RESEARCH ARTICLE

The exclusive effects of chaperonin on the behavior of proteins with 5_2 knot

Yani Zhao[1,2☯], Pawel Dabrowski-Tumanski[1,3☯], Szymon Niewieczerzal[1☯], Joanna I. Sulkowska[1,3]*

1 Centre of New Technologies, University of Warsaw, Warsaw, Poland, **2** Institute of Physics, Polish Academy of Sciences, Warsaw, Poland, **3** Faculty of Chemistry, University of Warsaw, Warsaw, Poland

☯ These authors contributed equally to this work.
* jsulkowska@chem.uw.edu.pl

Abstract

The folding of proteins with a complex knot is still an unresolved question. Based on representative members of Ubiquitin C-terminal Hydrolases (UCHs) that contain the 5_2 knot in the native state, we explain how UCHs are able to unfold and refold *in vitro* reversibly within the structure-based model. In particular, we identify two, topologically different folding/unfolding pathways and corroborate our results with experiment, recreating the chevron plot. We show that confinement effect of chaperonin or weak crowding greatly facilitates folding, simultaneously slowing down the unfolding process of UCHs, compared with bulk conditions. Finally, we analyze the existence of knots in the denatured state of UCHs. The results of the work show that the crowded environment of the cell should have a positive effect on the kinetics of complex knotted proteins, especially when proteins with deeper knots are found in this family.

Author Summary

Self-tying of knotted proteins remains a challenge both for theoreticians and experimentalist. In this work, we study the proteins with complex, the 5_2 knot, in a bulk and confined within a chaperonin box. We show that in our model we recreate the experimental results, identify two topologically distinct folding pathways and explain the beneficial role of confinement for complex knotted proteins. Encapsulation provides a possibility to fold via alternative pathway—folding via trefoil intermediate knot (N-terminal pathway) from entropic reason while folding via the C-terminal (direct tying) appears with the same probability. The results of this work show, how crowded environment in the real cell may enhance self-tying of proteins. The results are also the first step to the identification of possible oligomerization-prone forms of UCHs, which may cause neurodegenerative diseases.

Introduction

The role of knots in protein structures is still not fully understood. The topological complexity induces stability to the structure [1, 2] and enforces local motifs favorable for active sites of enzymes [3]. The latter fact may explain, why over 80% of known knotted proteins are enzymes with the active site located at the entangled region [4]. Nevertheless, folding process of knotted proteins is a fundamental and still not solved problem.

One of the families of knotted proteins is Ubiquitin C-terminal Hydrolase (UCH) of which the characteristic feature is the presence of a complex topological fingerprint $5_2 3_1 3_1$ [4] as shown in the Fig 1. This means that the entire protein forms a 5_2 knot as a whole, but some of its subchains form two trefoil knots (see Table 1). Each entry in the matrix indicates the knot type, formed by one continuous subchain, by one particular color; e.g. the unknot is denoted in white. Each such subchain starts with the N-terminal amino acid at position x and ends
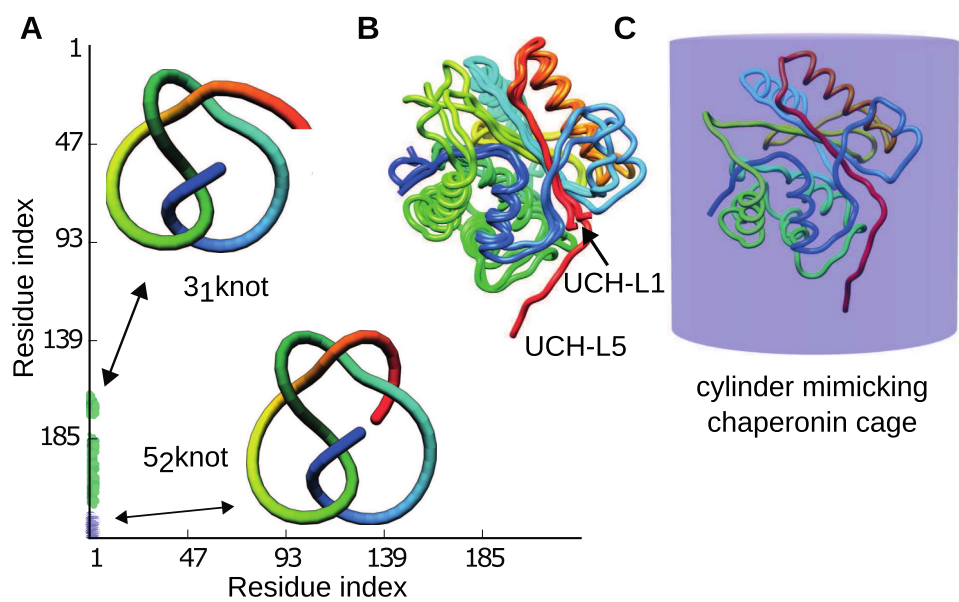
Fig 1. Molecular structures and a topological fingerprint for Ubiquitin C-terminal Hydrolases, UCH. A: The topological fingerprint—presented as a matrix—for UCH shows that this protein forms the Gordian knot (5_2) as a whole but some of its subchains form two trefoil slipknots (described in detail in the text). B: Ribbon representation of backbone superposition of UCH-L1 (PDB code 2LEN and 3IRT) and UCH-L5 (PDB code 4I6N), where red and blue color indicate the N- and C-terminus, respectively. C: Model of the chaperonin cage, a cylindrical box (closed from both sides, with height equal to the diameter), which encapsulates the UCH-L5 protein.

https://doi.org/10.1371/journal.pcbi.1005970.g001

Table 1. Protein structures investigated in this work. L_{seq} denotes the number of residues in the structure, "cont." the number of native contacts, N_{tail} and C_{tail} the length of N- and C-terminal 5_2 knot tails. The knotted core of the innermost 3_1 knot inside 5_2 knot is also shown. PDB code 4WLR denotes protein with the deepest 5_2 knot in the UCH family. More detailed comparison incl the sequence similarity is given in SI.

PDB id	UCH type	organism	L_{seq}	cont.	knot location		N_{tail}	C_{tail}
					3_1	5_2		
3IRT	L1	Homo sapiens	223	454	6-162	5-220	4	2
2LEN	L1	Homo sapiens	231	387	5-163	6-220	5	10
4I6N	L5	Trichinella spiralis	229	470	12-166	12-225	11	3
4I6N-m	–	–	222	457	5-159	5-218	4	3
4WLR	L5	Mus musculus	313	860	6-160	5-219	4	94

https://doi.org/10.1371/journal.pcbi.1005970.t001

with the C-terminal amino acid at position y, and the corresponding colored entry in the matrix is shown in position (x, y) (along respectively horizontal and vertical axes). Specifically, one can trace what is the topology of the subchain with one end in N-terminus. Pictorially, this is represented as the traveling down of the left-most vertical line in the matrix in Fig 1). In the beginning, successive subchains are unknotted, however reaching at least Ile163 the subchain becomes trefoil knotted (first green patch in the matrix). The subchain Met1-Tyr173 is still knotted, however then the chain winds back forming a slipknot loop and when the end of the subchain is in-between Glu174 and Pro180 (parts of C-terminal β-strands), such subchain is unknotted (the break between the green patches). Next, the subchain starting in N-terminus and ending in-between residues Pro182 and Ala216 is again trefoil knotted (bottom green patch) and finally the whole chain is 5_2 knotted (blue patch). The $5_2 3_1 3_1$ fingerprint is unique and conserved in all UCH members, which are separated by billion years of evolution and exhibit a very low sequence similarity (below 30%) [5]. Notably, the formation of the larger trefoil results in the formation of the inner-most (dipper) trefoil knot. Therefore, in subsequent analysis by "formation of 3_1 knot", we mean the formation of the larger (and hence both) trefoil knot.

The UCH superfamily is a group of deubiquitinating enzymes (DUBs). Their exact substrates have not yet been determined, however it seems that the role of UCHs is to detach the ubiquitin from small nucleophiles. Four of the UCH family members exist in humans: UCH-L1, UCH-L3, human UCH-L5 (UCH37) and BAP-1. They share a high degree of homology in their catalytic domains [6], surrounded by the deepest 3_1 knot. Moreover, UCHs have a tissue-specific expression in complex organisms such as humans and their activity is crucial from the therapeutical point of view. For example, UCH-L3 has been shown to be upregulated in breast cancer tissues [7], and a high expression of UCH-L5 is significantly associated with poor prognosis in human epithelial ovarian cancer [8]. On the other hand, UCH-L1 is one of the most common proteins in human brain (composing up to 1-2% of the brain the total protein content [9]), and it is highly expressed in pancreatic [10], esophageal [11], prostate [12], medullary thyroid [13], colorectal carcinomas [14] and HPV16-transformed cells [15]. Its misfolded forms were connected with neuronal disorders such as Parkinson's, Huntington's and Alzheimer's diseases [16], which justifies the importance of studying the UCH folding process.

In general, it is expected that folding of knotted proteins is governed mainly by the depth of the knot and the complexity of the topological fingerprint [17, 18]. Self-tying was observed theoretically for the smallest knotted proteins with DNA binding motif and a rather shallow 3_1 knot [19, 20]. They mainly fold by a slipknot conformation [17, 19]. Similarly, it was shown that proteins with a deep trefoil knot, such as YibK and YbeA, can self-tie [21]. The theoretical results obtained for these proteins with a structure based model additionally revealed that a knotting event is a rate-limiting step [22] and the folding efficiency can be controlled by non-native contacts [23] or consideration of cotranslational on-ribosome folding [24]. For the protein with a 6_1 knot, DehI [25], there were only a few successful folding pathways observed theoretically. Surprisingly, this protein folds via a simple mechanism: a large twisted loop formed on the backbone flips over another protein fragment previously arranged in a twisted loop, and in consequence, the six-fold knot is created in a single movement. Even though this protein is prone to aggregate, the experimental data support this mechanism [26]. These results suggest that bulk structure-based models can be used to investigate knotted proteins.

On the other hand, experimental data show that knotting process of trefoil-knotted YibK and YbeA bacterial proteins can be specifically and significantly accelerated by the GroEL-GroES chaperonin complex [21] encapsulating the folding protein. This agrees with the theoretical investigation showing that knotting probability of polymers increases in confinement [27]. Only due to the encapsulation (following [28]), successful reversible folding was observed

for members of knotted proteins with DNA binding motif VirC and DndE [29]. It is then natural to expect, that chaperonins encapsulating proteins may also facilitate folding and self-knotting of eucaryotic UCHs, although no experimental result in this topic is available yet. Nevertheless, one has to bear in mind, that encapsulation is the simplest possible model of chaperonin, lacking many "biological features"—specific binding to the cage, chaperonin conformational changes, etc.

However, any theoretical results have to be confronted with the experimental data concerning UCHs folding. It has been shown already that UCHs can fold and refold reversibly in two parallel pathways, each consisting of one slow and one fast phase, as determined from chevron plot [30, 31]. Despite a common mechanism, folding processes of different UCHs is characterized by different kinetic parameters. Such differences can stem from a various depth of knots in UCH family [4], ranging from rather shallow (from the N-terminus) to the deep knot, which we just found, as shown in Table 1. The most attention-drawing are the S18Y and I93M mutations, which were found to modify (either decrease or increase) the risk of Parkinson's disease [32], and the intermediates on folding pathway, as these are especially prone to oligomerization [33]. However, the results concerning these mutations are variable and differ in different studies. Despite the successful assignment of the majority of signals in NMR spectrum of the UCH-L1 [34], and in spite of studies of its tryptophan variants [33], the exact conformation of intermediates is still unclear. This may be due to a broad structural plasticity around the intermediate states [30]. The self-tying was postulated via the direct knotting event in accordance with the theoretical study of the on-lattice model of a designed by hand heteropolymer chain with 5_2 knot [35]. However, the optical tweezers stretching experiment showed that the threading significantly decelerates the folding [36]. Still, because the topology cannot be detected in the *in vitro* experiment, the mechanism of knot tying remains unresolved.

In this study, we asked following questions: What is the difference between the two experimentally observed, parallel folding pathways? What is the influence of a chaperonin cage (confinement) on the folding and self-tying of UCHs? And more generally, what are the dynamical properties of UCH in a bulk and in a confinement? To answer these questions, we performed a comprehensive study of representative UCH members in a model of chaperonin cage (mimicked by repulsive cavity) and in the bulk, using structure-based model simulations. To ensure robustness of the results, we investigated proteins from different organisms, with low sequence similarity, and different depth of the knot. The results show that the structure-based model was sufficient to knot and unknot each of the studied proteins with and without the presence of confinement. However, only in the confinement, the simulations in transition temperature were accessible. We performed a comprehensive analysis of knot occurrence during simulations, resulting in the identification of two topologically distinct pathways. To relate our results to the experiment, we reproduce the chevron plot for a representative protein member of UCH family, revealing the existence of fast and slow phases. Next, we studied short-lived knots on the folding and the unfolding pathway and revealed for the first time the existence of random knots in the unfolded protein chain. To our knowledge, this is the first theoretical study with a direct investigation of the influence of the excluded volume on proteins containing complex knots.

Materials and methods

Protein structures

To obtain robust results, three sequentially different members of UCH family denoted with their PDB codes 3IRT (UCH-L1), 2LEN (UCH-L1), and 4I6N (UCH-L5) were studied. Additionally, to check an influence of the length of knot tails, we constructed *in silico* mutant of

4I6N, denoted as 4I6N-m, obtained by removing 7 residues from the N-terminus. All investigated structures feature a left-handed 5_2 knot and complex topological fingerprint $5_2 3_1 3_1$. Alignment of the 3-dimensional structures of the studies proteins is shown in the Fig 1 and their most important topological and structural information is summarized in Table 1. Further structural and sequential comparison of chosen structures is presented in S1 Appendix, part 1. Not determined region (amino acids 142–152) in the structure of 4I6N was repaired using Modeller software [37], where the model with the lowest DOPE potential (Discrete Optimized Protein Energy) was chosen. The DOPE potential is one of the quantities assessing the structure correctness [38].

The model and simulations

The dynamics of investigated structures was studied in structure based $C\alpha$ model [39, 40] with standard parameters as proposed by the SMOG server [41]. The model included bonded interactions (bonds, planar and dihedral angles), bead excluded volume (Lennard-Jones repelling part) and non-bonded interactions described with a 10-12 Lennard-Jones potential. The non-bonded attraction was applied between residues forming contacts in the native structure, as defined in [42]. The number of the native contacts for all considered proteins is presented in Table 1. The folding/unfolding transitions were studied through constant temperature molecular dynamics simulations with the Nose-Hoover thermostat (coupling constant eq. 0.025) using Gromacs v4.5.4 package [43]. Temperature $\tilde{T}$ in Gromacs is defined by the equation $\tilde{T} = (k_b T)/(\epsilon \tilde{k}_b)$, where $\tilde{k}_b = 0.00831451$. Through the text, the temperature is denoted simply as T and the Boltzman constant $\tilde{k}_b$ as k_b. There were performed 200 simulations for each structure and temperature. The number of steps was in the range $10^7 - 1.6 \cdot 10^9$ steps depending on the condition.

Simulation of the confinement

The confinement is represented by a cylinder (Fig 1) with a diameter equal to its height equal 6.0 nm [44], introduced into the system as in [29]. The interactions between the inner wall of the cylinder and protein are purely repulsive (only a confinement effect). Such model was previously used to study confinement or crowding effect on protein folding [44–46].

Topological analysis

The data concerning the position of the knotted core and the length of the knot tails (Table 1) are taken from KnotProt server [4]. The knot type of each of the subchains of the protein is determined using the implementation of the HOMFLY-PT polynomial [47–49] and the chain closing method as in [5, 50]. The same algorithm was used to detect the entanglement along the protein backbone during simulations. The knot was regarded as present in the simulation if it was detected for at least 5 consecutive frames.

Reaction coordinates, (un)folding pathways

The similarity to the native state was measured by the fraction of native contacts, Q. At given conformation, each native contact was regarded as present, if the distance between a pair of C_α atoms was less than 1.2 times their native distance. The untied structure was regarded as unfolded, if $Q < 0.2$. By unfolding pathway, we mean the shortest part of trajectory connecting knotted structure with $Q > 0.9$ and unfolded structure. By folding pathway, we mean the shortest part of the trajectory connecting unknotted structure with $Q < 0.4$ and knotted

structure with $Q > 0.9$. Folding trajectories start from one of the previously generated 100 unknotted conformations with $Q < 0.2$. All initial structures belong to separate clusters with 0.1 nm cutoff, to remove any possible bias.

Visualisation

The structures were visualized using UCSF Chimera [51].

Results

The landscape of successful folding pathways that leads to the correctly knotted native conformation is observed for each of the proteins, in both conditions through our simulations. As the folding of all UCHs is similar [31], we concentrated our study on the most commonly analyzed protein—UCH-L1 (PDB code 3IRT), comparing results with other UCHs when needed. Our results naturally split into three parts: description of folding/unfolding landscape, kinetics of the process and an analysis of the random and short-lived knots.

The landscape of folding and unfolding of 5_2 knotted proteins from the UCH family

Folding of UCHs. Folding of UCHs is known to follow two parallel pathways [30, 31]. In our simulations, data analysis of time evolution of the topology revealed also two, topologically distinct pathways, shown schematically in Fig 2. The pathways in general differ in the order of the protein tails' secondary and tertiary structure formation:
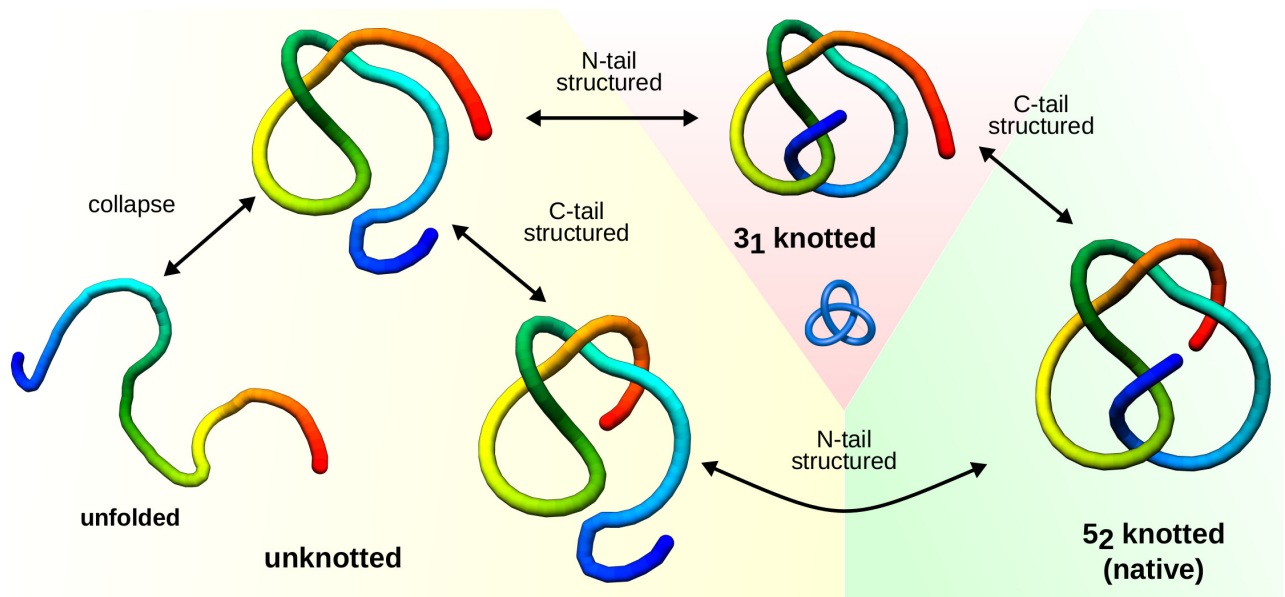
Fig 2. Schematic representation of possible folding pathways of UCHs. The folding starts with a collapse, next, depending on which terminus becomes structured first, two topologically distinct pathways are possible. Lower arrows: the pathway where N-terminus is structured after C-terminus (F_N), protein self-ties in one step (direct $0_1 \rightarrow 5_2$ transition). The upper arrows: pathway where C-terminus is structured last (F_C)—the protein follows $0_1 \rightarrow 3_1 \rightarrow 5_2$ pathway, where in the 3_1 knotted intermediate the N-terminus is in its native position, while C-terminus is not. Different background denotes different topology.

https://doi.org/10.1371/journal.pcbi.1005970.g002

- F_N—the N-terminus is structured last, in particular after C-terminus (resulting in a direct $0_1 \rightarrow 5_2$ transition);

- F_C—the C-terminus is structured last, in particular after N-terminus—the protein follows $0_1 \rightarrow 3_1 \rightarrow 5_2$ pathway, where in the 3_1 knotted intermediate the N-terminus is in its native position, while C-terminus is not.

The existence of 3_1 knotted intermediate was questioned before as in principle it should require two threading events (of both termini), being hence entropically unfavorable [33]. Observe however, that actually two threadings are required independently on the pathway, however, they can be realized in a relatively easy way, if the protein is not yet collapsed. In particular, the tail may first form some of its secondary structure, and then the knotting loop may build upon the structured terminus ("squeezing" the loop around it). The existence of the self-tying mechanism in the bulk, realized by two parallel pathways, is clearly in agreement with the experimental results.

To extract the dominating folding mechanism, we determined the probabilities (frequency of occurrences) of each pathway (F_N and F_C). The results for various temperatures in both conditions are presented in Table 2. For all accessible temperatures, the F_N pathway significantly dominates. The F_C folding pathway is more populated at lower temperatures and in the presence of confinement. At the highest accessible folding temperature in bulk (112), probabilities of the observed F_C pathway are equal to 2% (bulk) and 18.5% (confinement).

The preference for F_N pathway stems from the fact, that the C-terminus forms the β-strand, and therefore it is more strongly "pulled" towards its native position, compared to loose N-terminus. This effect can be also quantified in terms of the energy of the states. In our model, the mean potential energy (at a given temperature) of the structure with C-terminus formed (but not N-terminus) is always lower than the mean energy of the state with only N-terminus formed (for details see S1 Appendix, part 2). Therefore, the state with structured C-terminus is thermodynamically more favorable during folding and hence more commonly formed, leading to the higher prevalence of the F_N pathway. The increased probability of F_C pathway at lower temperatures may however indicate, that this pathway is kinetically more accessible. Indeed, for lower temperatures the core of the structure collapses relatively fast, making it hard for C-terminus to thread the loop due to entropic reasons. On the other hand, the N-terminus has to thread a relatively large loop formed by about 60 residues (helices H4-H7). Therefore, if only the helices are not yet attached to the core of the protein (which is commonly the case),

Table 2. Probabilities of (un)folding pathways in different conditions. P_{F_N} and P_{F_C} denote the folding probability of 3IRT via pathways F_N and F_C at different temperatures T. P_{misfold} denotes the probability of misfolded structures during the folding of 3IRT. P_{U_N} and P_{U_C} denote the unfolding probabilities of 3IRT via pathways U_N and U_C.

Folding							Unfolding					
T	P_{F_N} (%)		P_{F_C} (%)		P_{misfold} (%)		T	P_{U_N} (%)		P_{U_C} (%)	
	bulk	conf	bulk	conf	bulk	conf		bulk	conf	bulk	conf
105	82.0	50.5	14.0	18.0	4.0	31.5	118	96.5	–	3.5	–
107.5	89.5	–	9.5	–	1.0	–	120	–	92.0	–	8.0
110	97.5	75.5	2.5	16.5	0.5	8.0	122	–	94.0	–	6.0
112	98.0	80.0	2.0	18.5	0.0	1.5	125	97.5	97.5	2.5	2.5
115	–	85.5	–	14.0	–	0.5	130	87.0	92.0	13.0	8.0
118	–	93.5	–	6.5	–	0.0	135	89.0	88.5	11.0	11.5
120	–	94.0	–	6.0	–	0.0	140	96.0	95.0	4.0	5.0
							145	100.0	99.0	0.0	1.0

threading is more favorable for N-terminus. The presence of the confinement enhances folding by the F_C pathway narrowing the configurational space scanned by N-tail and forcing the N-terminus stay spatially close the loop it has to thread. On the other hand our preliminary results suggest, that too tight confinement forces the helices H4-H7 to form a tertiary structure (attached to the core), reducing the advantage for N-terminus and hindering (but still not forbidding) folding. Hence one can speculate that very tight confinement will again favor the F_N pathway. This could be possibly checked by folding the UCHs in chaperones with different cage size.

Knotting probability of UCH-L1 during folding. To qualitatively describe topological properties along the folding pathway, we introduce the total knotting probability $K(Q)$ (as a function of Q describing progress in folding). This probability is composed mostly of the probability of 3_1 ($K(Q)_{3_1}$) and 5_2 ($K(Q)_{5_2}$) knots presence. In Fig 3A we show $K(Q)$ calculated under confining conditions at three different temperatures—below (110) and around (115) estimated T_f for the bulk, and at estimated T_f for the confinement (120). The T_f values are estimated based on the chevron plot described in the following section. The shape of $K(Q)$ strongly depends on the temperature. Below T_f in the confinement (T eq. 110 and 115), there is a local maximum of $K(Q)$ at $Q \sim 0.7$, which decreases and shifts to the lower values of Q with increasing temperature. Decomposition of $K(Q)$ on the knot type directly shows that this maximum comes from the probability of the 3_1 knot ($K(Q)_{3_1}$), see Fig 3B. The height of the
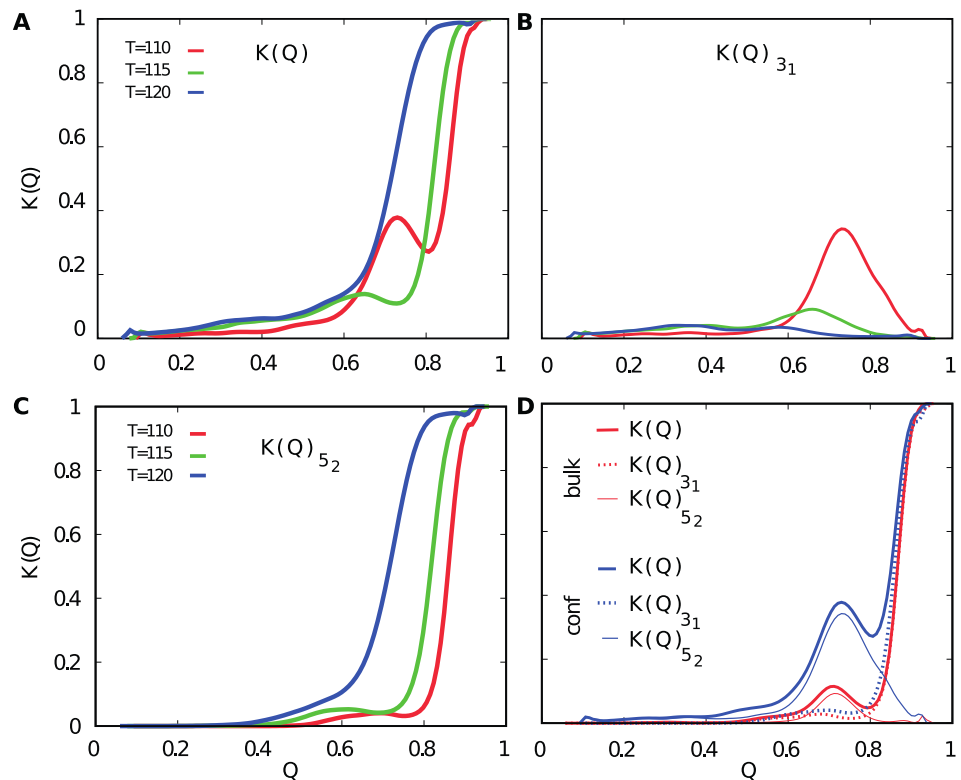
Fig 3. Plots of knotting probability as a function of Q during folding of UCH-L1 protein (PDB code 3IRT) in different conditions and temperatures. A: The total knotting probability $K(Q)_T$ (all possible type of knots) in the confinement. B: The knotting probability of a 3_1 knot ($K(Q)_{3_1}$) in the confinement. C: The knotting probability of 5_2 knot ($K(Q)_{5_2}$) in the confinement. D: The comparison of $K(Q)$, $K(Q)_{5_2}$ and $K(Q)_{3_1}$ in bulk and the confinement at $T = 110$.

https://doi.org/10.1371/journal.pcbi.1005970.g003

peak decreases at higher temperatures, which is in accordance with the fact that the F_C pathway (during which temporary 3_1 is formed) becomes less probable. Finally, at the higher temperature, corresponding to the T_f for confinement, the curve $K(Q)$ is almost monotonic and almost coincident with the shape of $K(Q)_{5_2}$ presented in the Fig 3C. This shows that in the majority of cases, the protein folds through intermediate configuration corresponding to the 5_2 knot. The difference between $K(Q)$ and $K(Q)_{5_2}$ for low values of Q stems from random, short-lived knots discussed in detail in the last section.

The position at which the total knotting probability $K(Q) = 0.5$, implies that in the majority of the successful folding routes, knotting is one of the latest steps ($Q > 0.7$ in all cases). With an increasing temperature $K(Q)$ and $K(Q)_{5_2}$ curves shift towards lower values of Q (Fig 3A and 3C, S1 Appendix, part 3) as a consequence of overall destabilization of the protein's structure.

Investigation of all considered knotting probabilities at the temperature close to the T_f for bulk (110) shows that the probability of forming the 5_2 knot is similar for both conditions (see Fig 3D). Moreover, much more populated F_C pathway in the confinement provides a much larger contribution of 3_1 knot probability. The introduction of confinement leads to the drastic change of the way how the folding process is realized. On the other hand, the very low probability of folding via the 3_1 knot in the bulk is postulated by the experimental results [33, 52].

Misfolding. The increased probability of F_C pathway at lower temperatures poses a question whether there is a temperature in which the F_C pathway dominates? Such considerations would shed light on the energetic barriers on both folding pathways. However, for low temperature, there is no sufficient thermal energy to break the contacts which formed to fast. This effect is especially evident in the confinement where the confinement enforces the long-distance contacts independently on the proper folding order (Table 2). Nevertheless, in both conditions for temperature close to T_f the fraction of misfolded structures is negligible. The misfolded structures can be divided into two topologically distinct types: 3_1 knotted (more common) and unknotted (less common) conformations. There is, however, no dominating structure of misfolded product. Some examples are shown in Fig 4. To check whether the misfolded structures are not just artifacts of a too simplified model, we rebuilt the side chains from $C\alpha$ trace using Modeller. We obtained all-atom models with DOPE potential lower than -17.000 suggesting, that the misfolded conformations could be actually found in real conditions (see also S1 Appendix, part 4).

Unfolding of UCHs. In our simulations, the unfolding process is, in general, the reversal of the folding, which is again consistent with experimental results [30]. In both conditions (bulk and confinement), unfolding has a common scenario in which an initial destabilization of the native structure results in a significant loss of the tertiary and secondary structure and untying. In particular, we observe two unfolding pathways differing in the occurrence of 3_1 knotted intermediate. Hence, the unfolding pathways are denoted throughout the text in analogy to their folding counterparts as U_N (direct $5_2 \rightarrow 0_1$) and U_C ($5_2 \rightarrow 3_1 \rightarrow 0_1$).

The unfolding process of UCH-L1 is realized mostly by the U_N pathway. Independently of the confining conditions and temperature, we identify more than 85% of unfolding trajectories in which the UCH-L1 unfolded in this way. Similarly to folding, this prevalence comes from higher stabilization of the C-terminus. This explains the fact, that the probability for U_N pathway rises in both conditions, confinement, and bulk, with an increase of temperature up to $T = 135$ (Table 2). Above this temperature, due to thermal fluctuations and unfolding rate, the possible 3_1 knot probably lives too short to be detected and both termini can be unplugged in comparably short time.
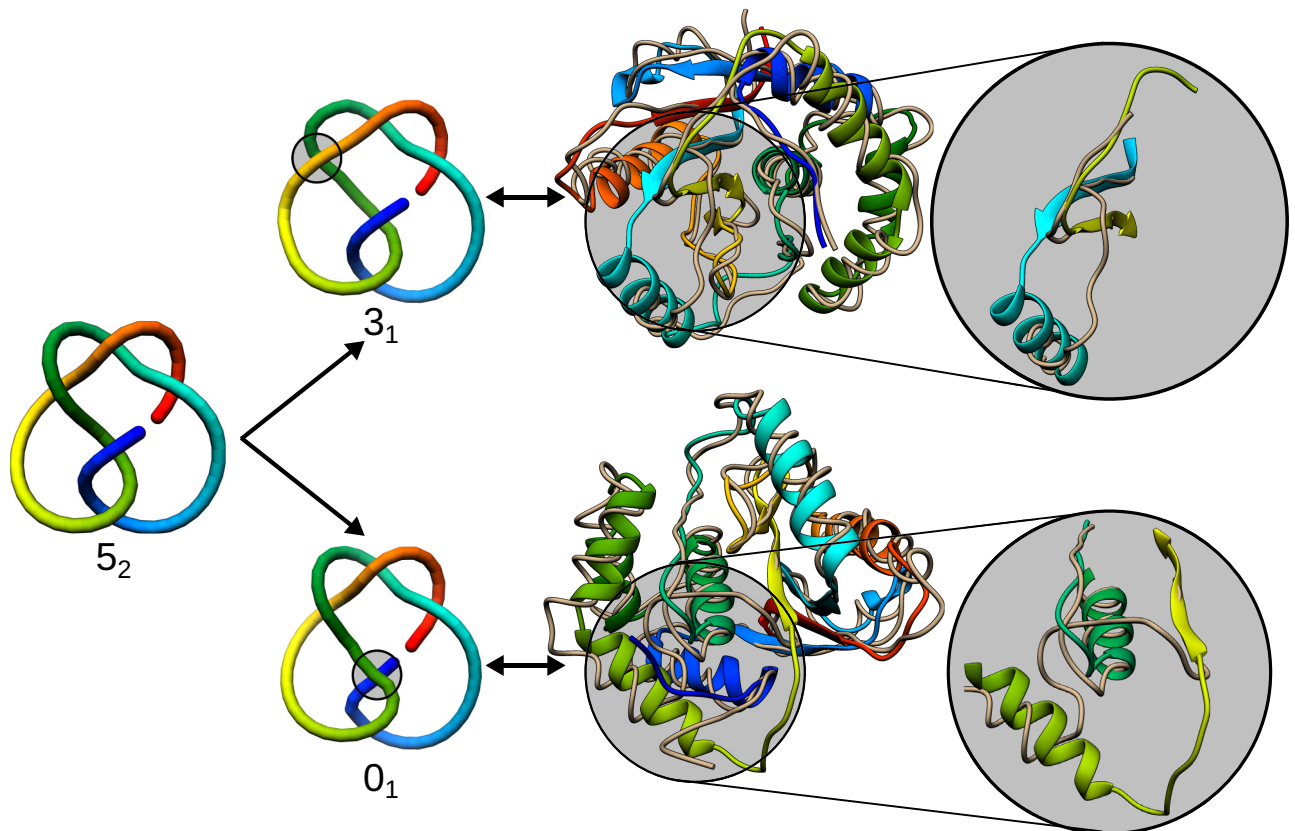
Fig 4. Misfolded structures obtained in UCH-L1 folding. Left: Changing the crossings in 5_2 knot results either in 3_1 or 0_1 knot (the crossings changed are marked with a circle). Middle: Comparison of the native structure of UCH-L1 (rainbow) with its misfolded structure (sandy)—upper panel—3_1, bottom—0_1 knot. Encircled is the crucial change in topology. Right: Enlargement of the topology-changing fragment. Note the different arrangement of sandy part of the chain.

The dependence of unfolding pathway on the knot tails stabilization is also visible comparing different UCHs. Although for all studied proteins, the U_N is the dominating pathway, for UCH-L5 (PDB id 4I6N) with the longest N-terminus, the probability of U_C pathway reaches 39% in bulk at T = 125, which is about 7 times more frequent than in the case of UCH-L1 at the same temperature. On the other hand, the structure with the shortest C-terminus (PDB id 2LEN) has the lowest probability of U_C pathway (for exact numbers see S1 Appendix, part 5).

Kinetics in the bulk and in the confinement

The UCHs are known to fold and refold along two parallel pathways, each featuring one slow and one fast phase [30, 31, 52]. Therefore, to correlate our model with experimental results we recreated the chevron plot for UCH-L1, with the temperature as a denaturant. In conventional chevron plots, the folding constants are calculated based on the time dependence of e.g. fluorescence. The fluorescence of a protein's tryptophan depends on its neighborhood. Hence, the fluorescence trace can be understood as a measure of similarity of the tryptophan neighborhood to the native structure. In our simulations, such a measure is given by the fraction of the native contacts—Q. Therefore, we calculated the average Q as a function of time (representative trace in Fig 5A). The average was taken over all simulations in a given T and in each condition (bulk/confinement). Next, we fitted the smoothed $Q_{aver}(t)$ with the sum of exponential functions. In particular, we fitted the trace with the sum of the highest number of exponents,
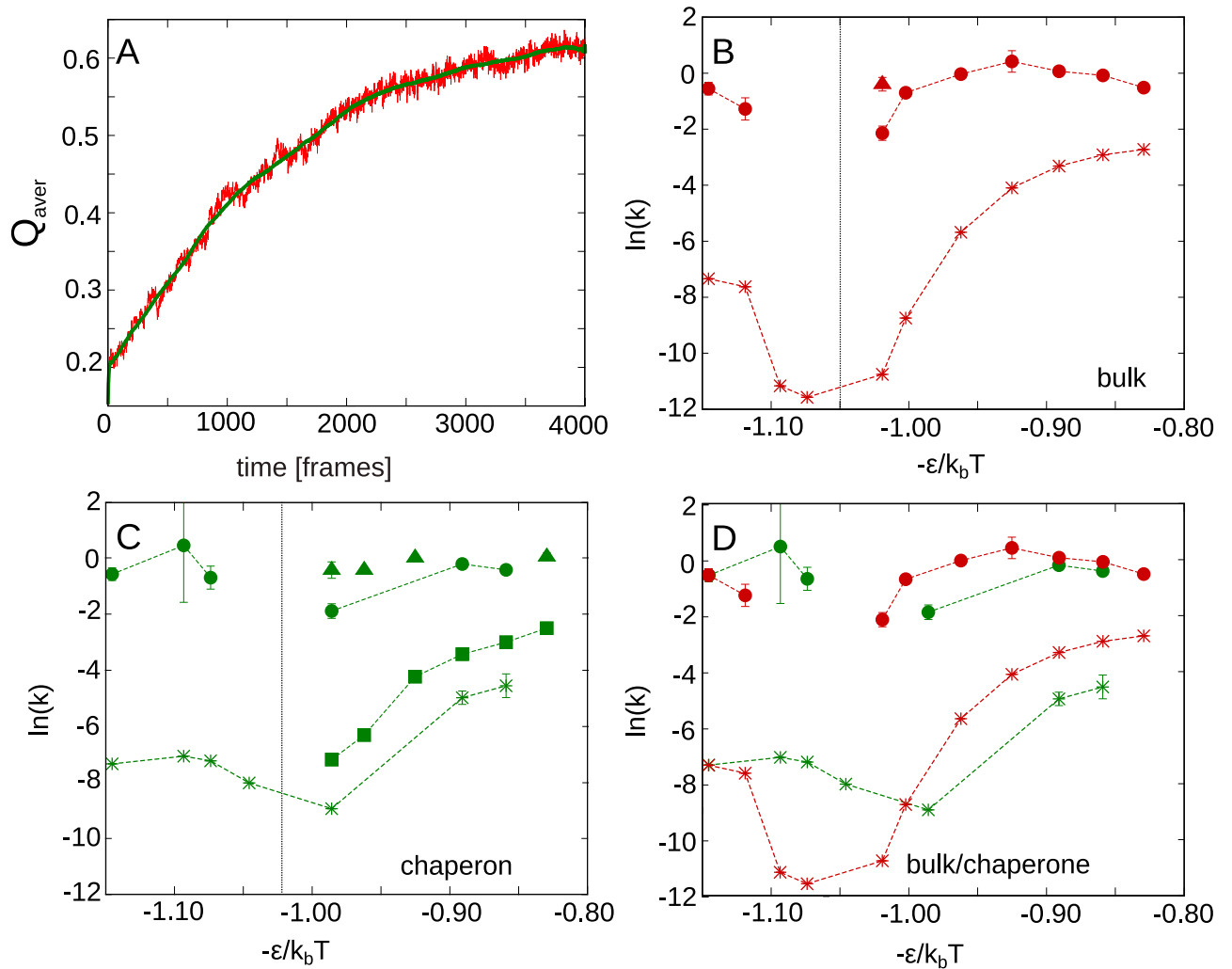
Fig 5. Simulated chevron plot for UCH-L1. A: Representative mean Q as the function of time (red) and smoothed curve (green). B: Chevron plot obtained for bulk. C: Chevron plot obtained for the confinement. B and C comes from fitting od sum of exponentials to plot in A. D: comparison of one slow and one fast phase for bulk (red) and confinement (green). The dashed lines present the expected chevron plot. Fitting the equation describing chevron plot results in an approximate T_f equal 114 ($-\epsilon/k_bT = -1.05$) for bulk and 120 ($-\epsilon/k_bT = -0.99$) for the confinement. In B, C, D the fitting error bars higher than 5% of value were shown.

https://doi.org/10.1371/journal.pcbi.1005970.g005

for which the fitting errors were lower than 5% of the value. The details of the plot along with the values of obtained constants and errors are presented in S1 Appendix, part 6. Although $Q_{aver}(t)$ is only a rough equivalent for the fluorescence, in almost all cases we were able to decompose the trace as a sum of 2-4 exponentials, i.e. to find up to 2 fast and 2 slow phases. These data are shown in Fig 5B (bulk) and 5C (confinement) in the form of chevron plot with the inverse of temperature (precisely $-\epsilon/k_bT$) mimicking the denaturant concentration [53]. For the consistency with conventional chevron plot, we plot the logarithm of $k = 1/\tau$ where τ is a characteristic time of a given phase. The obtained values create the trends characteristic for chevron plots, therefore they were connected by dashed lines. Note that in some cases the connection of values is arbitrary. For most cases, the fitting error was an order of magnitude smaller than the value obtained (Fig 5), and those, for which the error was higher (e.g. the point for confinement, $-\epsilon/k_bT = -1.10$), still corresponds to reasonable values.

The presence of the fast and slow phase during folding/unfolding process shows that our results are clearly consistent with previous experimental observations [30, 52]. However, in most cases, we were not able to determine four individual phases (as in the experiment). This may be due to similar characteristic times of separate phases, the model imperfection, or because of a much more complicated folding pathway. Indeed, the curvature of the limbs of the chevron plot indicates more complicated mechanism in each phase, again consistently with the experiment [33]. The detailed analysis of folding/unfolding pathway should be the next step in investigating of these proteins.

To determine the influence of the confinement, we compared the "most complete" kinetic trace for the slow and the fast phases for bulk and confinement (Fig 5D). The slowest phases can be fitted to an equation describing chevron plot, which yields an approximate T_f equal 114 ($-\epsilon/k_bT = -1.05$) for bulk and 120 ($-\epsilon/k_bT = -0.99$) for the confinement (for details see S1 Appendix, part 6). This indicates that confinement stabilizes UCH as it was observed for proteins with trivial topology [44, 54]. Moreover, the chevron plot indicates that the confinement significantly accelerates the folding process, especially the slow phase. In particular, the simulations in the T_f for bulk were not accessible computationally due to very slow rates, while they were accessible in T_f for the confinement. This enables us to calculate near-equilibrium $F(Q)$ dependence for confinement, which in principle could give additional information on UCHs folding. However in this case, due to the complexity of the folding landscape, the standard ways of its representation do not reveal any new information (S1 Appendix, part 7).

Regardless of the conditions, a collapse of the protein (the first phase of folding) occurs relatively fast, which stays in accordance with the experimental results, that the knotting (occurring in our model for $Q > 0.7$) should be the rate-limiting step [36]. Therefore in our case, the fast phase corresponds to arriving at collapsed, non-knotted form (first part of folding) and the slow phase should correspond to knotting and subsequential reaching of the native structure. The impact of the confinement on the slow phase indicates that the confinement facilitates knotting by restricting the conformational space of the termini. On the other hand, the confinement slows down the unfolding process by slowing down the unknotting—note the change of order in the curves in Fig 5D. The lower unfolding rates in the confinement may be also a result of the *retying* during unfolding (discussed in the next section). As it turns out that the once unfolded knot has a higher probability to retie in the confinement which results in higher knot stability and slower unfolding. Slow unfolding is again in agreement with intuition and experimental observation made for proteins with trivial topology [55].

To additionally investigate the influence of the confinement on both unfolded and folded state, we determined the average asphericity [56] parameter for bulk and confinement in both states. Again, the asphericity of folded state was comparable in both conditions, indicating, that the confinement does not influence the near-folded structures significantly. On the other hand, the asphericity of unfolded state was different in the confinement than in bulk, showing the influence of confinement on the unfolded basin (S1 Appendix, part 8).

Short-lived and random knots

The probability of knot presence in a polymer chain increases rapidly with its length. As a result, it is highly probable that the sufficiently long polymer will spontaneously form a knot. However, the fraction of knotted proteins is far lower than in the case of equally long polymers [57]. Moreover, the spontaneous self-tying of protein chains in the denatured state was not reported so far even in the natively deeply-knotted structures [22, 58, 59] or in the case of small knotted proteins in confinement [29]. However, in the case of UCHs we observe a
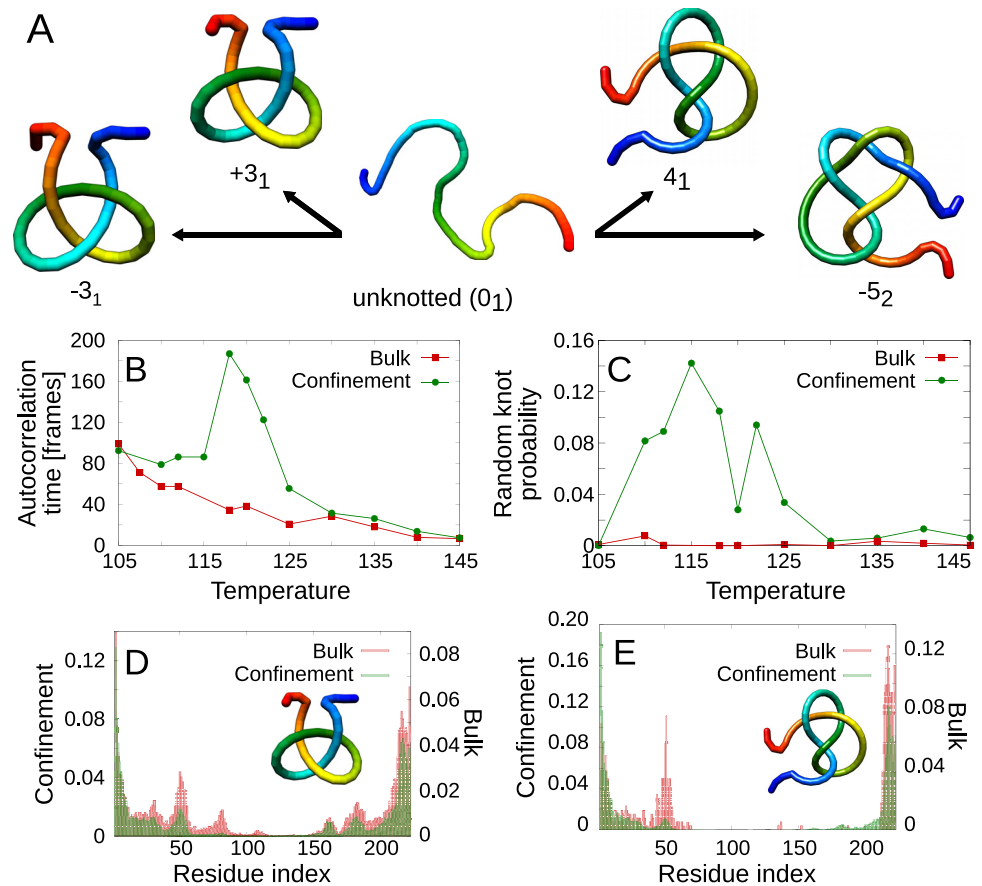
Fig 6. Random knots observed in the unfolded basin of UCH-L1. A: Schematic drawings of the type of observed random knots. B: Autocorrelation time in the unfolded basin as a function of temperature T. C: Random knot probability (for details see main text). D: preferred locations of a 3_1 and E: 4_1 knot's termini. In the plots, data for the confinement (green) and bulk (red).

significant fraction of (in most cases short-lived) knots, appearing during folding/unfolding pathway, or in the denatured state (with $Q < 0.2$).

UCHs can be knotted in the unfolded basin. Opposed to other proteins, even deeply knotted ones, in a case of UCHs, in the denatured state we observe various knot types: native-like $+3_1$ and -5_2 knots, but also non-native -3_1 and 4_1 knots (Fig 6A). In most cases, these are short-lived knots (S1 Appendix, part 9) and their probability decreases with knot complexity (in accordance with the theoretical expression for knot probability [60]). To compare the conditions, we calculated the fraction of simulations in which at least one random knot occurred. Similarly to previous results, we observe a much higher number of trajectories with random knots in the confinement. Indeed, in $T = 120$ the random knot in the denatured state occurred in 92% of trajectories (for raw data on random knots see S1 Appendix, part 9).

The investigation of the number of short-lived knots within a single trajectory is also possible. However, this has to be done with a greater care in order to count only those knots which are not time-correlated. Therefore, at first, we calculated the autocorrelation time as a function of temperature (Fig 6B). This allowed us to calculate the mean number of random knots occurring during the trajectory. In particular, in bulk at low temperatures, there is hardly 1 random knot (in the unfolded state) per 2 trajectories, and at higher temperatures, the probability is even smaller. On the other hand, in the confinement, there are up to 4 random knots on

average in one trajectory for $T = T_f$. The temperature dependence of these two effects (number of random knots per trajectory and number of trajectories with random knots) are however strictly connected to the time the protein spends in the unfolded basin. To quantify that effect, we calculated the total time the protein is unfolded at a given temperature and divided it into pieces of the length of autocorrelation time for given temperature. Next, we calculated what is the fraction of the number of intervals, in which the protein was knotted, compared to the whole number of such intervals at a given temperature. This quantity we will further call the random knot probability and we present the results in Fig 6C.

Again, the probability of knot formation is significantly larger in the confinement and is highest close to $T = T_f$. Under such conditions, the temporary conformations are relatively stable (in contrary to high temperatures), and the protein can search larger part of conformational space before folding (in contrary to lower temperatures). Furthermore, the confinement favors more complex knots. There is only a small number of cases where the 4_1 or 5_2 knot occurs in bulk, while there is a meaningful probability of a 4_1 and a considerable number of 5_2 knotted structures in the confinement (for examples see S1 Appendix, part 9). This stays in accordance with the similar behavior of polymers, for which the confinement favors more complex knots [61, 62].

To check the geometry of the short-lived knots we determined the preferred location of a knot's ends (Fig 6D and 6E). It turned out that the confinement favors shallow knots (both 3_1 and 4_1), while in the bulk there is a noticeably higher fraction of structures with a deeper knot, with the N-tail threaded by approximately 50 residues. Such knots occur by interaction of helix H8 (indices 193-208) with β strand B2 (46-54), with the formation of the twisted loop. This loop may be threaded by the N-terminus, induced by the interaction with helices H4-H7 (91-147). Depending on the threading direction, the resulting structure may have a 3_1, or 4_1 knot topology (examples in S1 Appendix, part 9). The fact that the deep knots occur more often in bulk may be somewhat intriguing as the confinement should favor formation of the loop-forming, long-distance B2-H8 interaction. As that surely is the case, the confinement probably also squeezes the newly formed loop, hindering threading due to entropic reasons.

The closer analysis of the autocorrelation time (Fig 6B) reveals that the confinement induces longer correlation times, however, this effect decreases with rising temperature (Fig 6E). Longer correlation time means that in our simulations, the relative change of the structure is slower in the confinement, than in bulk. On the other hand, folding in the confinement is in general faster, therefore the existence of the confinement either stabilizes the native contacts or hinders approximation of parts of the chain, which do not form such contacts. In order to discriminate between the two, first, we compared the mean probability of contact breaking in the confinement and in bulk (details in S1 Appendix, part 10). It turned out that the probability of native contact breaking is lower in confinement, which means that the contacts are more stabilized in the confinement relative to bulk. To investigate the destabilization of structures, with spatially close pieces of chain, which do not form native contacts, we looked at every pair of sufficiently close beads (less than 6Å) which do not form a contact in the native structure. We checked, how often such pair separates for distance larger than 6Å in-between consecutive frames. It turns out that the probability of such pair separation is similar (for intermediate temperatures), or significantly higher (for larger T) in the confinements than in bulk. This proves that the confinement works in both ways—stabilizes the native interaction in the same time destabilizing the structures in which pieces of chains which do not form native contacts are spatially close.

Short-lived knots may appear during folding or unfolding. During unfolding, the spontaneous *retying* phenomena can be frequently observed. By *retying*, we define an event, when the partially unfolded, untied protein with the still preserved native-like bias reestablishes the
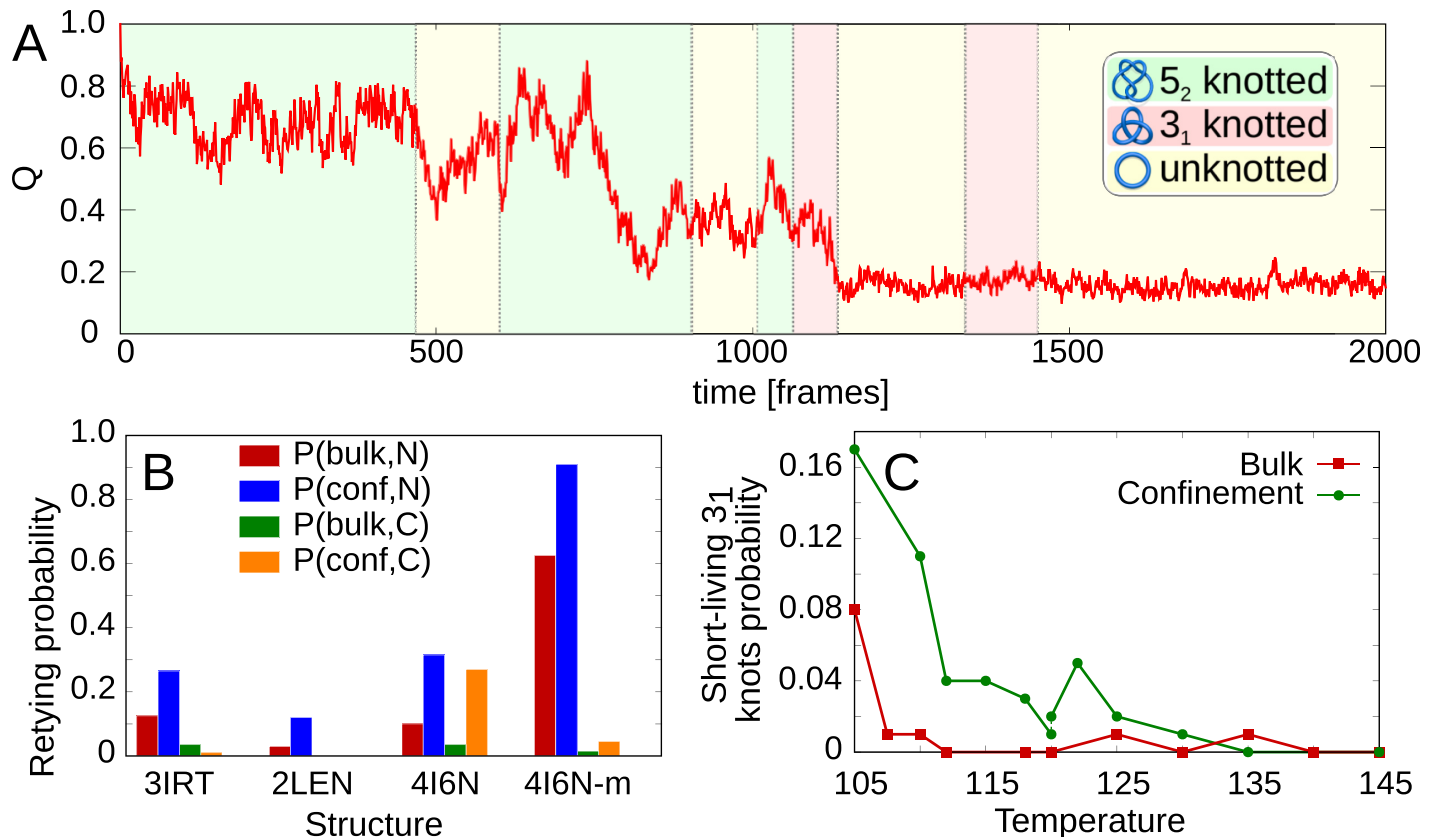
Fig 7. Short-lived knots during folding and unfolding. A: Exemplary unfolding trajectory with the topology indicated by color stripes. B: The retying probability for structures differing in knot tail length. C: The probability of formation of short-lived knots on folding/unfolding pathway. For the confinement at $T = 120$ bottom dot denotes short-lived knots on folding, top—unfolding pathway.

knot as shown in Fig 7A. All the investigated UCHs can retie to a 3_1 or 5_2 knot. The *retying* probability, P_{retie} of a protein, depends on the tails length (Fig 7B, data in S1 Appendix, part 11). For example, the structure with the longest C-terminal knot tail (UCH-L1, PDB code 2LEN) exhibits the probability of retying by C-terminus equal to zero in both conditions. Similarly, the reduction of the N-terminus by 7 residues (from structure 4I6N to 4I6N-m in UCH-L5) significantly increases the N-terminus retying probability. On the other hand, the 3IRT structure (UCH-L1) has the same N-terminus knot tail length as 4I6N-m (UCH-L5), but significantly lower retying probability. This comes from the distribution of native contacts in both structures. The number of the native contacts formed between the N-terminus and its knotting loop for 3IRT (9) is almost two times smaller than for 4I6N-m (17). Those extra native contacts provide an additional force increasing the probability of the retying of 4I6N-m. Independent of the tail length however, the confinement increases the probability of retying. This may also be one of the reasons of deceleration of the unfolding in the confinement (compare Fig 5D), and therefore increased knot stability.

The short-lived knots occur also on a folding pathway. This may be especially bothering, as a non-natively knotted structure may be prone to oligomerization, causing in turn serious neurodegenerative diseases [33]. On the folding pathway, we observed an existence of temporary 3_1, 4_1 and 5_2 knots, which untie after a short time period. The 3_1 and 5_2 knots are usually a result of the premature placement of N- or C-terminus in its native position, in which (in early phases of folding) the proper terminus is not stabilized enough, therefore it can slip out of the

loop. Consequently, the 3_1 short-lived knots are more common in confinement and at low temperature, as in such conditions the F_C pathway is more probable.

The 4_1 knot on the folding pathway occurs rarely. Consistently with the findings presented above, this type of a knot is present more frequently in confinement at low temperatures—3% of folding pathways contain a short-lived 4_1 knot in the confinement at $T = 105$. For higher temperatures the probability of obtaining a 4_1 knot diminishes, being negligible for $T = T_f$ in both conditions. The 4_1 knot is very shallow and occurs usually by threading a loose N-terminus through the loop formed by unstructured helices H4-H7 (91-147), similarly as for random knots in the denaturated state.

Conclusions

We found that confinement leads to faster and more efficient folding of UCH proteins for two reasons. First, encapsulation provides the possibility to fold via an alternative pathway. More precisely, the confinement facilitates folding via the trefoil knot (the F_C pathway) for entropic reasons, while it does not affect folding via the F_N pathway (direct tying, the N-terminus folds the last). This surprising behavior is supported by the experimental observation of uneven influence of chaperonin [63] on the substrate protein rhodanese, which decelerates the folding of the C-terminal domain, but leaves the folding rate of the N-terminal domain unaffected [64]. Second, at the same confinement stabilizes native interactions and destabilizes non-native ones in comparison to bulk, and thus it reduces the height of the free energy barrier and accelerates the folding rate as it was observed for a protein with trivial topology [55, 65, 66]. The two pathways clearly distinguishable in our analysis are in accordance with two pathways identified in experiment [30]. However, there is still no technique, which could determine the topology of the protein during folding *in vitro*, which prevents from direct validation of our results. Some insights can however be given by the study of the tryptophan variants of UCH [33]. In particular, it was shown that the pathways differ in the structure of intermediates for which highly stable central β-sheet core and flanking α-helices and loop regions are formed differently. This is in accordance with our results, however detailed analysis of folding pathway, with special emphasis put on the location of tryptophan mutations, is required to precisely compare the experimental and theoretical results.

On the other hand, our results show that the confinement introduced by the chaperone-like cage decelerates the unfolding of UCH proteins. Firstly due to a decrease in the effective mobility of the protein backbone upon encapsulation and under topological constrains, which reduce the rate at which new configuration can be explored (especially in twisted loops). This argument without topological contribution is used to explain lower folding rate for a protein with the trivial topology [64]. Secondly, the decelerated unfolding in the confinement is caused by the retying phenomena. It is worth pointing out that the retying phenomena can be used by other knotted proteins with a rather shallow knot, e.g. carbonic anhydrase, to stabilize the structure in a crowded environment (moreover carbonic anhydrase structures with deeper knots also start to be crystallized).

Under the confinement, a significant number of short living knots is observed in the denatured state and in folding and unfolding routes of UCH proteins, what has not been reported for other knotted proteins. More and more complicated knots are more common to occur upon encapsulation, which is in the agreement with the polymer theory [61, 62]. These knots seem to have only positive effect, i.e. their formation accelerates folding. In principle, deeply knotted structures could lead to misfolding, but contrary to the situation in the bulk, they are not formed due to the constrained configurational space.

In summary, we took advantages of structure based model and knot theory, and made the step forward in characterizing folding/unfolding routes for UCH proteins identified experimentally in [30]. We identified possible oligomerization-prone forms of UCHs, which may cause neurodegenerative diseases. We found that weak confinement smooths the rough and not continuous free energy landscape of UCH proteins in a subtle way, e.g. enhancing an indirect tying route. However, at low temperature or strong confinement slower folding should be again observed due to restriction on indirect tying. The deceleration under strong confinement was suggested for a protein with the trivial topology in [63].

Supporting information

S1 Appendix. Supporting tables and figures.
(PDF)

Author Contributions

Conceptualization: Yani Zhao, Szymon Niewieczerzal, Joanna I. Sulkowska.

Data curation: Pawel Dabrowski-Tumanski.

Formal analysis: Yani Zhao, Pawel Dabrowski-Tumanski, Szymon Niewieczerzal.

Funding acquisition: Joanna I. Sulkowska.

Investigation: Yani Zhao, Pawel Dabrowski-Tumanski, Szymon Niewieczerzal, Joanna I. Sulkowska.

Methodology: Pawel Dabrowski-Tumanski, Szymon Niewieczerzal.

Project administration: Joanna I. Sulkowska.

Software: Szymon Niewieczerzal, Joanna I. Sulkowska.

Supervision: Szymon Niewieczerzal, Joanna I. Sulkowska.

Validation: Pawel Dabrowski-Tumanski, Szymon Niewieczerzal, Joanna I. Sulkowska.

Visualization: Pawel Dabrowski-Tumanski.

Writing – original draft: Yani Zhao, Szymon Niewieczerzal.

Writing – review & editing: Pawel Dabrowski-Tumanski, Joanna I. Sulkowska.

References

1. Sułkowska JI, Sułkowski P, Szymczak P, Cieplak M. Stabilizing effect of knots on proteins. Proceedings of the National Academy of Sciences. 2008; 105(50):19714–19719. https://doi.org/10.1073/pnas.0805468105

2. Sayre TC, Lee TM, King NP, Yeates TO. Protein stabilization in a highly knotted protein polymer. Protein Engineering Design and Selection. 2011; 24(8):627–630. https://doi.org/10.1093/protein/gzr024

3. Dabrowski-Tumanski P, Stasiak A, Sulkowska JI. In Search of Functional Advantages of Knots in Proteins. PloS one. 2016; 11(11):e0165986. https://doi.org/10.1371/journal.pone.0165986 PMID: 27806097

4. Jamroz M, Niemyska W, Rawdon EJ, Stasiak A, Millett KC, Sułkowski P, et al. KnotProt: a database of proteins with knots and slipknots. Nucleic acids research. 2015; 43(D1):D306–D314. https://doi.org/10.1093/nar/gku1059 PMID: 25361973

5. Sułkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A. Conservation of complex knotting and slipknotting patterns in proteins. Proceedings of the National Academy of Sciences. 2012; 109(26): E1715–E1723. https://doi.org/10.1073/pnas.1205918109

6. Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, Sixma TK, et al. A genomic and functional inventory of deubiquitinating enzymes. Cell. 2005; 123(5):773–786. https://doi.org/10.1016/j.cell.2005.11.007 PMID: 16325574

7. Miyoshi Y, Nakayama S, Torikoshi Y, Tanaka S, Ishihara H, Taguchi T, et al. High expression of ubiquitin carboxy-terminal hydrolase-L1 and-L3 mRNA predicts early recurrence in patients with invasive breast cancer. Cancer science. 2006; 97(6):523–529. https://doi.org/10.1111/j.1349-7006.2006.00202.x PMID: 16734731

8. Wang L, Chen YJ, Xu K, Wang YY, Shen XZ, Tu RQ. High expression of UCH37 is significantly associated with poor prognosis in human epithelial ovarian cancer. Tumor Biology. 2014; 35(11):11427–11433. https://doi.org/10.1007/s13277-014-2446-3 PMID: 25123264

9. Wilkinson KD, Lee K, Deshpande S, Duerksen-Hughes P, Boss JM, Pohl J. The neuron-specific protein PGP 9.5 is a ubiquitin carboxyl-terminal hydrolase. Science. 1989; 246(4930):670–673. https://doi.org/10.1126/science.2530630 PMID: 2530630

10. Tezel E, Hibi K, Nagasaka T, Nakao A. PGP9. 5 as a prognostic factor in pancreatic cancer. Clinical Cancer Research. 2000; 6(12):4764–4767. PMID: 11156232

11. Takase T, Hibi K, Yamazaki T, Nakayama H, Taguchi M, Kasai Y, et al. PGP9. 5 overexpression in esophageal squamous cell carcinoma. Hepato-gastroenterology. 2002; 50(53):1278–1280.

12. Leiblich A, Cross SS, Catto JW, Pesce G, Hamdy FC, Rehman I. Human prostate cancer cells express neuroendocrine cell markers PGP 9.5 and chromogranin A. The Prostate. 2007; 67(16):1761–1769. https://doi.org/10.1002/pros.20654 PMID: 17929277

13. Takano T, Miyauchi A, Matsuzuka F, Yoshida H, Nakata Y, Kuma K, et al. PGP9. 5 mRNA could contribute to the molecular-based diagnosis of medullary thyroid carcinoma. European Journal of Cancer. 2004; 40(4):614–618. https://doi.org/10.1016/j.ejca.2003.11.016 PMID: 14962731

14. Mizukami H, Shirahata A, Goto T, Sakata M, Saito M, Ishibashi K, et al. PGP9. 5 methylation as a marker for metastatic colorectal cancer. Anticancer research. 2008; 28(5A):2697–2700. PMID: 19035297

15. Wan F, Miao X, Quraishi I, Kennedy V, Creek KE, Pirisi L. Gene expression changes during HPV-mediated carcinogenesis: A comparison between an in vitro cell model and cervical cancer. International Journal of Cancer. 2008; 123(1):32–40. https://doi.org/10.1002/ijc.23463 PMID: 18398830

16. Setsuie R, Wada K. The functions of UCH-L1 and its relation to neurodegenerative diseases. Neurochemistry international. 2007; 51(2):105–111. https://doi.org/10.1016/j.neuint.2007.05.007 PMID: 17586089

17. Sułkowska JI, Noel JK, Ramırez-Sarmiento CA, Rawdon EJ, Millett KC, Onuchic JN. Knotting pathways in proteins. Biochem Soc Trans. 2013; 41(2):523–527. https://doi.org/10.1042/BST20120342 PMID: 23514147

18. Jackson SE, 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. https://doi.org/10.1016/j.sbi.2016.10.002 PMID: 27794211

19. Noel JK, Sułkowska JI, Onuchic JN. Slipknotting upon native-like loop formation in a trefoil knot protein. Proceedings of the National Academy of Sciences. 2010; 107(35):15403–15408. https://doi.org/10.1073/pnas.1009522107

20. Noel JK, Onuchic JN, Sulkowska JI. Knotting a protein in explicit solvent. The Journal of Physical Chemistry Letters. 2013; 4(21):3570–3573. https://doi.org/10.1021/jz401842f

21. Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nature chemical biology. 2012; 8(2):147–153. https://doi.org/10.1038/nchembio.742

22. Sułkowska JI, Sułkowski P, Onuchic J. Dodging the crisis of folding proteins with knots. Proceedings of the National Academy of Sciences. 2009; 106(9):3119–3124. https://doi.org/10.1073/pnas.0811147106

23. Wallin S, Zeldovich KB, Shakhnovich EI. The folding mechanics of a knotted protein. Journal of molecular biology. 2007; 368(3):884–893. https://doi.org/10.1016/j.jmb.2007.02.035 PMID: 17368671

24. Chwastyk M, Cieplak M. Cotranslational folding of deeply knotted proteins. Journal of Physics: Condensed Matter. 2015; 27(35):354105. PMID: 26292194

25. Bölinger D, Sułkowska JI, Hsu HP, Mirny LA, Kardar M, Onuchic JN, et al. A Stevedore's protein knot. PLoS Comput Biol. 2010; 6(4):e1000731–e1000731. https://doi.org/10.1371/journal.pcbi.1000731 PMID: 20369018

26. Wang I, Chen SY, Hsu STD. Folding analysis of the most complex Stevedore?s protein knot. Scientific Reports. 2016; 6.

27. D'Adamo G, Micheletti C. Molecular crowding increases knots abundance in linear polymers. Macromolecules. 2015; 48(17):6337–6346. https://doi.org/10.1021/acs.macromol.5b01323

28. Soler MA, Rey A, Faísca PF. Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding. Physical Chemistry Chemical Physics. 2016; 18(38):26391–26403. https://doi.org/10.1039/C6CP05086G PMID: 27722468

29. Niewieczerzal S, Sulkowska JI. Knotting and unknotting proteins in the chaperonin cage: Effects of the excluded volume. PLoS One. 2017; 12:e0176744. https://doi.org/10.1371/journal.pone.0176744 PMID: 28489858

30. Andersson FI, Pina DG, Mallam AL, Blaser G, Jackson SE. Untangling the folding mechanism of the 52-knotted protein UCH-L3. Febs Journal. 2009; 276(9):2625–2635. https://doi.org/10.1111/j.1742-4658.2009.06990.x PMID: 19476499

31. Lee YTC, Chang CY, Chen SY, Pan YR, Ho MR, Hsu STD. Entropic stabilization of a deubiquitinase provides conformational plasticity and slow unfolding kinetics beneficial for functioning on the proteasome. Scientific Reports. 2017; 7:45174. https://doi.org/10.1038/srep45174 PMID: 28338014

32. Andersson FI, Werrell EF, McMorran L, Crone WJ, Das C, Hsu STD, et al. The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1. Journal of molecular biology. 2011; 407(2):261–272. https://doi.org/10.1016/j.jmb.2010.12.029 PMID: 21251915

33. Zhang H, Jackson SE. Characterization of the Folding of a 5 2-Knotted Protein Using Engineered Single-Tryptophan Variants. Biophysical Journal. 2016; 111(12):2587–2599. https://doi.org/10.1016/j.bpj.2016.10.029 PMID: 28002735

34. Andersson FI, Jackson SE, Hsu STD. Backbone assignments of the 26 kDa neuron-specific ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1). Biomolecular NMR assignments. 2010; 4(1):41–43. https://doi.org/10.1007/s12104-009-9203-3 PMID: 20012716

35. Soler MA, Nunes A, Faísca PF. Effects of knot type in the folding of topologically complex lattice proteins. The Journal of chemical physics. 2014; 141(2):07B607_1. https://doi.org/10.1063/1.4886401

36. Ziegler F, Lim NC, Mandal SS, Pelz B, Ng WP, Schlierf M, et al. Knotting and unknotting of a protein in single molecule experiments. Proceedings of the National Academy of Sciences. 2016; p. 201600614.

37. Webb B, Sali A. Protein structure modeling with MODELLER. Protein Structure Prediction. 2014; p. 1–15. https://doi.org/10.1007/978-1-4939-0366-5_1

38. Shen My, Sali A. Statistical potential for assessment and prediction of protein structures. Protein science. 2006; 15(11):2507–2524. https://doi.org/10.1110/ps.062416606

39. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. Journal of molecular biology. 2000; 298(5):937–953. https://doi.org/10.1006/jmbi.2000.3693 PMID: 10801360

40. Sułkowska JI, Cieplak M. Selection of optimal variants of Gō-like models of proteins through studies of stretching. Biophysical journal. 2008; 95(7):3174–3191. https://doi.org/10.1529/biophysj.107.127233 PMID: 18567634

41. Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, et al. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. PLoS Comput Biol. 2016; 12(3):e1004794. https://doi.org/10.1371/journal.pcbi.1004794 PMID: 26963394

42. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. Journal of molecular biology. 1999; 290(1):253–266. https://doi.org/10.1006/jmbi.1999.2829 PMID: 10388571

43. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of chemical theory and computation. 2008; 4 (3):435–447. https://doi.org/10.1021/ct700301q PMID: 26620784

44. 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 (20):11367–11372. https://doi.org/10.1073/pnas.1831920100

45. Xu WX, Wang J, Wang W. Folding behavior of chaperonin-mediated substrate protein. Proteins: Structure, Function, and Bioinformatics. 2005; 61(4):777–794. https://doi.org/10.1002/prot.20689

46. Lucent D, Vishal V, Pande VS. Protein folding under confinement: a role for solvent. Proceedings of the National Academy of Sciences. 2007; 104(25):10430–10434. https://doi.org/10.1073/pnas.0608256104

47. Freyd P, Yetter D, Hoste J, Lickorish WR, Millett K, Ocneanu A. A new polynomial invariant of knots and links. Bulletin of the American Mathematical Society. 1985; 12(2):239–246. https://doi.org/10.1090/S0273-0979-1985-15361-3

48. Przytycki JH, Traczyk P. Invariants of links of Conway type. Kobe Journal of Mathematics. 1988; 4:115–139.

49. Ewing B, Millett KC. A load balanced algorithm for the calculation of the polynomial knot and link invariants. The mathematical heritage of CF Gauss. 1991; p. 225–266. https://doi.org/10.1142/9789814503457_0017

50. Millett KC, Rawdon EJ, Stasiak A, Sułkowska JI. Identifying knots in proteins; 2013.

51. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004; 25(13):1605–1612. https://doi.org/10.1002/jcc.20084 PMID: 15264254

52. Lou SC, Wetzel S, Zhang H, Crone EW, Lee YT, Jackson SE, et al. The Knotted Protein UCH-L1 Exhibits Partially Unfolded Forms under Native Conditions that Share Common Structural Features with Its Kinetic Folding Intermediates. Journal of molecular biology. 2016; 428(11):2507–2520. https://doi.org/10.1016/j.jmb.2016.04.002 PMID: 27067109

53. Shortle D, Chan HS, Dill KA. Modeling the effects of mutations on the denatured states of proteins. Protein Science. 1992; 1(2):201–215. https://doi.org/10.1002/pro.5560010202 PMID: 1304903

54. Rathore N, Knotts TA, de Pablo JJ. Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations. Biophysical journal. 2006; 90(5):1767–1773. https://doi.org/10.1529/biophysj.105.071076 PMID: 16361344

55. Mittal J, Best RB. Thermodynamics and kinetics of protein folding under confinement. Proceedings of the National Academy of Sciences. 2008; 105(51):20233–20238. https://doi.org/10.1073/pnas.0807742105

56. Van Giessen A, Szleifer I. Monte Carlo simulations of chain molecules in confined environments. The Journal of chemical physics. 1995; 102(22):9069–9076. https://doi.org/10.1063/1.468855

57. Lua RC, Grosberg AY. Statistics of knots, geometry of conformations, and evolution of proteins. PLoS Comput Biol. 2006; 2(5):e45. https://doi.org/10.1371/journal.pcbi.0020045 PMID: 16710448

58. 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(44):17789–17794. https://doi.org/10.1073/pnas.1201807109

59. Covino R, Škrbić T, Faccioli P, Micheletti C, et al. The role of non-native interactions in the folding of knotted proteins: insights from molecular dynamics simulations. Biomolecules. 2013; 4(1):1–19. https://doi.org/10.3390/biom4010001 PMID: 24970203

60. Deguchi T, Tsurusaki K. A statistical study of random knotting using the Vassiliev invariants. Journal of Knot Theory and Its Ramifications. 1994; 3(03):321–353. https://doi.org/10.1142/S0218216594000241

61. Micheletti C, Marenduzzo D, Orlandini E, Summers D. Knotting of random ring polymers in confined spaces. The Journal of chemical physics. 2006; 124(6):064903. https://doi.org/10.1063/1.2162886

62. Micheletti C, Marenduzzo D, Orlandini E, Sumners D. Simulations of knotting in confined circular DNA. Biophysical journal. 2008; 95(8):3591–3599. https://doi.org/10.1529/biophysj.108.137653 PMID: 18621819

63. Jewett A, Baumketner A, Shea JE. 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(36):13192–13197. https://doi.org/10.1073/pnas.0400720101 PMID: 15331776

64. Hofmann H, Hillger F, Pfeil SH, Hoffmann A, Streich D, Haenni D, et al. Single-molecule spectroscopy of protein folding in a chaperonin cage. Proceedings of the National Academy of Sciences. 2010; 107(26):11793–11798. https://doi.org/10.1073/pnas.1002356107

65. Zhou HX, Dill KA. Stabilization of proteins in confined spaces. Biochemistry. 2001; 40(38):11289–11293. https://doi.org/10.1021/bi0155504 PMID: 11560476

66. Jewett AI, Shea JE. Reconciling theories of chaperonin accelerated folding with experimental evidence. Cellular and molecular life sciences. 2010; 67(2):255–276. https://doi.org/10.1007/s00018-009-0164-6 PMID: 19851829