Proceedings of the National Academy of Sciences* **2016**, 113, 7533-7538.
14. Zhang, H.; Jackson, S. E., Characterization of the Folding of a 5₂-Knotted Protein Using Engineered Single-Tryptophan Variants. *Biophysical Journal* **2016**, 111, 2587-2599.
15. Bölinger, D.; Sułkowska, J. I.; Hsu, H.-P.; Mirny, L. A.; Kardar, M.; Onuchic, J. N.; Virnau, P., A Stevedore's Protein Knot. *PLOS Computational Biology* **2010**, 6, e1000731.
16. Wang, I.; Chen, S.-Y.; Hsu, S.-T. D., Folding analysis of the most complex Stevedore's protein knot. *Scientific Reports* **2016**, 6, 31514.
17. Wallin, S.; Zeldovich, K. B.; Shakhnovich, E. I., The Folding Mechanics of a Knotted Protein. *Journal of Molecular Biology* **2007**, 368, 884-893.
18. Sułkowska, J. I.; Sułkowski, P.; Onuchic, J., Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences* **2009**, 106, 3119-3124.
19. Faísca, P. F. N.; Travasso, R. M. D.; Charters, T.; Nunes, A.; M., C., The folding of knotted proteins: insights from lattice simulations. *Physical Biology* **2010**, 7, 016009.
20. a Beccara, S.; Škrbić, T.; Covino, R.; Micheletti, C.; Faccioli, P., Folding Pathways of a Knotted Protein with a Realistic Atomistic Force Field. *PLOS Computational Biology* **2013**, 9, e1003002.
21. Najafi, S.; Potestio, R., Folding of small knotted proteins: Insights from a mean field coarse-grained model. *The Journal of Chemical Physics* **2015**, 143, 243121.
22. Mallam, A. L.; Jackson, S. E., A Comparison of the Folding of Two Knotted Proteins: YbeA and YibK. *Journal of Molecular Biology* **2007**, 366, 650-665.
23. Mallam, A. L.; Morris, E. R.; Jackson, S. E., Exploring knotting mechanisms in protein folding. *Proceedings of the National Academy of Sciences* **2008**, 105, 18740-18745.
24. Mallam, A. L.; Rogers, J. M.; Jackson, S. E., Experimental detection of knotted conformations in denatured proteins. *Proceedings of the National Academy of Sciences* **2010**, 107, 8189-8194.
25. Lim, N. C. H.; Jackson, S. E., Mechanistic Insights into the Folding of Knotted Proteins In Vitro and In Vivo. *Journal of Molecular Biology* **2015**, 427, 248-258.
26. Noel, J. K.; Sułkowska, J. I.; Onuchic, J. N., Slipknotting upon native-like loop formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences* **2010**, 107, 15403-15408.
27. Noel, J. K.; Onuchic, J. N.; Sulkowska, J. I., Knotting a Protein in Explicit Solvent. *The Journal of Physical Chemistry Letters* **2013**, 4, 3570-3573.

28. Sułkowska, J. I.; Noel, J. K.; Onuchic, J. N., Energy landscape of knotted protein folding. *Proceedings of the National Academy of Sciences* **2012**, 109, 17783-17788.
29. Chwastyk, M.; Cieplak, M., Cotranslational folding of deeply knotted proteins. *Journal of Physics: Condensed Matter* **2015**, 27, 354105.
30. He, C.; Li, S.; Gao, X.; Xiao, A.; Hu, C.; Hu, X.; Hu, X.; Li, H., Direct observation of the fast and robust folding of a slipknotted protein by optical tweezers. *Nanoscale* **2019**, 11, 3945-3951.
31. Soler, M. A.; Faísca, P. F. N., How Difficult Is It to Fold a Knotted Protein? In Silico Insights from Surface-Tethered Folding Experiments. *PLOS ONE* **2012**, 7, e52343.
32. Soler, M. A.; Faísca, P. F. N., Effects of Knots on Protein Folding Properties. *PLOS ONE* **2013**, 8, e74755.
33. Škrbić, T.; Micheletti, C.; Faccioli, P., The Role of Non-Native Interactions in the Folding of Knotted Proteins. *PLOS Computational Biology* **2012**, 8, e1002504.
34. Bustamante, A.; Sotelo-Campos, J.; Guerra, D. G.; Floor, M.; Wilson, C. A. M.; Bustamante, C.; Báez, M., The energy cost of polypeptide knot formation and its folding consequences. *Nature Communications* **2017**, 8, 1581.
35. Xu, Z.; Horwich, A. L.; Sigler, P. B., The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. *Nature* **1997**, 388, 741.
36. Hayer-Hartl, M.; Bracher, A.; Hartl, F. U., The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences* **2016**, 41, 62-76.
37. Mallam, A. L.; Jackson, S. E., Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol* **2012**, 8, 147-153.
38. Chan Hue, S.; Dill Ken, A., A simple model of chaperonin‐mediated protein folding. *Proteins: Structure, Function, and Bioinformatics* **1996**, 24, 345-351.
39. Betancourt, M. R.; Thirumalai, D., Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity11Edited by A. R. Fersht. *Journal of Molecular Biology* **1999**, 287, 627-644.
40. Lucent, D.; Vishal, V.; Pande, V. S., Protein folding under confinement: A role for solvent. *Proceedings of the National Academy of Sciences* **2007**, 104, 10430-10434.
41. Baumketner, A.; Jewett, A.; Shea, J. E., Effects of Confinement in Chaperonin Assisted Protein Folding: Rate Enhancement by Decreasing the Roughness of the Folding Energy Landscape. *Journal of Molecular Biology* **2003**, 332, 701-713.
42. Jewett, A. I.; Baumketner, A.; Shea, J.-E., Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: Creation of an alternate fast folding pathway. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, 13192-13197.
43. Tian, J.; Garcia, A. E., Simulation Studies of Protein Folding/Unfolding Equilibrium under Polar and Nonpolar Confinement. *Journal of the American Chemical Society* **2011**, 133, 15157-15164.
44. Marino, K. A.; Bolhuis, P. G., Confinement-Induced States in the Folding Landscape of the Trp-cage Miniprotein. *The Journal of Physical Chemistry B* **2012**, 116, 11872-11880.

45. Sirur, A.; Best, R. B., Effects of Interactions with the GroEL Cavity on Protein Folding Rates. *Biophysical Journal* **2013**, 104, 1098-1106.
46. Mittal, J.; Best, R. B., Thermodynamics and Kinetics of Protein Folding under Confinement. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, 20233-20238.
47. Piana, S.; Shaw, D. E., Atomic-Level Description of Protein Folding inside the GroEL Cavity. *The Journal of Physical Chemistry B* **2018**, 122, 11440-11449.
48. Cheng, C.; Wu, J.; Liu, G.; Shi, S.; Chen, T., Effects of Non-native Interactions on Frustrated Proteins Folding under Confinement. *The Journal of Physical Chemistry B* **2018**, 122, 7654-7667.
49. Soler, M. A.; Rey, A.; Faisca, P. F. N., Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding. *Physical Chemistry Chemical Physics* **2016**, 18, 26391-26403.
50. Li, W.; Terakawa, T.; Wang, W.; Takada, S., Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proceedings of the National Academy of Sciences* **2012**, 109, 17789.
51. Niewieczerzal, S.; Sulkowska, J. I., Knotting and unknotting proteins in the chaperonin cage: Effects of the excluded volume. *PLOS ONE* **2017**, 12, e0176744.
52. Zhao, Y.; Dabrowski-Tumanski, P.; Niewieczerzal, S.; Sulkowska, J. I., The exclusive effects of chaperonin on the behavior of proteins with 52 knot. *PLOS Computational Biology* **2018**, 14, e1005970.
53. Heiko, L.; Alexander, S.; N., O. J., Robustness and generalization of structure-based models for protein folding and function. *Proteins: Structure, Function, and Bioinformatics* **2009**, 77, 881-891.
54. Baweja, L.; Roche, J., Pushing the Limits of Structure-Based Models: Prediction of Nonglobular Protein Folding and Fibrils Formation with Go-Model Simulations. *The Journal of Physical Chemistry B* **2018**, 122, 2525-2535.
55. Noel, J. K.; Onuchic, J. N. The Many Faces of Structure-Based Potentials: From Protein Folding Landscapes to Structural Characterization of Complex Biomolecules. In *Computational Modeling of Biological Systems: From Molecules to Pathways*, Dokholyan, N. V., Ed.; Springer US: Boston, MA, 2012, pp 31-54.
56. Tozzini, V., Coarse-grained models for proteins. *Current Opinion in Structural Biology* **2005**, 15, 144-150.
57. Li, M. S.; Co, N. T.; Reddy, G.; Hu, C.-K.; Straub, J. E.; Thirumalai, D., Factors Governing Fibrillogenesis of Polypeptide Chains Revealed by Lattice Models. *Physical Review Letters* **2010**, 105, 218101.
58. Krobath, H.; Estácio, S. G.; Faísca, P. F. N.; Shakhnovich, E. I., Identification of a conserved aggregation-prone intermediate state in the folding pathways of Spc-SH3 amyloidogenic variants. *Journal of molecular biology* **2012**, 422, 705-722.
59. Enciso, M.; Rey, A., Sketching protein aggregation with a physics-based toy model. *The Journal of Chemical Physics* **2013**, 139, 115101.
60. Holzgräfe, C.; Irbäck, A.; Troein, C., Mutation-induced fold switching among lattice proteins. *The Journal of Chemical Physics* **2011**, 135, 195101.
61. Das, S.; Eisen, A.; Lin, Y.-H.; Chan, H. S., A Lattice Model of Charge-Pattern-Dependent Polyampholyte Phase Separation. *The Journal of Physical Chemistry B* **2018**.

62. Wagner, C.; Kiefhaber, T., Intermediates can accelerate protein folding. *Proceedings of the National Academy of Sciences* **1999**, 96, 6716.
63. Taketomi, H.; Ueda, Y.; Gō, N., STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION. *International Journal of Peptide and Protein Research* **1975**, 7, 445-459.
64. Dill, K. A., Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, 24, 1501-1509.
65. Kyte, J.; Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **1982**, 157, 105-132.
66. Prieto, L.; de Sancho, D.; Rey, A., Thermodynamics of Go-type models for protein folding. *J Chem Phys* **2005**, 123, 154903.
67. Prieto, L.; Rey, A., Influence of the native topology on the folding barrier for small proteins. *J Chem Phys* **2007**, 127, 175101.
68. Prieto, L.; de Sancho, D.; Rey, A., Thermodynamics of Gō-type models for protein folding. *The Journal of Chemical Physics* **2005**, 123, 154903.
69. Prieto, L.; Rey, A., Influence of the native topology on the folding barrier for small proteins. *The Journal of Chemical Physics* **2007**, 127, 175101.
70. Faisca, P. F. N.; Ball, R. C., Thermodynamic control and dynamical regimes in protein folding. *The Journal of Chemical Physics* **2002**, 116, 7231-7237.
71. Faísca, P. F. N.; Travasso, R. D. M.; Parisi, A.; Rey, A., Why Do Protein Folding Rates Correlate with Metrics of Native Topology? *PLOS ONE* **2012**, 7, e35599.
72. Heinrich, K.; Patrícia, F. N. F., Interplay between native topology and non-native interactions in the folding of tethered proteins. *Physical Biology* **2013**, 10, 016002.
73. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **1953**, 21, 1087-1092.
74. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, 314, 141-151.
75. Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A., Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *Journal of Chemical Theory and Computation* **2007**, 3, 26-41.
76. Tubiana, L.; Polles, G.; Orlandini, E.; Micheletti, C., KymoKnot: A web server and software package to identify and locate knots in trajectories of linear or circular polymers. *The European Physical Journal E* **2018**, 41, 72.
77. Tubiana, L.; Orlandini, E.; Micheletti, C., Probing the Entanglement and Locating Knots in Ring Polymers: A Comparative Study of Different Arc Closure Schemes. *Progress of Theoretical Physics Supplement* **2011**, 191, 192-204.
78. Lu, D.; Liu, Z.; Wu, J., Structural Transitions of Confined Model Proteins: Molecular Dynamics Simulation and Experimental Validation. *Biophysical Journal* **2006**, 90, 3224-3238.
79. Faisca, P. F.; Travasso, R. D.; Parisi, A.; Rey, A., Why do protein folding rates correlate with metrics of native topology? *PLoS One* **2012**, 7, e35599.
80. Takagi, F.; Koga, N.; Takada, S., How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proceedings of the National Academy of Sciences* **2003**, 100, 11367-11372.

81. Thirumalai, D., From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I France* **1995**, 5, 1457-1467.
82. King, N. P.; Jacobitz, A. W.; Sawaya, M. R.; Goldschmidt, L.; Yeates, T. O., Structure and folding of a designed knotted protein. *Proceedings of the National Academy of Sciences* **2010**, 107, 20732.
83. Spiess, C.; Meyer, A. S.; Reissmann, S.; Frydman, J., Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends in Cell Biology* **2004**, 14, 598-604.

Figures

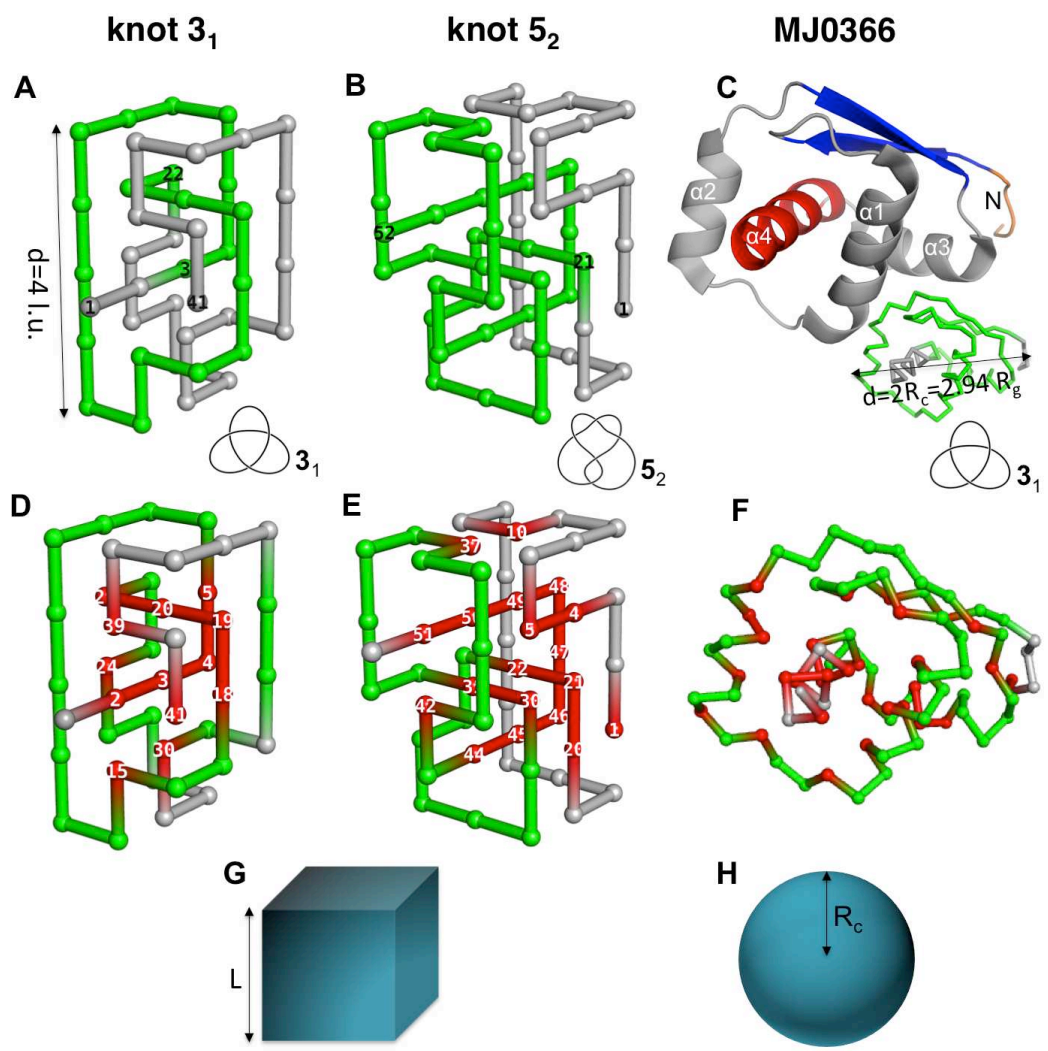


Figure 1. **Model systems**. Three-dimensional representation of the lattice model systems used in this study embedding a shallow trefoil knot (A) and a shallow twisted-three knot (B). We term these lattice systems as knot 3_1 and knot 5_2, respectively. The knotted core, i.e. the minimal segment of the backbone that contains the knot, is highlighted in green. (C) Cartoon representation of the native structure of protein MJ0366 (PDB ID: 2efv), which embeds a shallow trefoil knot: The C-terminal helix (helix 4) threads a loop formed by helices 1 and 2, which is locked by the two-stranded beta-sheet. The knotted core of protein MJ0366 (highlighted in green in the ribbon representation) extends from residue 11 to 76. In panels D-F the hydrophobic residues of sequence Ctc01 and protein MJ0366 are highlighted in red. In sequence Ctc01 the hydrophobic residues correspond to most connected beads (i.e. those that establish 3 or 4 native contacts), which are identified by their number along the chain. The lattice knot 3_1 and knot 5_2 are confined to a cubic region (F) whose linear size L is limited to d = 4 lattice units, while the radii of the confinement sphere (G), R_c, used in the off-lattice model protein MJ0366 is limited to $d = 2.94Rg$.

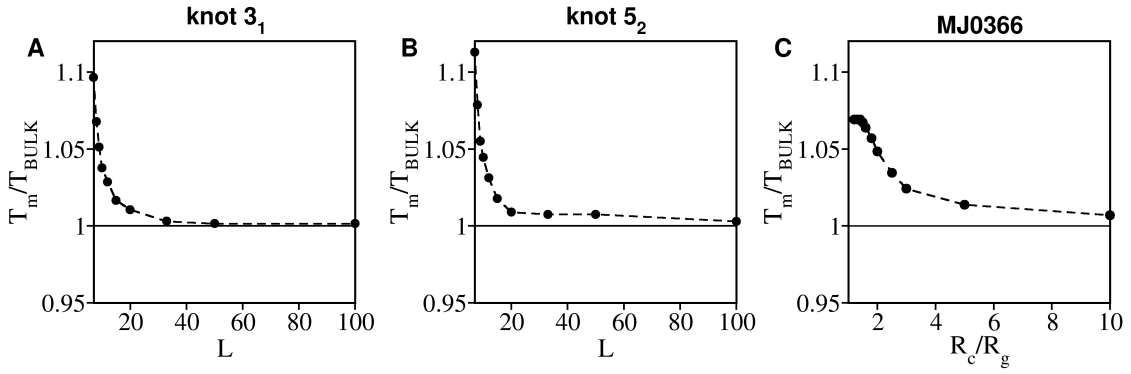


Figure 2. **Thermal stability under steric confinement**. Dependence of the melting temperature, T_m, (normalized to the melting temperature observed under bulk conditions, T_{BULK}), on the size of the confining box for lattice knot 3_1 (A), lattice knot 5_2 (B), and on the radii of the confining sphere (normalized to the gyration radius of the native state) for protein MJ0366 (C).

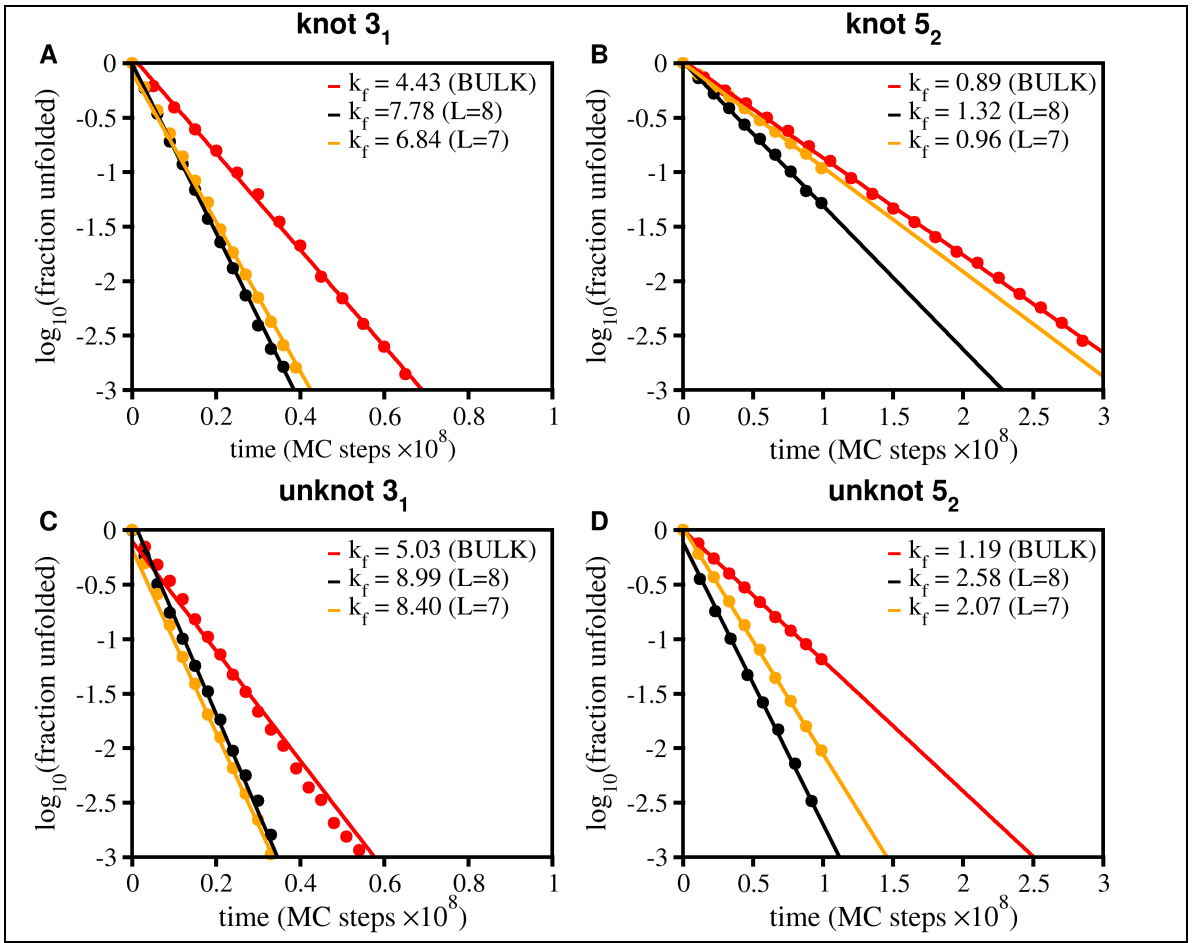


Figure 3. **Folding kinetics.** The folding rate k_f, corresponds to the slope of each represented curve and was determined at T_m. Panels (A) and (B) report the folding rate of the knotted lattice systems, and panels (C) and (D) report the folding rate of their unknotted counterparts.

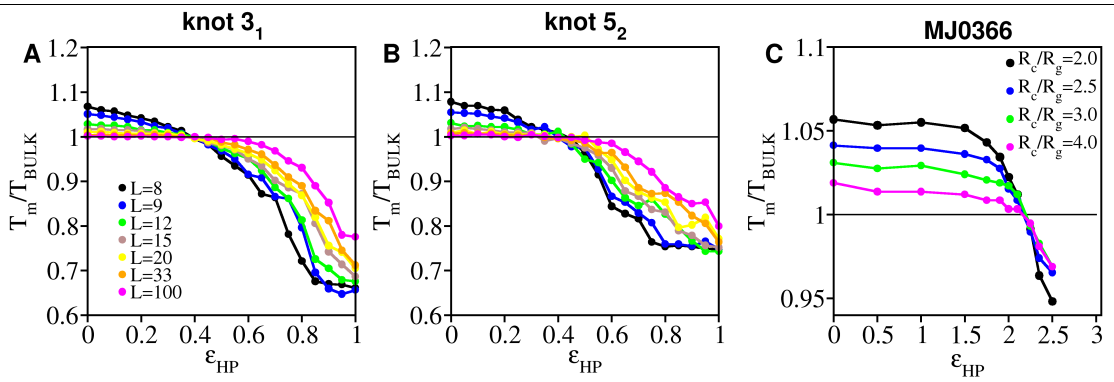


Figure 4. **Thermal stability under hydrophobic confinement**. Dependence of the melting temperature *Tm* (normalized to the melting temperature observed in bulk conditions, T_{BULK}) on the hydrophobic intermolecular parameter, ε_{HP}, for lattice sequence Ctc01 of knot 31 (a), sequence Ctc01 of knot 52 (b) and protein MJ0366 (C) under different confinement conditions.

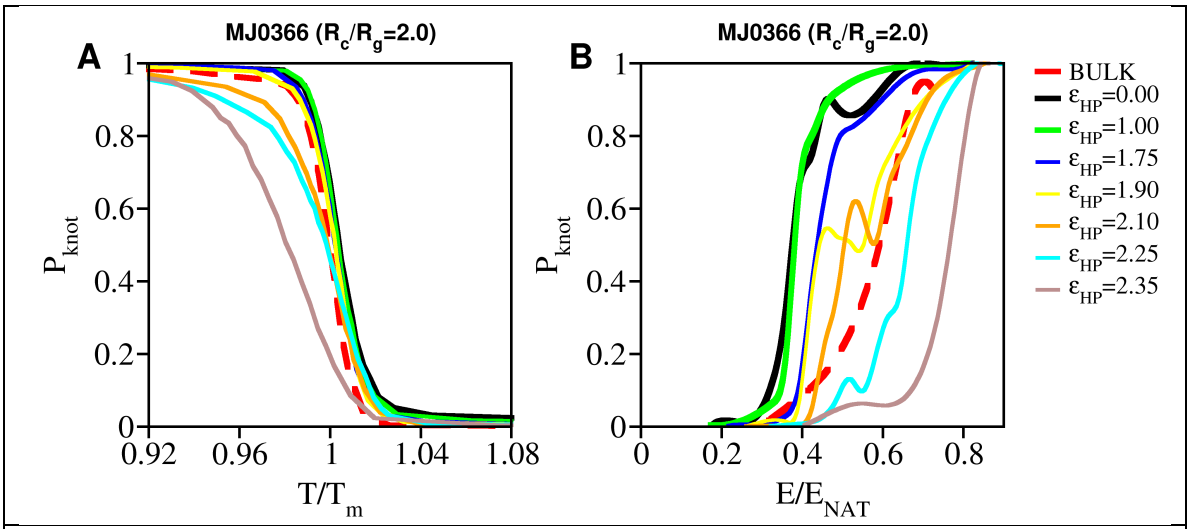


Figure 5. **Knotting probability of protein MJ0366 under hydrophobic confinement**. Dependence of knotting probability, P_{knot}, on temperature (normalized to the melting temperature, T_m) for several strengths of the hydrophobic intermolecular parameter ε_{HP} (A), and dependence of P_{knot} on the reaction coordinate at T_m (B) in a sphere of radius R_c=2R_g. The up-down trends on the knotting probability curves reflect the rather low population of conformations in the region corresponding roughly to the top of the free energy barrier, being numerical artifacts of the calculation method.

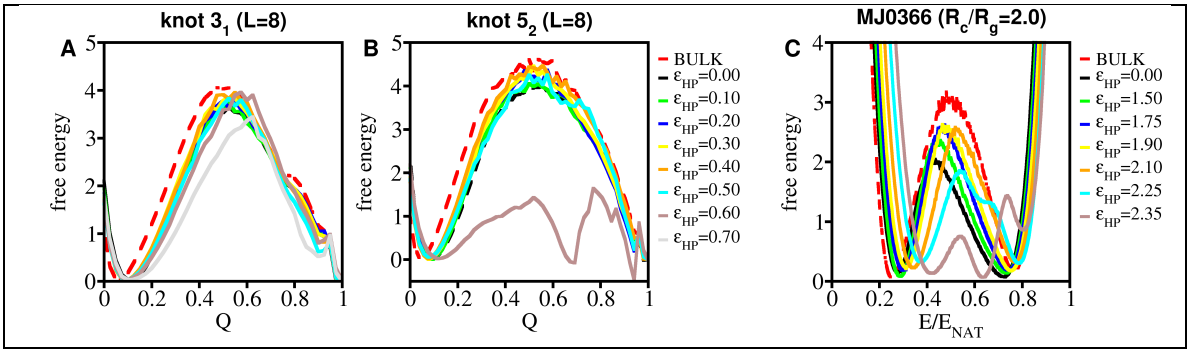


Figure 6. Free energy landscapes under hydrophobic confinement. Free energy projected on the reaction coordinate Q for sequence Ctc01 of lattice knot 3_1 (A) and lattice knot 5_2 (B) in a box of size L=8 at T_m for several strengths of the hydrophobic intermolecular parameter ε_{HP}. Free energy landscape evaluated of protein MJ0366 in a sphere of radius R_c=2R_g evaluated at T_m.

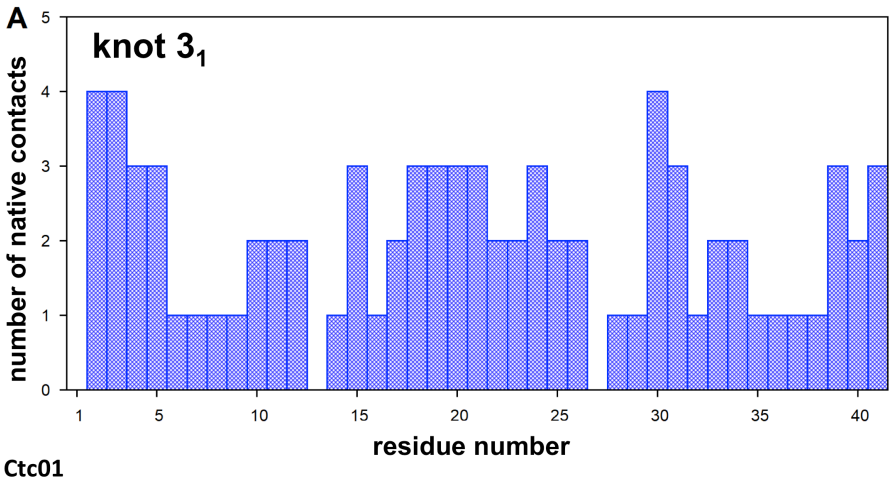
Supplementary information file for

Hydrophobic confinement modulates thermal stability and assists knotting in the folding of tangled proteins

João N Especial[1], Ana Nunes[1,2], Antonio Rey[3]* & Patrícia FN Faísca[1,2]*
[1]Departamento de Física and [2]BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, 1749-016 Lisboa, Portugal
[3]Departamento de Química Física, Facultad de Ciencias Químicas, Universidad Complutense, 28940 Madrid, Spain

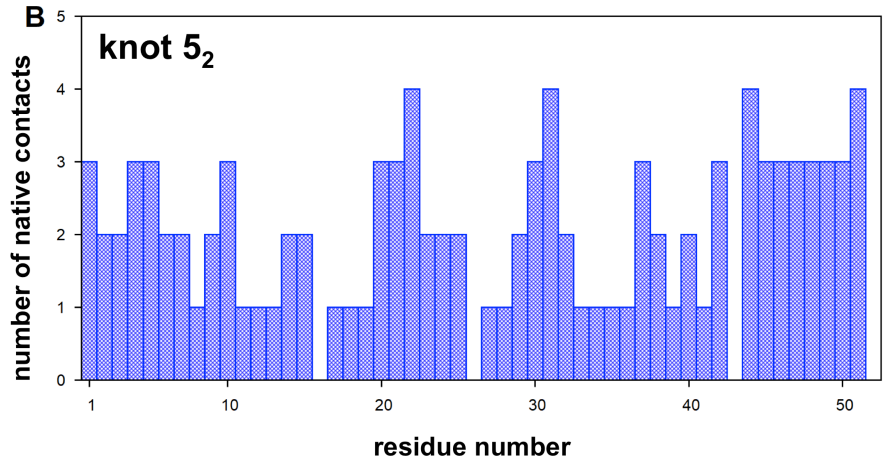
Corresponding authors:
Patrícia F.N. Faísca, Email: pffaisca@fc.ul.pt
Antonio Rey, Email: areygayo@ucm.es

This file includes:
Supplementary Figures SI:1-SI:7



Ctc01
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29
-30-31-32-33-34-35-36-37-38-39-40-41

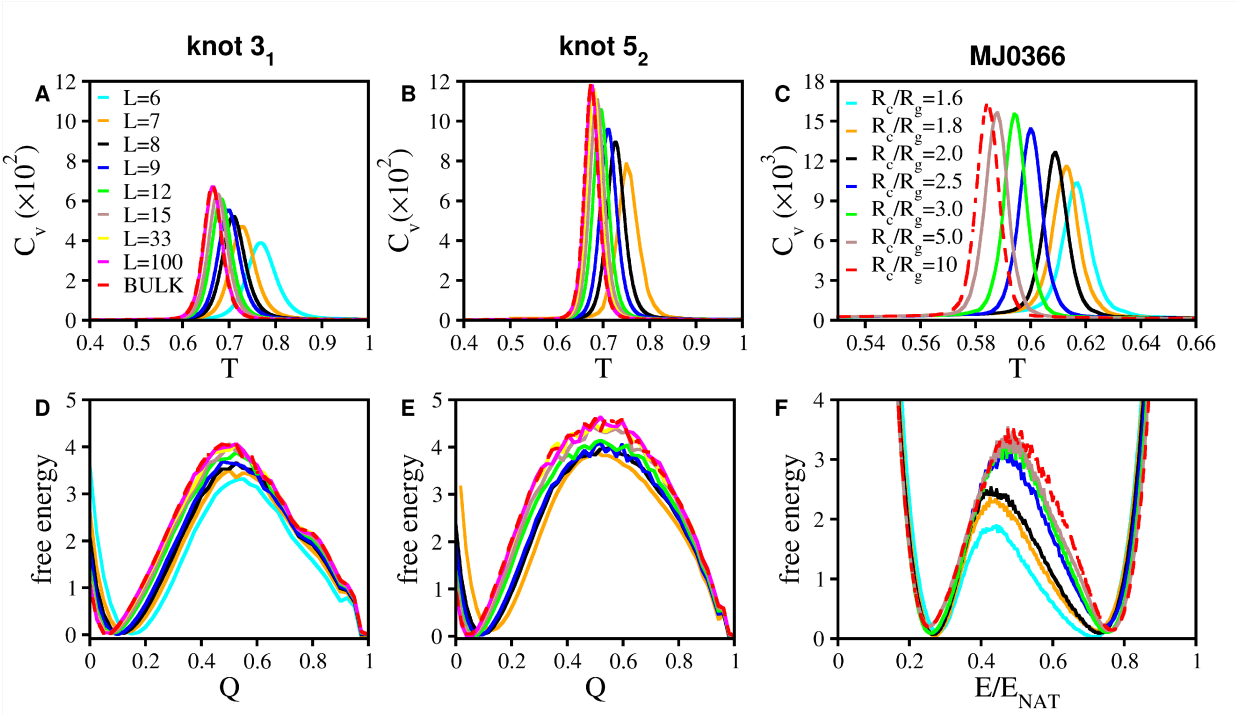
Ctc14
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29
-30-31-32-33-34-35-36-37-38-39-40-41



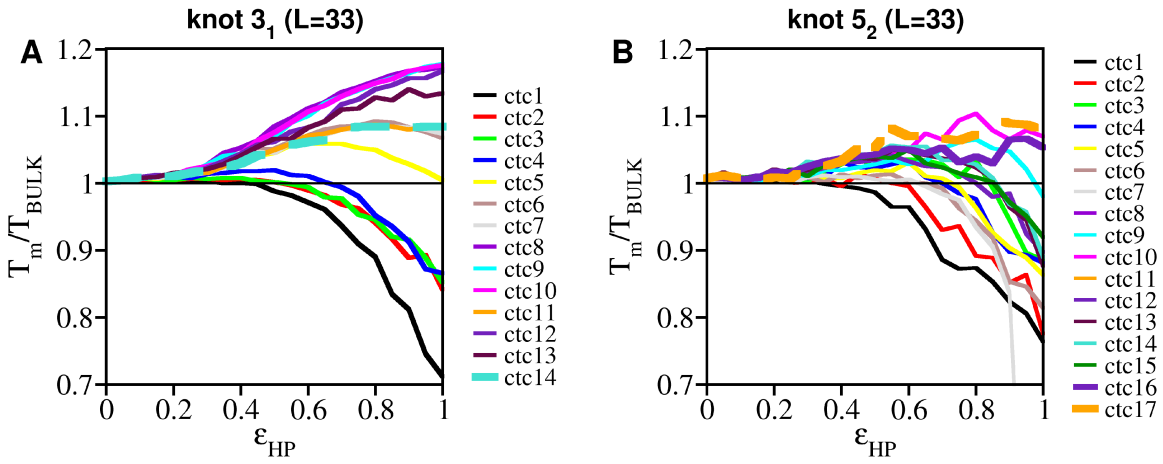
Ctc01
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29
-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51

Ctc17
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29
-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51

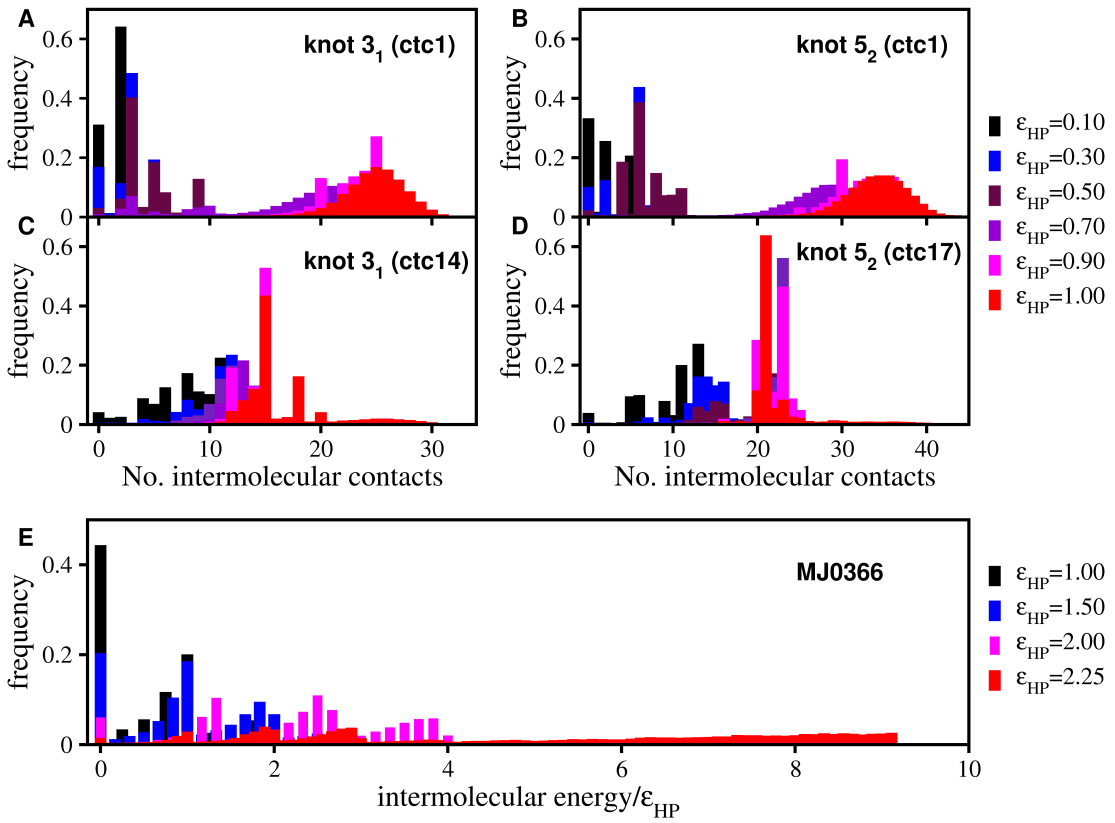
SI Figure 1 - Number of native contacts per residue in lattice knot 3_1 (A) and lattice knot 5_2 (B). The hydrophobic beads are colored red in the designed protein sequences. In sequence Ctc01 the most connected beads (establishing 3 and 4 native contacts) are the ones considered hydrophobic.



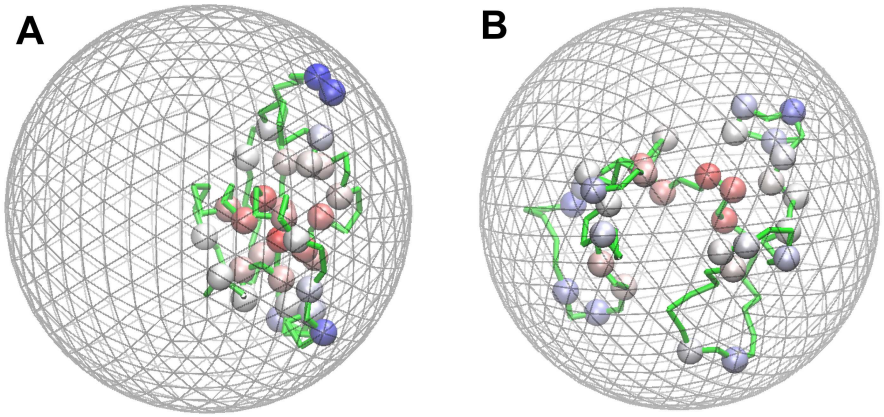
SI Figure 2 - **Folding thermodynamics under steric confinement**. Heat capacity as a function of temperature for knot 3_1 (A), knot 5_2 (B), and protein MJ0366 (C). The melting temperature T_m is the temperature at which the C_v peaks. The projection of the free energy on the selected reaction coordinate at T_m for knot 3_1 (D), knot 5_2 (E), and protein MJ0366 (F).



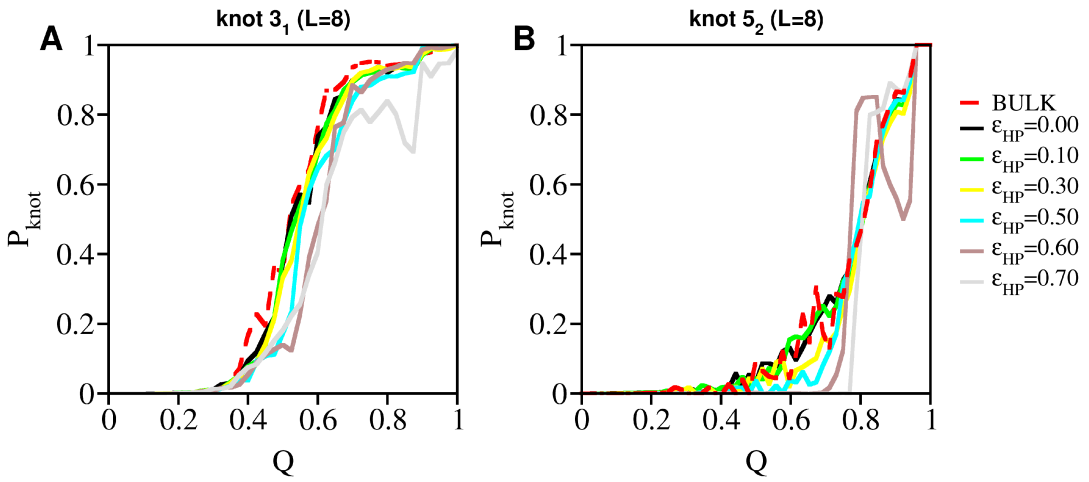
SI Figure 3 - Thermal stability under hydrophobic confinement. Dependence of the melting temperature, T_m, (normalized to the melting temperature observed in bulk conditions, T_{BULK}) on the hydrophobic intermolecular parameter, ε_{HP}, for lattice sequence Ctc01 of knot 3_1 (a), sequence Ctc01 of knot 5_2 (b), and protein MJ0366 (C) under different confinement conditions.



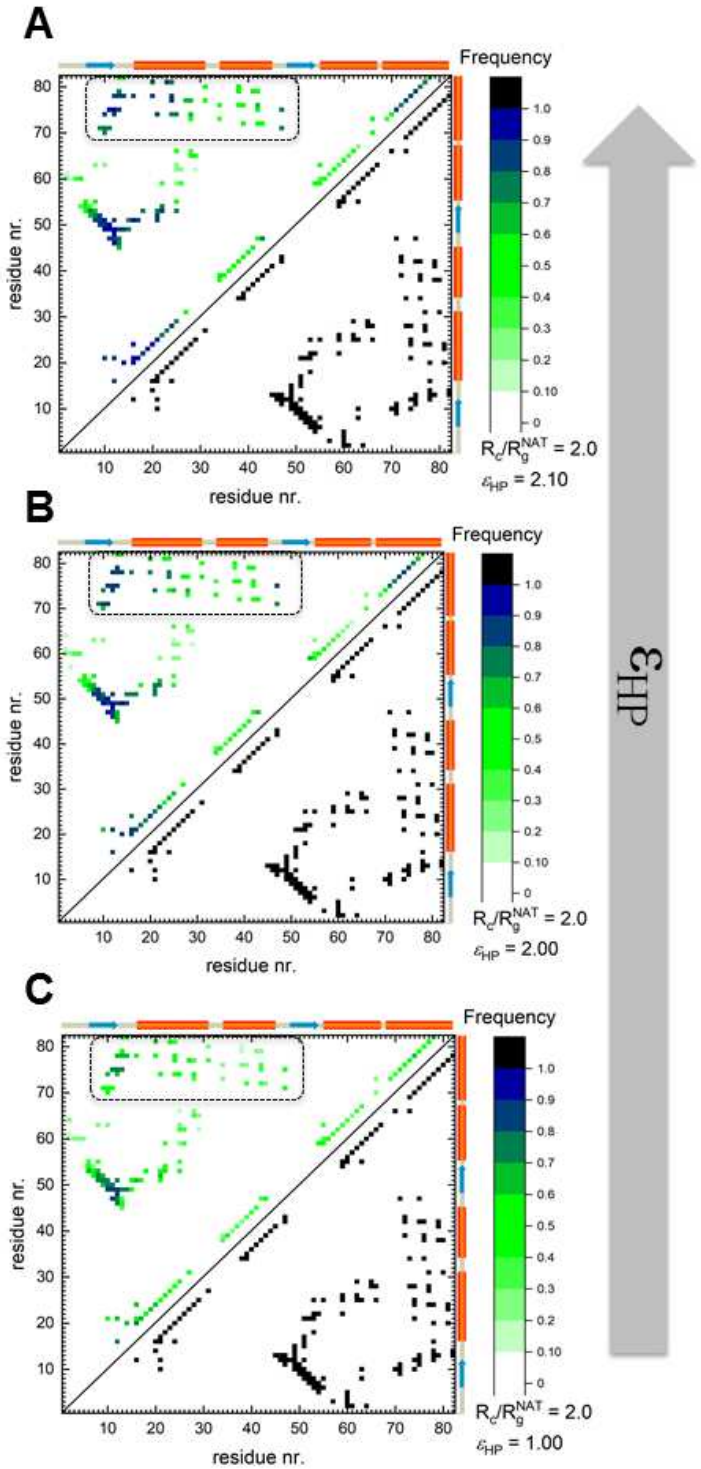
SI Figure 4 - **Distribution of intermolecular interactions**. Probability histograms for the distribution of the number of established intermolecular contacts at fixed temperature (T=0.6) for model sequence Ctc01 of lattice knot 3_1 (A), sequence Ctc01 of lattice knot 5_2 (B), and for sequences ctc14 (knot 3_1) (C) and ctc17 (knot 5_2) in a box of size L=8. The probability histogram for intermolecular energy for protein MJ0366 at fixed temperature (T=0.58) in a sphere of radius R_c=2R_g (E). Depending on ε_{HP}, the simulation temperature can be above or below T_m (see main text).



SI Figure 5 - **Conformations populated under hydrophobic confinement.** The confining sphere has radius R_C/R_s=2.0, and hydrophobic residues are represented as spheres. Residues closer to the cavity's surface are colored in blue while those further away from the surface are colored in white. The conformation (A) is populated at low T ($< T_m$) while conformation (B) is populated at high T ($> T_m$).



SI: Figure 6 - **Knotting probability of lattice sequence ctc0 under hydrophobic confinement**. Dependence of P_{knot} on the reaction coordinate Q for knot 3_1 (A) and knot 5_2 (B), for several strengths of the hydrophobic intermolecular parameter, ε_{HP}, in a box of size L=8 at T_m.



SI: Figure 7 – **Transition state structure**. Effect of the strength of intermolecular interactions with the cavity, ε_{HP}, on the structure of the transition state ensemble: ε_{HP}=2.1 (A), ε_{HP}=2.0 (B) and ε_{HP}=1.0 (C). The dotted region in the probability (native contact) maps indicate the native contacts establishing between the C-terminal helix 4 and its surroundings. In all cases the R_C/R_g=2.0.