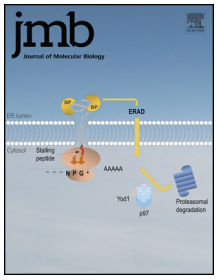

Proteins' Knotty Problems

Aleksandra I. Jarmolinska[1,2], Agata P. Perlinska[1,2], Robert Runkel[3], Benjamin Trefz[3,4], Helen M. Ginn[5], Peter Virnau[3] and Joanna I. Sulkowska[1]

*1 - **Centre of New Technologies,** University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland*
*2 - **College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences,** Banacha 2c, 02-097 Warsaw, Poland*
*3 - **Department of Physics,** Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz, Germany*
*4 - **Graduate School Material Science in Mainz,** Staudinger Weg 9, 55128 Mainz, Germany*
*5 - **Division of Structural Biology,** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

***Correspondence to Joanna I. Sulkowska:** jsulkowska@cent.uw.edu.pl*
https://doi.org/10.1016/j.jmb.2018.10.012
Edited by James Bowie

Abstract

Knots in proteins are increasingly being recognized as an important structural concept, and the folding of these peculiar structures still poses considerable challenges. From a functional point of view, most protein knots discovered so far are either enzymes or DNA-binding proteins. Our comprehensive topological analysis of the Protein Data Bank reveals several novel structures including knotted mitochondrial proteins and the most deeply embedded protein knot discovered so far. For the latter, we propose a novel folding pathway based on the idea that a loose knot forms at a terminus and slides to its native position. For the mitochondrial proteins, we discuss the folding problem from the perspective of transport and suggest that they fold inside the mitochondria. We also discuss the evolutionary origin of a novel class of knotted membrane proteins and argue that a novel knotted DNA-binding protein constitutes a new fold. Finally, we have also discovered a knot in an artificially designed protein structure.

Introduction

Topological structures are an emerging field in modern chemistry [1]. Indeed, one of the laureates of the Nobel Prize for Chemistry in 2016, J.-P. Sauvage, not only pioneered the synthesis of catenanes [2], which were later used in the construction of nano-machines, but also constructed the first artificial molecular trefoil knot in 1989 [3]. Knots also play an important role in biochemistry and have been observed to occur naturally in DNA [4–7] and proteins [8–13].

The study of knots in proteins in particular is a rather novel area of research because for a long time biologists thought it impossible to fold a protein into a knot. A trefoil in a carbonic anhydrase was the only known exception to the rule [14], but this knot is rather loose and dissolves when a few amino acids are cut from the C-terminus. Apart from that, the first systematic scan of the Protein Data Bank in 1994 [8] did not yield any novel knots, and it took two more years until the first deeply knotted structure was located in *S*-adenosylmethionine synthetase [9]. This discovery was rather remarkable as it is difficult to identify knots in proteins with the naked eye (see, e.g., Fig. 1). Most protein knots (like the ones in this manuscript) have indeed been identified by computer programs [10–13,15], some of which were ironically written to dismiss knotted structures in homology modeling. In recent years, however, the rather challenging task of identifying knots has become accessible to non-experts with the emergence of dedicated web servers [16–18].

Although knots are only mathematically well defined in closed curves, termini of proteins are typically located at the surface and can be connected in an unambiguous way [11,19]. So far, about a dozen distinct families of knotted proteins have been identified [13,20,21]. Most protein knots are simple

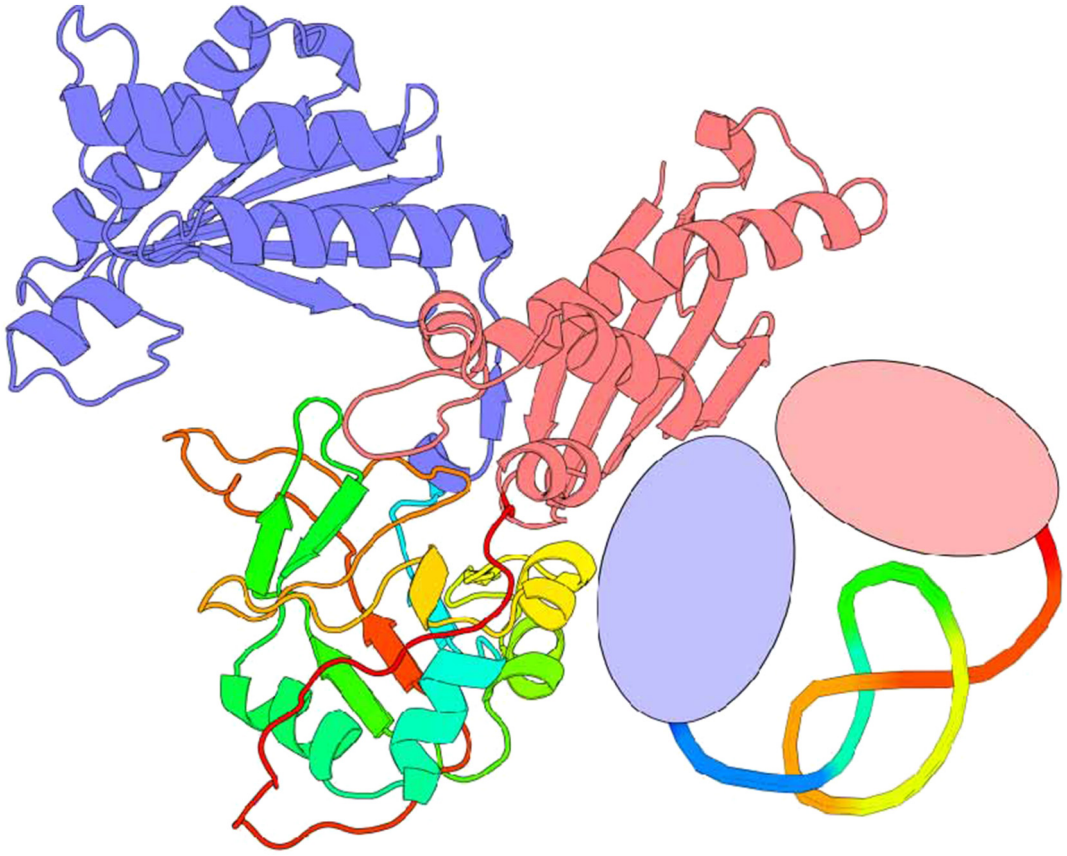

Fig. 1. Cartoon representation of the protein Tp0642 with the deepest knot discovered so far. As an inset, a reduced schematic representation of the protein backbone is shown (PDB code: 5jir, chain A). Colors change from red (N-terminus) to blue (C-terminus) to guide the eye.

trefoils (3_1) [18]; that is, they exhibit three minimal crossings in an appropriate projection onto a plane. Two proteins contain the figure-eight (4_1) knot, there is one 5_2 knot [11,15], and one knot with six crossings (6_1) [12]. Note that the number of protein knots is small, especially if contrasted with 1221 different folds in the current version (2.06) of SCOPe [22].

Although protein knots are rare, they appear in all kingdoms of life and tend to be preserved throughout evolution—pointing to an origin, which goes back hundreds of millions, potentially billions of years [13]. In recent years, significant progress has been achieved, like the first creation of an artificial protein knot [23] as well as experimental [24–33] and numerical studies [12,34–42] on folding. The latter in particular suggests that the folding of complicated protein knots can typically be achieved in a few or even a single global movement of a subchain [43] or as recently suggested directly on the ribosome [44].

Little is still known about the function of the knot in particular proteins. The only knotted protein, to which an unknotted homolog exists [11,45], a transcarbamylase [46], catalyzes a chemical reaction quite similar to its unknotted counterpart. For these proteins, simulations indicate that the knotted one is more mechanically resistant [34]. A few pairs of proteins that are analogous in function but have a different topology have been identified in methyltransferases [47]. The comparison of knotted and unknotted proteins responsible for methylation of tRNA has shown that, although the knot is thought to be a rigid part of the structure, it responds to the motions of the whole protein [48]. The trefoil knot in the TrmD protein, for example, is capable of transfering the signal coming from the substrates further through the protein by its internal movements. Mutations of key residues in the knot suppress its motion and make the protein unable to conduce the methyl transfer. Knot's involvement in the enzymatic function suggests that the search for its role should lay beyond the structure-induced durability. Apart from that, there have been speculations [11,13,34] and supporting numerical simulations [13,34,49–51] that knotting in proteins induces resistance to thermal and proteasomal degradation. Corresponding experimental investigations have, however, only been started recently[†] [31,32].

In this paper, we discuss in depth novel protein structures, which contain knots using topological and

structural analysis, numerical simulations and bioinformatic tools. After providing a comprehensive table of knotted proteins, we report the discovery of knotted mitochondrial proteins and propose potential folding pathways based on extensive numerical simulations. We argue that the folding into the knotted native state happens after the proteins are transported to the mitochondrion. We also report the discovery of a novel knotted protein from *Treponema pallidum*. This protein contains the deepest embedded knot known to date, which persists even when 100 amino acids are cleaved from both termini. Based on our numerical simulations, we suggest a novel folding pathway for this protein: A loose knot forms at one of the termini and then slides to its native position during folding. In addition, we discuss in detail two novel knotted membrane structures. These structures belong to two subbranches of Ca^{2+} antiporters but share barely more than 10% of their sequence. A reconstruction of a phylogenetic tree suggests that all members of this fold should be knotted. Also, we argue that a novel knotted fold in a DNA-binding protein constitutes a novel fold distinct from the knotted ribbon–helix–helix (RHH) superfamily. Finally, we show a new example of an artificially knotted protein, a single-chain TFIIA protein that retained its function, although its topology has changed.

Novel Protein Knots

Table 1 provides a non-redundant list of protein knots classified according to their function, with recent additions being marked in bold.[‡] The table includes a representative PDB code, the length of the chain, the knot type and the length of the knotted core, that is, the minimum chain segment, which still contains the knot [19]. This range together with chain length indicates how deeply a knot is embedded in the structure. We also provide information on the chirality of the knots. A knot can either be left

Table 1. A comprehensive list of knotted protein families with representative structures

Protein family	PDB code	Chain length	Knot (fingerprint)	Knotted core
Enzyme				
Methyltransferase (α/β knot)	1ns5A	153	**3_1+**	71–119
tRNA methyltransferase	1uakA	243	**3_1+**	87–130
rRNA methyltransferase	2egvA	229	**3_1+**	161–205
protein methyltransferase	5h5fA	224	**3_1+**	83–131
Carbonic anhydrase II	1lugA	259	**3_1+**	27–256
SAM synthetase	1fugA	383	**3_1+**	11–266
Transcarbamylase fold	1js1X	324	**3_1+**	172–237
N-acetylglucosamine deacetylase	5bu6A	264	**3_1+**	45–294
Pyridine synthase	5wa4A	310	**3_1+**	168–350
CII Ketol acid reductoisomerase	1yveL	513	**4_1**(3_1+)	320–525
Chromophore binding domain	2o9cA	319	**4_1**(4_1)	33–274
Ubiquitin C-terminal hydrolase	2etlA	223	**5_2-** $(3_1-,3_1-)$	5–219
α-Haloacid dehalogenase I	3bjxA	311	**6_1+** $(6_1+,4_1,3_1+)$	63–276
(**Mitochondrial apoptosis-inducing factor** 1[a])	5fmhA	486	**3_1-** (3_1-)	509–580
DNA binding				
Zinc-finger fold	2k0aA	109	**3_1-**	22–70
RHH superfamily				
MJO366	2efvA	82	**3_1-**	16–81
VirC2	2rh3A	121	**3_1-**	88–191
DndE	4lrvA	107	**3_1-**	9–100
Unknown function				
Protein from *T. pallidum*	5jirA	413	**3_1+**	202–313
Artificial proteins				
Artificially (designed) knotted protein	3mlgA	169	**3_1-**	32–133
Single-chain TFIIA	5m4sA	209	**3_1+**	50–172
Ribosome subunits				
Mitochondrial ribosomal protein	4v1aw	387	**3_1-** $(3_1-,3_1-,3_1-)$	226–289
Membrane proteins				
Calcium exchanger protein:				
NCX	5hwyA	300	**3_1+**	44–230
CAX	4kppA	395	**3_1+**	78–274

Novel knotted protein families are depicted in bold. PDB code includes the chain identifier the last letter. The first (bold) entry in the knot fingerprint column refers to the knot type of the whole chain including its handedness (+ for right handed, – for left handed). In brackets, we depict additional slipknots, which occur (and disappear) when certain subchains are considered (compare with the fingerprint matrix in Fig. 3). Knotted core positions are taken from the KnotProt [18] and renumbered in agreement with indices from PDB. Chain length indicated here is the length of the structurally resolved part of the protein and may be significantly shorter than the whole known sequence.

[a] Likely knotted, but 5fmhA is missing 15 amino acids in the center of the knot.

handed (–) or right handed (+) or achiral as described, for example, in Ref. [20].

A complete picture of knotting along the protein backbone chain is visual via the so-called fingerprint matrix introduced in Refs. [13,52]. This matrix represents the length of the knotted core, types, chirality and probability of forming knots in all subchains of a given protein backbone (see, e.g., Fig. 3). For example, the whole chain in protein with PDB code 4v1aw forms a trefoil knot. Embedded in the structure are, however, three additional trefoil knots (denoted in brackets in Table 1), which can be located if certain subchains are considered. Such chains are referred to as internal slipknots [49,53] and dissolve when the chain as a whole is taken into account. This happens, for example, when the ends of a slipknot are extended in a way that they are fed back into a loop of the knot, dissolving it in the process. So far, only knotted structures acting as enzymes and DNA-binding proteins were known. In the following, we also discuss novel knotted structures appearing in the ribosome and membrane proteins.

Topologically Knotted Mitochondrial Proteins

We identify for the first time two mitochondrial protein families, which both contain a left-handed trefoil knot.

In the following, we discuss in detail structures and folding pathways based on numerical simulations and argue that the proteins likely fold to their knotted native state inside the mitochondrion.

The apoptosis-inducing factor 1

The first new knotted fold is observed in the apoptosis-inducing factor 1 (Aifm1) PDB code: 5fmh,[§] which is a mitochondrial flavoprotein located in the inner membrane that, after cell death induction, is distributed to the nucleus to mediate chromatinolysis [54].

Folding

Based on our coarse-grained structure-based folding simulations [55,56], Aifm1 can be divided into four domains [including the knotted C-terminal domain (CTD)] that fold independently (proposed folding pathway is presented in Fig. 2). Interestingly, it was shown experimentally that Aifm1's import to the mitochondrion can occur only for a non-native state and the folding occurs after the transportation into the intermembrane space [57]. Aifm1 contains (near the N-terminal end) a short transmembrane motif which is responsible for anchoring the protein to the inner membrane [58].

Such immobilization of one of the ends of an unfolded protein reduces its conformational sampling and likely affects the folding. Moreover, our results, regarding the folding pathway of Aifm1, indicate that the CTD can fold into the knotted structure independently from the N-terminal part of the protein. This suggests that the whole folding

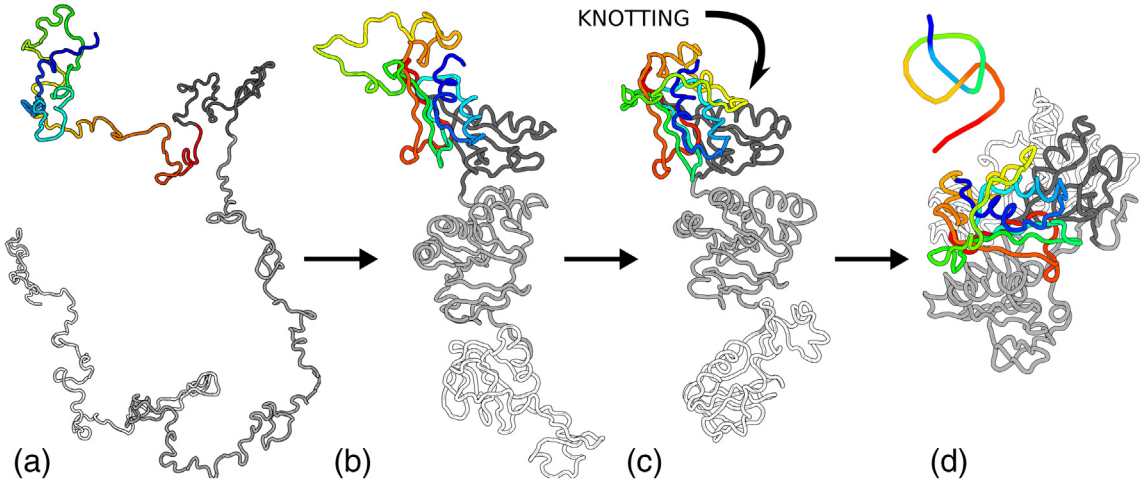

Fig. 2. Proposed folding pathway of a mitochondrial apoptosis-inducing factor (PDB code: 5fmh, chain a). Based on folding simulations, this protein can be divided into three trivial domains (colored white, medium gray, dark gray from N-terminus) and the C-terminal knotted one (colored red to blue from the N-terminus). All these domains fold independently, with the knotted one first forming a nucleation site consisting of a β-sheet and α-helices* with a freely following loop (b), which then covers this nucleation site, thus forming the knot (c). Domains then join together to form the native conformation (d). *Note that simulations were performed with a coarse-grained structure. Thus, here, we indicate that helix- and sheet-like structures were formed by regions of the protein natively involved in the formation of corresponding secondary structure elements.

process of this protein is performed in the intermembrane space of the mitochondrion, including knotting.

Importance of the knotted region

Aifm1 is released from the mitochondrion after a proteolytic cut (see Fig. 4 in SI), and this truncated form was shown to be apoptogenic [57]. Interestingly, the CTD domain, which is also knotted, plays a crucial role in this process [59–61] via DNA-binding motifs: K510 and K518 [62], and an extended proline-rich sequence (from amino acid 544 to 555 [63]), which both are part of the knotted core, specifically the loop covering the knot. Deletion or mutagenesis of the proline-rich binding motif disrupts interaction of Aifm1 with histone H2AX and stops the apoptosis [63]. Knots are often tied to the stabilization of otherwise flexible domains [34,64] (which is in agreement with the B-factor of the Aifm1 crystal structure). Because the important binding motifs can be found in the knotted CTD domain, its knot-induced stabilization could also facilitate the proper alignment of the functionally crucial elements.

It is also worth noting that Aifm1 as a flavoprotein binds to flavin derivatives FAD, although binding regions do not coincide with the knot (as shown in the SI Fig. 4). We believe that this departure from the usual active site-knot relationship may be due to Aifm1 being a signaling protein, and not an enzyme.

The human mitochondrial ribosome

The second knotted fold is found in two homologous proteins from the human mitochondrial ribosome type L37 and S30 (respectively, PDB codes: 4v1a, chain w and 5aj4, chain Bw). Mitochondrial ribosomes are responsible for the synthesis and insertion of membrane proteins that are critical for energy conversion and ATP production inside mitochondria [65].

The atomistic structures for both of these proteins came from *de novo* tracing, since the ribosome structure was obtained with cryo-electron microscopy and there were no homologous structures for either of these proteins in the PDB [65,66]. This is the first knotted protein predicted *de novo* (based on our review of the structures predicted during the CASP competition). Matrix representations of the knot reveal a complex topological fingerprint (see Fig. 3) for these proteins. Each protein forms a left-handed 3_1 knot as the whole, and moreover, three distinct regions in the fingerprint matrix represent subchains with (other) 3_1 knots.

The main 3_1 knot is rather shallow from the C-terminus, with only 5 excessive amino acids. To see whether the knot can untie due to thermal fluctuations, we have performed a few all-atom explicit solvent molecular dynamics simulations. Over the time of every simulation (the longest for 1 μs), the C-terminus remains stable (Fig. 5 in SI), which is an indication for the stability of this knotted topology, although some loops unfold spontaneously (Fig. 6 in SI).

Potential folding pathways of the human mitochondrial ribosomal protein

It is difficult to imagine how the protein with such a complex topological fingerprint (as can be seen in Fig. 3) could fold into a stable structure. Current theoretical results [12,36,38–40,43,67,68] indicate that proteins can self-tie in a single movement, which simplifies the folding landscape significantly [10]. This move corresponds to threading one of the protein termini across the native twisted loop [38]. In the case of deeply slip knotted proteins, two folding pathways are observed. In one, the loop is threaded across the knotted loop, giving rise to a slipknot, similarly as in the unfolding mechanism [49]. In another one, both the knot and the slipknot are formed almost simultaneously due to rotation of the loop by almost 360° [38]. This one move makes the protein structure nontrivial. In all of the studied examples, knotting is one of the latest steps during folding. These results suggest that the mitochondrial ribosomal protein with the complex topological fingerprint could fold just in two steps: a knot formation and then rotating (threading) one loop over/through another loop.

Based on our molecular dynamics folding simulations with a structure-based coarse-grained model of

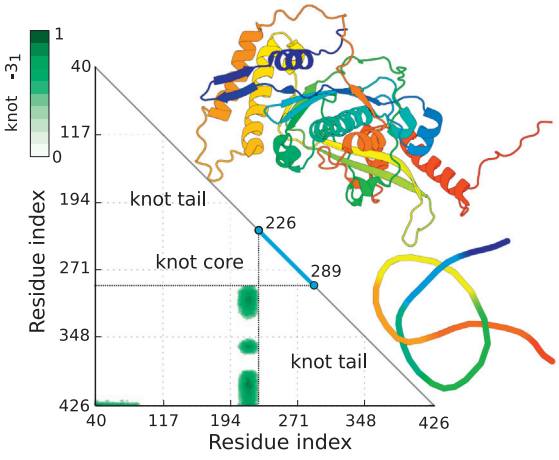

Fig. 3. Protein from the large subunit of the mammalian mitochondrial ribosome (PDB code: 4v1a, chain w). Cartoon representation of a protein, a matrix of protein knotting determined by KnotProt and as an inset, a reduced schematic representation of the protein backbone is shown. Colors change from red (N-terminus) to blue (C-terminus) to guide the eye. The matrix shows how clipping the C-terminus and/or N-terminus from the entire polypeptide chain (with a global trefoil knot) leads to the formation of subchains that form three internal 3_1 slipknots.

mitochondrial ribosomal protein (PDB code: 4v1a, chain w), we propose its potential folding pathway, as schematically drawn in Fig. 4.

As can be seen there, whether the final structure is correctly knotted depends on the position of the N-terminus. If the N-terminal end does not thread through at the beginning of the outer knot-forming loop formation, the protein ends up in a topological trap, where this terminus cannot be threaded to form the knot (see Fig. 4e). This configuration, however, leads to the native slipknot topology. On the other hand, if the N-terminus does thread (Fig. 4b), a native 3_1 knot is formed soon (Fig. 4c) by threading of the C-terminus (shorter knot tail). Thus, this intermediate configuration is quite stable. The ultimate step, that is threading of the last of the inner loops, turns out to be the rate-limiting step in the simulations. Consequently, it was only possible to fold the knotted structure in the simulation time frame once we have shortened the N-terminus by 34 amino acids (full statistics from simulations are available in Table 1 in SI). Based on the analysis of this data, we think that removing 34 amino acids (which are present only as excluded volume interactions in the structure-based model) from the freely moving N-terminus makes it less motile and leaves more space for the final loop threading.

We do not observe spontaneous knotting at the early stages of folding. However, probably due to complexity of the topological fingerprint of this protein, we observe a significant number of topological traps during folding. These traps decrease the probability of folding this protein in the simulations. (Types and frequency of topological traps are available in Table 2 in SI.)

Import to mitochondria

Interestingly, mitochondrial proteins from this study contain a transit peptide on the N-terminus and are therefore produced outside the mitochondrion and have to be transferred through the mitochondrial membranes. Only small natively folded proteins are eligible for such a transport, most of the proteins should be in an (at least partially) unfolded state [69]. Since the knotted mitochondrial proteins from this study are at least 400 amino acids long, they cannot be transported in their native state. However, two different methods of transportation for such proteins are feasible. In the first scenario, the protein is being transported in an unknotted state to the mitochondrion either cotranslationally or with aid of the chaperones to avoid aggregation (as was shown for other proteins [70,71]).

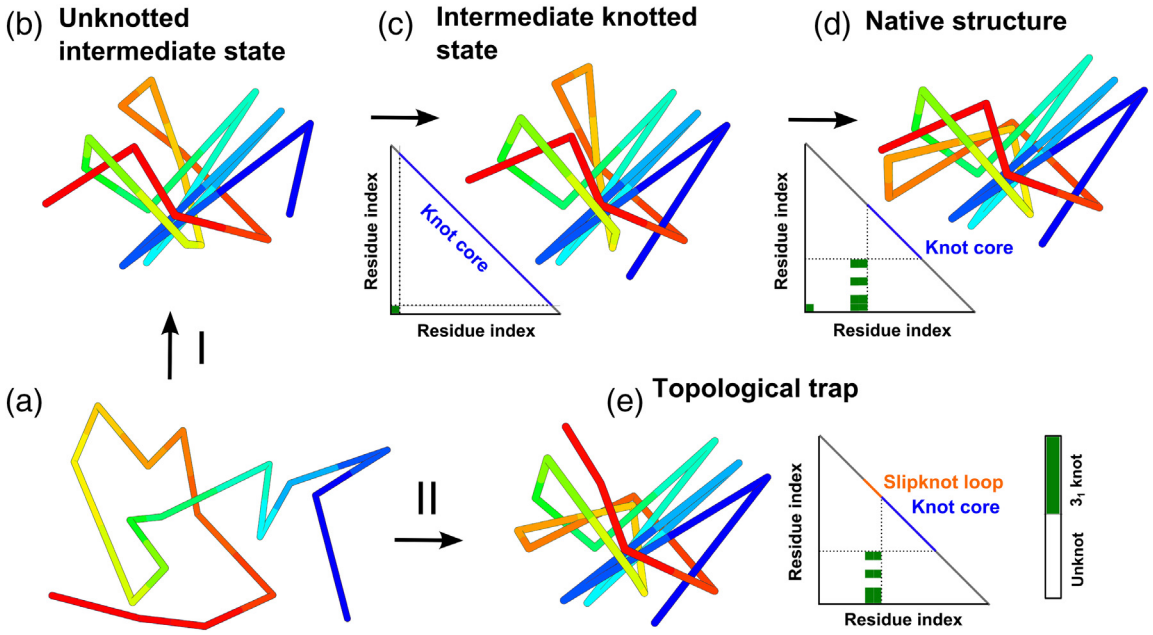

Fig. 4. Proposed folding pathways of mitochondrial knotted protein (PDB code: 4v1a, chain w), based on numerical simulations. Pathway I (which leads to the proper structure) starts by wrapping the outer loop around the N-terminus (B), then C-terminus is threaded, which forms the 3_1 knot (C), and finally another loop threads through the outer knot loop, which results in the expected complex topological fingerprint K $3_1 3_1 3_1 3_1$. (D) Pathway II, in which the N-terminus does not thread through the outer knot loop, results in a topological trap (E), S$3_1 3_1 3_1$. The letters K and S followed by knot type notations indicate whether the entire polypeptide chain of this protein forms a knot K or a slipknot S, respectively, and the knot types formed by the subchains of this protein [13]. Colors change from red (N-terminus) to blue (C-terminus) to guide the eye. Approximate knot fingerprint matrices for each non-trivial state are shown.

Our results suggest that the mitochondrial proteins from this study follow this pathway and are unknotted prior to entrance and folding in mitochondrion. Numerical simulations indicate that Aifm1 is a protein with a downhill folding pathway that can easily knot from the C-terminus, and in agreement with other studies [57], we suggest that folding occurs after its transportation to the mitochondrion.

The second approach to mitochondrial transport assumes that the folded knotted protein unfolds due to the encounter with the membrane. After reaching its location in the mitochondrion, it folds to the native structure. Movement of a protein chain through a membrane pore is mediated by an import motor, which applies force to unfold and pull the protein. In case of an already knotted protein, pulling would tighten the knot [72], thus enabling the passage [49,50]. Such import method is possible only if the membrane channels will not be occluded by a knot existing on the polypeptide chain. A tightly knotted protein chain has a radius of gyration of 7.2 ± 0.2 Å (3_1 knot [73]), whereas mitochondrial channels' diameters vary from 13 Å for TIM23 (translocase of the inner membrane [74]) to 22 Å for TOM (translocase of the outer membrane [75]). Moreover, knotted proteins can easily refold when starting from the unfolded state with the knot present [38]. All of the above suggests that this approach is in principle feasible, but unlikely at least for proteins analyzed here (due to their topological complexity).

The mitochondrial ribosomal protein has a rather complicated knotting pathway that requires precise structural configurations. If the protein was knotted (with all of the inner loops threaded) in the cytoplasm, it would have to unfold substantially to cross the membranes, since it seems that only the tightly knotted structure is eligible for transport. We therefore suspect that this protein is also unknotted during import in order to avoid the necessity to repeat the complicated folding pathway.

The Most Deeply Embedded Knot Is (Tp0642) from *T. pallidum*

An interesting novel fold with a trefoil knot appeared in the crystal structure of a not yet annotated protein from *T. pallidum* (Fig. 1). While its function has not been ascertained yet, it has been suggested [76] that it can contribute to the unusual cell envelope structure, which in turn can be important for the pathogenesis for this syphilis-causative bacterium.

There are no homologous structures deposited yet, and all of the known sequential homologs, which are also composed of the three domains found in this protein, belong to species from genera *Treponema*. Especially interesting from the perspective of knotting of this protein are the aforementioned three domains. The structure of this protein is highly modular [76], with the knot appearing in the middle domain. Our review of all knotted proteins has shown that this protein has the deepest knot (in depth notes on that can be found in SI), with tails consisting of at least 138 amino acids (Fig. 1). Currently, it is believed that knot formation is always tied to the last stages of protein folding [20,35,43]. However, in the case of this protein, knotting should occur rather early in the folding process, as it is unlikely to tie a knot after folding the two external (terminal) domains. Our extensive coarse-grained structure-based simulations indeed indicate that when terminal domains are formed first, the length of backbone between them is too short to tie the middle domain. More precisely, it is impossible to thread one of folded domains through the twisted loop of the middle domain, without unfolding significant part of folded domain. Hence, folding should start from one of the terminal domains. The middle domain is under constant tension, so no stable nucleation site can be formed. An almost perfectly folded protein with a slipknot topology in the middle domain is shown in Fig. 11 in SI.

Surprisingly, when the starting structure contains a large loop with one of the termini close to threading, it is possible for the protein to fold correctly. As a first step, a shallow knot is formed (through aforementioned threading). Then, the knot slides toward its native position in the center of the protein. At this point, all the domains fold independently in random order. A full pathway is presented in Fig. 5. Note that although in Fig. 5 the folding occurs from the C-terminus, a folding from loosely formed very shallow knot at the N-terminus would be just as likely from a computational point of view. For simulations starting from unfolded conformation containing a knot around the native position, the probability of attaining the native structures is around 80% (see SI Table 2).

The geometry and depth of the knot of Tp0642 can be compared to knotted proteins YibK and YbeA fused to ThiS domain (92 residues long) on each terminus [77]. It was suggested [77] that YibK and YbeA can form the correct topology even when encumbered with an additional ThiS domain on each terminus. As the ThiS domain is expected to fold as soon as possible, this would suggest the necessity of threading a fully folded domain through the loop of the forming knot. The simulation presented in this paper however suggested that such knotting at least for Tp0642 is unlikely. Instead, a recent publication suggests that the key to correct folding of such proteins may lie in the interaction with the translating ribosome, and the folding process of the knot itself may be co-translational [44].

Knotted Membrane Proteins

An exciting discovery is the first occurrence of a deep trefoil knot in a membrane protein, namely, H^+

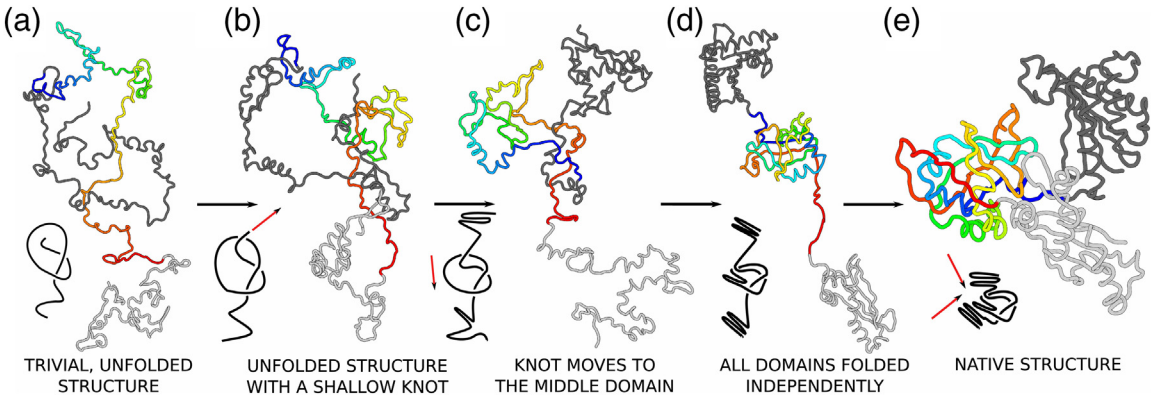

Fig. 5. Example of the simulated folding pathway of a knotted protein (PDB code: 5jir, chain A) composed of three domains. The knot is located in the middle domain (colored red to blue from N-terminus); N-terminal domain shown in light gray, C-terminal in dark gray. Starting from an unfolded state with a loop, a shallow knot is formed, which then slides to its native position in the middle domain. All domains fold independently, and are then joined into the final, native-like conformation.

[78] (Fig. 6a) and Na^+ [80] (Fig. 6b) calcium antiporters.[||]

Ca^{2+}/cation (CaCA) antiporters are transmembrane proteins, with most members of this family regulating the calcium level in the cytosol, through enabling the exchange of Ca^{2+} and various cations across membranes. CaCA proteins are highly diverse in terms of both function (with examples of H^+/Mn^{2+} and H^+/Mg^{2+} also classified there [82]), and sequence—with the proteins studied here barely sharing more than 10% of their sequence [83].

Even with aforementioned high variability of the sequence, CaCA proteins can be classified into five major groups: apart from H^+/Ca^{2+} (CAX) and Na^+/Ca^{2+} (NCX) exchangers, there are also K^+-dependent Na^+/Ca^{2+} exchangers (NCKX), cation/Ca^{2+} exchangers (CCX) (which exchange cations other than Na^+, H^+ and K^+) and the bacterial homologous gene (YrbG). In addition, two more subfamilies can be isolated: NCX-related H^+/Mg^{2+} (MHX) exchangers, and, possibly due to convergent evolution, Na^+/Ca^{2+} (NCL) exchangers [80], which nevertheless are more closely related to CAX.

Differences between CAX- and NCX-related families remain even when the internal domains are analyzed separately—that is when both repeats of the α-helical domain from each protein are aligned as a separate entity—both domains from CAX group closely together. The only group that breaks the CAX/NCX barrier are the NCL exchangers, for which—appropriately to their NCX-like function and CAX relatedness—the N-terminal domain gets grouped with the N-termini from CAX, and the C-terminal one with CTDs of NCX and similar proteins (full results

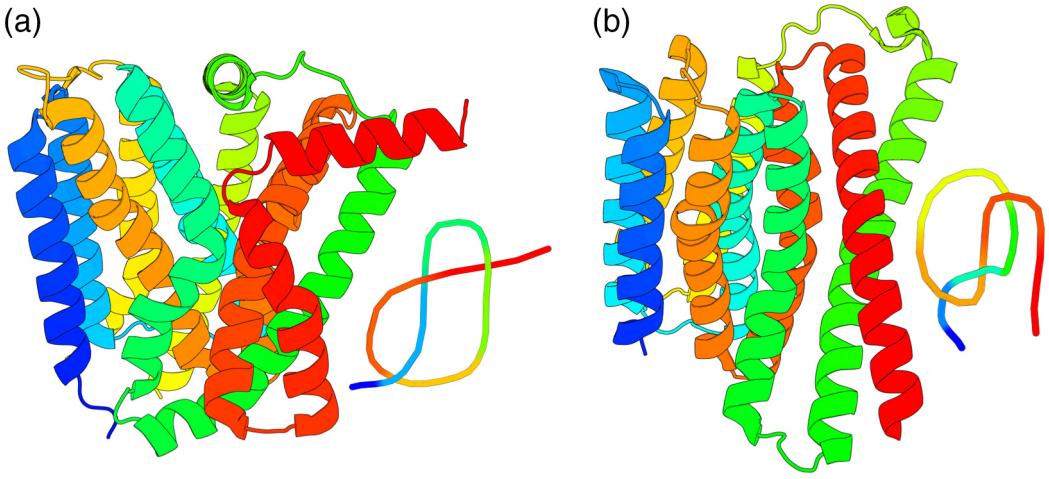

Fig. 6. Structure of knotted transmembrane proteins—Ca^{2+}/cation (CaCA) antiporters. CAX (a) [79] (PDB code: 4k1c) and NCX (b) [80] (PDB code: 5hwy) antiporter. As an inset, a reduced schematic representation of the protein backbone is shown. Colors change from red (N-terminus) to blue (C-terminus) to guide the eye.

including phylogenetic tree in SI). All this, coupled with disappearance of some CaCA proteins in selected lineages, occludes the phylogenetic relationships within this family.

There are currently four known CaCA exchanger structures, three H^+ antiporters (PDB codes: 4k1c, 4kjr, 4kpp) and one Na^+ antiporter (PDB code: 5hwy), all of them exhibiting a deep right-handed trefoil knot in the preserved core region—that is 10 transmembrane helices present in all members of the family [82]. Figure 7 shows a phylogenetic tree of the CaCA superfamily constructed (following Ref. [82]) from an alignment of the conserved core domains from antiporters from rice and *Arabidopsis* and selected other species (as per Ref. [82]). Highlighted are the locations of the known structures.

One of the known structures is missing from the tree—the H^+/Ca^{2+} from *Archaeoglobus fungidus* (PDB code: 4kpp), as its placement on the tree may be somewhat controversial (full tree is shown in Fig. 12 in SI). While it has been shown experimentally [78] to functionally belong to the CAX subfamily, sequential alignment, and the following phylogenetical tree, cluster it to the NCKX group. Based on bioinformatical analysis and study of the literature, we believe this placement to be correct and that the similarity in function to the CAX proteins is due to a convergent evolution. *A. fungidus* is an archaeal species, and it has been suggested [84] that Archaea/Bacteria split predates the emergence of CAX antiporters. In addition, there are no confirmed CAX proteins in UniProt belonging to any archaeal species. Similarly, database search using the

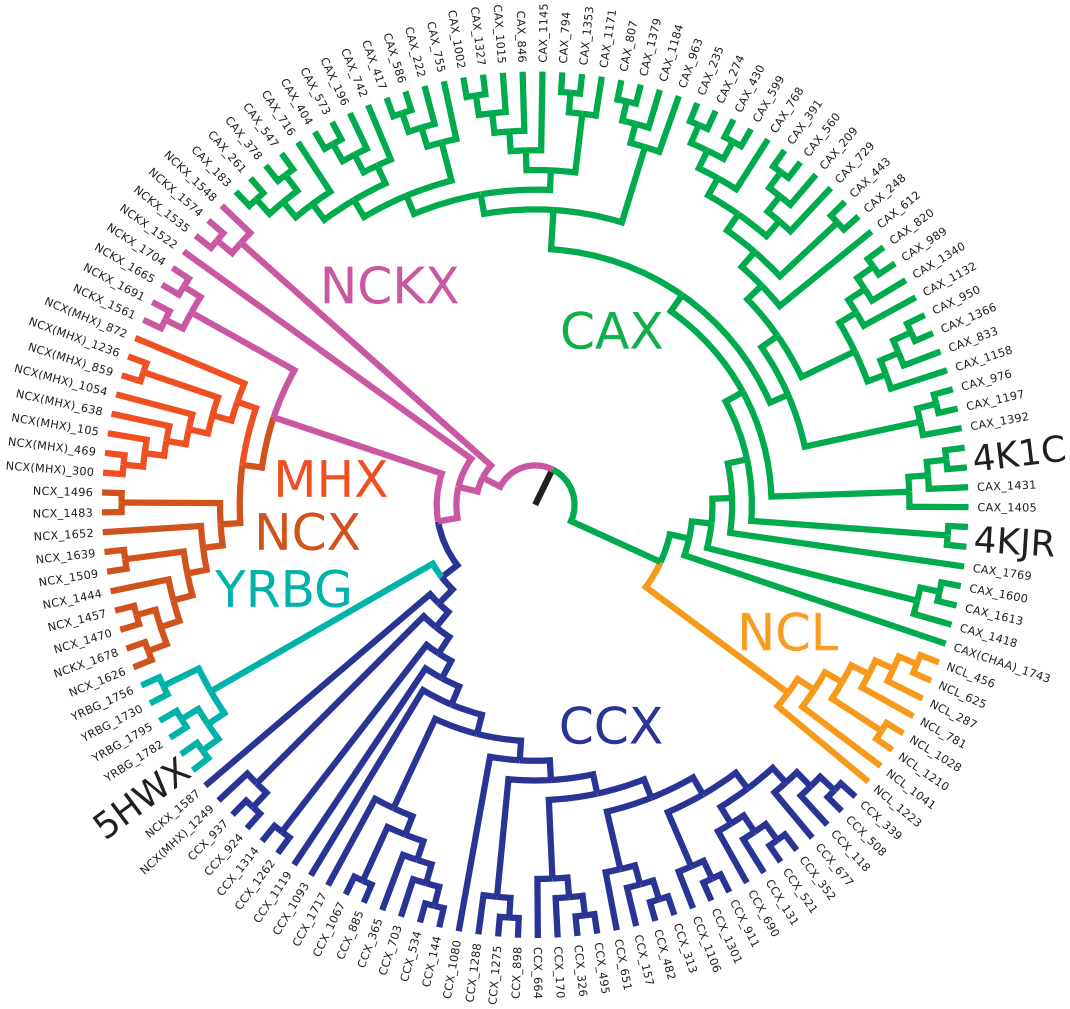

Fig. 7. Phylogenetic tree for Ca^{2+} antiporters. Subfamilies are shown with different colors. Known structures are indicated amongst tree leaves by a larger font. As all of the known structures contain a topological knot and they are spread throughout the tree, a reasonable conclusion, to be confirmed or rejected as new structures are found, is that the common ancestor of the family, and the root of the tree, was also knotted, and as such, all other members of this family will contain a knot.

sequence of the studied protein yields almost exclusively sodium-calcium exchangers.

Even with these phylogenetical complications, one thing appears to remain constant throughout this group of cation exchangers—aforementioned conserved core of transmembrane helices. There are noticeable distinctions in helix shapes in the functional canal between H^+ and Na^+ antiporters, which further shows that these two groups should be classified as two separate main branches of the phylogenetical tree. Since the general shape of the conserved core remains the same (with RMSD of 4 Å over 270 residues), currently known structures can be used to reliably model other CaCA exchangers (including, in Ref. [82], an MHX protein from *Arabidopsis thaliana*, other examples described in SI). Consequently, since the knotted topology is conserved between the two divergent subgroups, this similarity is an indication that the formation of the knotted structure occurred early in evolutionary terms and can be traced back to the common ancestor of all CaCA antiporters and may thus be present in all members of the superfamily (assuming that it was not lost again in some of the subbranches). This hypothesis will of course be confirmed or rejected as soon as more structures of the CaCA superfamily are determined.

A Small Knotted Protein with a New Fold

DndE belongs to the class of DNA-binding proteins and is involved in the phosphorothioation of DNA. In this physiological process, which is widespread in bacterial genomes, the canonical structure of DNA is modified by incorporating sulfur into the backbone as part of the restriction modification system. DndE folds into a tetramer conformation and is believed to bind nicked DNA [85]. The backbone of this protein has a left-handed trefoil knot (PDB code: 4lrv [85], Fig. 15 in SI). The crystallized form of DndE consists of only 110 amino acids and is one of the shortest known knotted proteins.

Structure-wise, DndE is very similar to the trefoil-knotted members of the RHH superfamily (MJO366, PDB code: 2efv; VirC2, PDB code: 2rh3) as classified by SCOPe [22]. In particular, all three knotted proteins DndE, MJO366 and VirC2 share a similarity with an unknotted homodimeric transcription regulator SvtR (PDB code: 2kel)—also with the RHH motif. Further analysis of available sequential and biological data, however, shows some key differences that highlight the novelty of DndE fold.

One major difference between typical RHH proteins and DndE is the structure–function correspondence, that is, the binding mode. Double-stranded β-sheet created by two RHH motifs fits into the major groove of the DNA molecule. On the other hand, it was shown [85] that DndE forms a tetramer, which binds the nucleic acid through a positively charged cleft made up from α-helices from all subunits (Fig. 15 in SI).

This relocation of the function along the structure corresponds to changes in sequence—sequence alignment shows that residues, corresponding to the binding ribbons from RHH motif, appear in an α-helix in DndE (Fig. 16 in SI). In particular, the two β-strands that are responsible for DNA binding in RHH proteins are all but absent in the DndE structure (Fig. 15 in SI).

Clear sequential and structural similarities between DndE and RHH proteins, when combined with aforementioned functional differences, may suggest that proteins here described exemplify consecutive steps in the evolution of a DNA-binding protein. Two dimerizing RHH subunits (SvtR), which by domain duplication form a knotted monomer (VirC2), then, through oligomerization, gain a new, as of yet unknown [85] function.

Artificially Knotted Protein

Single-chain TFIIA (PDB code: 5m4s) is an artificially knotted version of the human TFIIA protein, which in its native state consists of three separate polypeptide chains (α, β, γ). The single-chain structure was obtained by connecting the chains using two linkers. The way the linkers connect the chains results in the formation of the trefoil knot. Moreover, the single-chain structure perfectly reconstructs the mutual arrangement of the chains in the multi-chain TFIIA, implicating that the presence of the knot did not distort the conformation of the protein or its function [86]. In fact, the formation of the single chain enforced the close contact between the motifs and possibly facilitated the correct folding. Note that this protein is only the second artificially created protein, which contains a knot. While in Ref. [23] the protein was designed to contain a knot, the presence of the knot in TFIIA was not noted in the original publication [86].

Conclusions

In this paper, we provide a comprehensive overview of knotted protein structures and discuss in detail new additions by combining topological analysis, numerical simulations and bioinformatics. These novel folds include two mitochondrial proteins (an apoptosis-inducing factor and a ribosomal subunit), both of which are expressed in the cytoplasm and must therefore cross a membrane to get to their designated location (likely in an unknotted state as indicated by our simulations). Based on structure analysis, we suggest that the knot topology in an apoptosis-inducing factor on one hand stabilizes flexible CTD domain and on the other hand aligns the crucial DNA-binding elements.

We also present the most deeply knotted structure known to date in *T. pallidum* and suggest a folding pathway in which a loose knot forms at a terminus and slides across the protein to its native location. Another exciting discovery is the occurrence of a deep trefoil knot in a membrane protein, namely, the cation/calcium antiporters, which, based on our phylogenetic analysis, is likely present in all members of this family. Finally, we present arguments for a DNA-binding protein DndE to be considered as a new type of knotted fold and show new artificially knotted protein.

In summary, we believe that such an interdisciplinary approach is required to advance our understanding of these peculiar structures to understand the role of entanglement for hosting organisms. Knotted proteins appear in all kingdoms of life and, as demonstrated in this paper, are not limited to certain functional classes. Topological structures hold great promise not only in synthetic chemistry, and a better understanding of topology in proteins may initiate similar trends in bio-engineering.

Methods

Knot detection

To determine whether a protein is knotted and possesses a complex topological fingerprint, the knot detection from the KnotProt [18] has been used.[¶] Several hundreds of random closures are chosen in a way that the end points are connected to two randomly chosen points on a sphere encompassing the whole protein. The two new end points are connected on the surface of the sphere. The closed backbone is then reduced using a KMT algorithm [87], and the Alexander polynomial [88] is computed. The most frequently occurring knot type is the proteins dominant knot type [89]. To determine the chirality of the knot, the HOMFLY polynomial [90] is calculated.

To determine the size of a knot in a protein with N C_α-atoms, all subchains $S(j,k)$ are analyzed with $1 \leq j < k \leq N$. The size of the knot is then given by the smallest distance of k and j for which the knot still exists and is presented via the matrix presentations of protein knotting [13]. KnotProt renumbers the residues so that the first amino acid in the crystal structure has number 1, but to give a clearer picture, all plots within this paper were annotated with actual, sequence-based residue numbers.

Molecular dynamics simulations

All molecular dynamics simulations were performed in GROMACS 5.0.2 package (details are summarized in the SI). To study thermal stability of the protein, we used all-atom, explicit-solvent simulation with CHARMM force field. Folding kinetics were studied with a structure-based Cα model [91] with various strengths of the attractive potential implemented in the SMOG server [55]. Details regarding each simulation are presented in the SI, including the changes we made in the potential. Other simulation parameters were kept default as in the SMOG server. The simulations were conducted using leap-frog stochastic dynamics integrator and time step equal to 0.0005τ. The cutoff distance for the short-range neighbor list was set to 2 nm, as well as the cutoff for both the electrostatic and van der Waals interactions. The folding pathways were studied based on unknotted and unfolded conformations obtained in unfolding simulations.

Phylogenetic analysis

Phylogenetic analysis was done using ClustalO to create a multiple sequence alignment, from which the tree was found with ParsimonyTreeConstructor from BioPython (only columns populated in at least 60% of sequences were used). Additional information on the method, as well as other phylogenetic analyses for this data, can be found in SI.

Acknowledgments

Scientific work financed from the budget for science in the years 2016–2019 (No. 0003/ID3/2016/64 Ideas Plus), National Science Centre (No. 2012/07/E/NZ1/01900) and EMBO (No. 2057 Installation Grant) to J.I.S. B.T would like to acknowledge the financial support from graduate school MAINZ and is a recipient of the fellowship through the Excellence Initiative (DFG/GSC 266).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2018.10.012.

Received 22 July 2018;
Received in revised form 17 October 2018;
Accepted 17 October 2018
Available online 1 November 2018

Keywords:
knotted proteins;
folding

†Private communication with Sophie Jackson.

‡Table 1 does not include structures that we believe have been assigned incorrectly—see SI.

§As this structure is not complete (15 amino acids from the knot forming loop were not crystallized), the topology cannot be determined with certainty. Regardless, the most natural position of this loop (around the protein, as we believe it could not be threaded through the structure), coupled with aforementioned extended C-terminus, indicates the presence of a shallow 3_1 knot (Fig. 3 in SI).

‖The occurrence of a knot in the CAX antiporter (PDB code: 4kpp) and the DNA binding protein (PDB code: 4lrv) was first reported in a presentation by one of us (P.V.) at the Banff workshop on entanglement in biology in November 2013. These knots were discovered independently by the group of J.I.S. and listed shortly afterward online on their webserver [18]. Likewise, the NCX exchanger was also listed online directly after deposition. The membrane knots were mentioned in a review by Lim and Jackson [81] with reference to the webserver. A recent review by Jackson *et al.* [35] also lists (PDB code: 4lrv).

¶The researchers from Mainz have used a different closure method and algorithm to determine knots and their sizes as explained in Ref. [11]. All data in the paper refer to the analysis with KnotProt.

Abbreviations used:

Aifm1, apoptosis-inducing factor 1; CTD, C-terminal domain; CaCA, Ca^{2+}/cation; CAX, H^+/Ca^{2+} exchanger; NCX, Na^+/Ca^{2+} exchanger; RHH, ribbon–helix–helix.

References

[1] R.S. Forgan, J.-P. Sauvage, J.F. Stoddart, Chemical topology: complex molecular knots, links, and entanglements, Chem. Rev. 111 (9) (2011) 5434–5464.

[2] C. Dietrich-Buchecker, J.-P. Sauvage, J. Kintzinger, A new family of molecules—metallo-catenanes, Tetrahedron Lett. 24 (46) (1983) 5095–5098.

[3] C. Dietrich-Buchecker, J.-P. Sauvage, A synthetic molecular trefoil knot, Angew. Chem. Int. Ed. Engl. 28 (2) (1989) 189–192.

[4] H. MacGregor, M. Vlad, Interlocking and knotting of ring nucleoli in amphibian oocytes, Chromosoma 39 (2) (1972) 205.

[5] L. Liu, L. Perkocha, R. Calendar, J. Wang, Knotted dna from bacteriophage capsids, Proc. Natl. Acad. Sci. U. S. A. 78 (9) (1981) 5498–5502.

[6] J. Arsuaga, M. Vazquez, S. Trigueros, D. Sumners, J. Roca, Knotting probability of dna molecules confined in restricted volumes: DNA knotting in phage capsids, Proc. Natl. Acad. Sci. U. S. A. 99 (8) (2002) 5373–5377.

[7] J.T. Siebert, A.N. Kivel, L.P. Atkinson, T.J. Stevens, E.D. Laue, P. Virnau, Are there knots in chromosomes? Polymers 9 (8) (2017) 317.

[8] M.L. Mansfield, Are there knots in proteins? Nat. Struct. Mol. Biol. 1 (4) (1994) 213–214.

[9] F. Takusagawa, S. Kamitori, A real knot in protein, J. Am. Chem. Soc. 118 (37) (1996) 8945–8946.

[10] W.R. Taylor, A deeply knotted protein structure and how it might fold, Nature 406 (6798) (2000) 916–919.

[11] P. Virnau, L.A. Mirny, M. Kardar, Intricate knots in proteins: function and evolution, PLoS Comput. Biol. 2 (9) (2006), e122.

[12] D. Bölinger, J.I. Sułkowska, H.-P. Hsu, L.A. Mirny, M. Kardar, J.N. Onuchic, P. Virnau, A stevedore's protein knot, PLoS Comput. Biol. 6 (4) (2010), e1000731.

[13] J.I. Sułkowska, E.J. Rawdon, K.C. Millett, J.N. Onuchic, A. Stasiak, Conservation of complex knotting and slipknotting patterns in proteins, Proc. Natl. Acad. Sci. 109 (26) (2012) E1715–E1723.

[14] J.S. Richardson, Beta-sheet topology and the relatedness of proteins, Nature 268 (1977) 495–500.

[15] R.C. Lua, A.Y. Grosberg, Statistics of knots, geometry of conformations, and evolution of proteins, PLoS Comput. Biol. 2 (5) (2006), e45.

[16] G. Kolesov, P. Virnau, M. Kardar, L.A. Mirny, Protein knot server: detection of knots in protein structures, Nucleic Acids Res. 35 (suppl_2) (2007) W425–W428.

[17] Y.-L. Lai, S.-C. Yen, S.-H. Yu, J.-K. Hwang, pknot: the protein knot web server, Nucleic Acids Res. 35 (suppl – 2) (2007) W420–W424.

[18] M. Jamroz, W. Niemyska, E.J. Rawdon, A. Stasiak, K.C. Millett, P. Sułkowski, J.I. Sulkowska, Knotprot: a database of proteins with knots and slipknots, Nucleic Acids Res. 43 (D1) (2014) D306–D314.

[19] E.J. Rawdon, K.C. Millett, J.I. Sułkowska, A. Stasiak, Knot localization in proteins, Biochem. Soc. Trans. 41 (2) (2013) 538–541.

[20] P. Virnau, A. Mallam, S. Jackson, Structures and folding pathways of topologically knotted proteins, J. Phys. Condens. Matter 23 (3) (2010), 033101.

[21] P. Dabrowski-Tumanski, J.I. Sulkowska, To tie or not to tie? That is the question, Polymers 9 (9) (2017) 454.

[22] N.K. Fox, S.E. Brenner, J.-M. Chandonia, SCOPe: structural classification of proteins? Extended, integrating SCOP and ASTRAL data and classification of new structures, Nucleic Acids Res. 42 (D1) (2013) D304–D309.

[23] N.P. King, A.W. Jacobitz, M.R. Sawaya, L. Goldschmidt, T.O. Yeates, Structure and folding of a designed knotted protein, Proc. Natl. Acad. Sci. 107 (48) (2010) 20732–20737.

[24] A.L. Mallam, J.M. Rogers, S.E. Jackson, Experimental detection of knotted conformations in denatured proteins, Proc. Natl. Acad. Sci. 107 (18) (2010) 8189–8194.

[25] A.L. Mallam, S.E. Jackson, Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins, Nat. Chem. Biol. 8 (2) (2012) 147–153.

[26] F. Ziegler, N.C. Lim, S.S. Mandal, B. Pelz, W.-P. Ng, M. Schlierf, S.E. Jackson, M. Rief, Knotting and unknotting of a protein in single molecule experiments, Proc. Natl. Acad. Sci. 113 (27) (2016) 7533–7538 (201600614).

[27] I. Wang, S.-Y. Chen, S.-T.D. Hsu, Unraveling the folding mechanism of the smallest knotted protein, MJ0366, J. Phys. Chem. B 119 (12) (2015) 4359–4370.

[28] S.-C. Lou, S. Wetzel, H. Zhang, E.W. Crone, Y.-T. Lee, S.E. Jackson, S.-T.D. Hsu, The knotted protein UCH-l1 exhibits partially unfolded forms under native conditions that share common structural features with its kinetic folding intermediates, J. Mol. Biol. 428 (11) (2016) 2507–2520.

[29] T. Bornschlögl, D.M. Anstrom, E. Mey, J. Dzubiella, M. Rief, K.T. Forest, Tightening the knot in phytochrome by single-molecule atomic force microscopy, Biophys. J. 96 (4) (2009) 1508–1514.

[30] I. Wang, S.-Y. Chen, S.-T.D. Hsu, Folding analysis of the most complex stevedores protein knot, Sci. Rep. 6 (2016) 31514.

[31] Á. San Martín, P. Rodriguez-Aliaga, J.A. Molina, A. Martin, C. Bustamante, M. Baez, Knots can impair protein degradation by ATP-dependent proteases, Proc. Natl. Acad. Sci. 114 (37) (2017) 9864–9869.
[32] M.K. Sriramoju, Y. Chen, Y.-T.C. Lee, S.-T.D. Hsu, Topologically knotted deubiquitinases exhibit unprecedented mechanostability to withstand the proteolysis by an AAA+ protease, Sci. Rep. 8 (1) (2018) 7076.
[33] A. Bustamante, J. Sotelo-Campos, D.G. Guerra, M. Floor, C.A. Wilson, C. Bustamante, M. Báez, The energy cost of polypeptide knot formation and its folding consequences, Nat. Commun. 8 (1) (2017) 1581.
[34] J.I. Sułkowska, P. Sułkowski, P. Szymczak, M. Cieplak, Stabilizing effect of knots on proteins, Proc. Natl. Acad. Sci. 105 (50) (2008) 19714–19719.
[35] S.E. Jackson, A. Suma, C. Micheletti, How to fold intricately: using theory and experiments to unravel the properties of knotted proteins, Curr. Opin. Struct. Biol. 42 (2017) 6–14.
[36] J.I. Sułkowska, J.K. Noel, J.N. Onuchic, Energy landscape of knotted protein folding, Proc. Natl. Acad. Sci. 109 (44) (2012) 17783–17788.
[37] M. Chwastyk, M. Cieplak, Cotranslational folding of deeply knotted proteins, J. Phys. Condens. Matter 27 (35) (2015) 354105.
[38] J.I. Sułkowska, P. Sułkowski, J. Onuchic, Dodging the crisis of folding proteins with knots, Proc. Natl. Acad. Sci. 106 (9) (2009) 3119–3124.
[39] P.F. Faísca, R.D. Travasso, T. Charters, A. Nunes, M. Cieplak, The folding of knotted proteins: insights from lattice simulations, Phys. Biol. 7 (1) (2010), 016009.
[40] S.A. Beccara, T. Škrbić, R. Covino, C. Micheletti, P. Faccioli, Folding pathways of a knotted protein with a realistic atomistic force field, PLoS Comput. Biol. 9 (3) (2013), e1003002.
[41] Y. Zhao, P. Dabrowski-Tumanski, S. Niewieczerzal, J.I. Sulkowska, The exclusive effects of chaperonin on the behavior of proteins with 52 knot, PLoS Comput. Biol. 14 (3) (2018), e1005970.
[42] M. Sanders, H. Findlay, P. Booth, Lipid bilayer composition modulates the unfolding free energy of a knotted *α*-helical membrane protein, Proc. Natl. Acad. Sci. 115 (8) (2018) E1799–E1808 (201714668).
[43] J.I. Sułkowska, J.K. Noel, C.A. Ramírez-Sarmiento, E.J. Rawdon, K.C. Millett, J.N. Onuchic, Knotting pathways in proteins, Biochem. Soc. Trans. 41 (2) (2013) 523–527.
[44] P. Dabrowski-Tumanski, M. Piejko, S. Niewieczerzal, A. Stasiak, J.I. Sulkowska, Protein knotting by active threading of nascent polypeptide chain exiting from the ribosome exit channel, J. Phys. Chem. B. (2018).
[45] R. Potestio, C. Micheletti, H. Orland, Knotted vs. unknotted proteins: evidence of knot-promoting loops, PLoS Comput. Biol. 6 (7) (2010), e1000864.
[46] D. Shi, H. Morizono, X. Yu, L. Roth, L. Caldovic, N.M. Allewell, M.H. Malamy, M. Tuchman, Crystal structure of *n*-acetylornithine transcarbamylase from xanthomonas campestris a novel enzyme in a new arginine biosynthetic pathway found in several eubacteria, J. Biol. Chem. 280 (15) (2005) 14366–14369.
[47] K.L. Tkaczuk, S. Dunin-Horkawicz, E. Purta, J.M. Bujnicki, Structural and evolutionary bioinformatics of the spout superfamily of methyltransferases, BMC Bioinf. 8 (1) (2007) 1.
[48] T. Christian, R. Sakaguchi, A.P. Perlinska, G. Lahoud, T. Ito, E.A. Taylor, S. Yokoyama, J.I. Sulkowska, Y.-M. Hou, Methyl transfer by substrate signaling from a knotted protein fold, Nat. Struct. Mol. Biol. 23 (10) (2016) 941–948.
[49] J.I. Sułkowska, P. Sułkowski, J.N. Onuchic, Jamming proteins with slipknots and their free energy landscape, Phys. Rev. Lett. 103 (26) (2009) 268103.
[50] P. Szymczak, Tight knots in proteins: can they block the mitochondrial pores? Biochem Soc Trans 4 (2013) 620–640.
[51] M. Wojciechowski, À. Gómez-Sicilia, M. Carrión-Vázquez, M. Cieplak, Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins, Mol. BioSyst. 12 (9) (2016) 2700–2712.
[52] N.P. King, E.O. Yeates, T.O. Yeates, Identification of rare slipknots in proteins and their implications for stability and folding, J. Mol. Biol. 373 (1) (2007) 153–166.
[53] T.O. Yeates, T.S. Norcross, N.P. King, Knotted and topologically complex proteins as models for studying folding and stability, Curr. Opin. Chem. Biol. 11 (6) (2007) 595–603.
[54] P. Ferreira, R. Villanueva, M. Martínez-Jlvez, B. Herguedas, C. Marcuello, P. Fernandez-Silva, L. Cabon, J.A. Hermoso, A. Lostao, S.A. Susin, et al., Structural insights into the coenzyme mediated monomer–dimer transition of the pro-apoptotic apoptosis inducing factor, Biochemistry 53 (25) (2014) 4204–4215.
[55] J.K. Noel, M. Levi, M. Raghunathan, H. Lammert, R.L. Hayes, J.N. Onuchic, P.C. Whitford, Smog 2: a versatile software package for generating structure-based models, PLoS Comput. Biol. 12 (3) (2016), e1004794.
[56] J.I. Sułkowska, M. Cieplak, Selection of optimal variants of gō-like models of proteins through studies of stretching, Biophys. J. 95 (7) (2008) 3174–3191.
[57] S.A. Susin, H.K. Lorenzo, N. Zamzami, I. Marzo, B.E. Snow, G. M. Brothers, J. Mangion, E. Jacotot, P. Costantini, M. Loeffler, et al., Molecular characterization of mitochondrial apoptosis-inducing factor, Nature 397 (6718) (1999) 441–446.
[58] H. Otera, S. Ohsakaya, Z.-I. Nagaura, N. Ishihara, K. Mihara, Export of mitochondrial aif in response to proapoptotic stimuli depends on processing at the intermembrane space, EMBO J. 24 (7) (2005) 1375–1386.
[59] C. Yu, L. Jia, Y. Meng, J. Zhao, Y. Zhang, X. Qiu, Y. Xu, W. Wen, L. Yao, D. Fan, et al., Selective proapoptotic activity of a secreted recombinant antibody/AIF fusion protein in carcinomas overexpressing HER2, Gene Ther. 13 (4) (2006) 313–320.
[60] C. Delettre, V.J. Yuste, R.S. Moubarak, M. Bras, J.-C. Lesbordes-Brion, S. Petres, J. Bellalou, S.A. Susin, AIFSH, a novel apoptosis-inducing factor (AIF) pro-apoptotic isoform with potential pathological relevance in human cancer, J. Biol. Chem. 281 (10) (2006) 6413–6427.
[61] C. Delettre, V.J. Yuste, R.S. Moubarak, M. Bras, N. Robert, S.A. Susin, Identification and characterization of AIFSH2, a mitochondrial apoptosis-inducing factor (AIF) isoform with nadh oxidase activity, J. Biol. Chem. 281 (27) (2006) 18507–18518.
[62] H. Ye, C. Cande, N.C. Stephanou, S. Jiang, S. Gurbuxani, N. Larochette, E. Daugas, C. Garrido, G. Kroemer, H. Wu, Dna binding is required for the apoptogenic action of apoptosis inducing factor, Nat. Struct. Mol. Biol. 9 (9) (2002) 680–684.
[63] C. Artus, H. Boujrad, A. Bouharrour, M.-N. Brunelle, S. Hoos, V.J. Yuste, P. Lenormand, J.-C. Rousselle, A. Namane, P. England, et al., AIF promotes chromatinolysis and caspase-independent programmed necrosis by interacting with histone H2AX, EMBO J. 29 (9) (2010) 1585–1599.
[64] P. Dabrowski-Tumanski, A. Stasiak, J.I. Sulkowska, In search of functional advantages of knots in proteins, PLoS One 11 (11) (2016), e0165986.

[65] B.J. Greber, D. Boehringer, M. Leibundgut, P. Bieri, A. Leitner, N. Schmitz, R. Aebersold, N. Ban, The complete structure of the large subunit of the mammalian mitochondrial ribosome, Nature 515 (7526) (2014) 283–286.
[66] A. Brown, A. Amunts, X.-c. Bai, Y. Sugimoto, P.C. Edwards, G. Murshudov, S.H. Scheres, V. Ramakrishnan, Structure of the large ribosomal subunit from human mitochondria, Science 346 (6210) (2014) 718–722.
[67] J.K. Noel, J.N. Onuchic, J.I. Sulkowska, Knotting a protein in explicit solvent, J. Phys. Chem. Lett. 4 (21) (2013) 3570–3573.
[68] S. Niewieczerzal, J.I. Sulkowska, Knotting and unknotting proteins in the chaperonin cage: effects of the excluded volume, PLoS One 12 (5) (2017), e0176744.
[69] J. Rassow, F.-U. Hartl, B. Guiard, N. Pfanner, W. Neupert, Polypeptides traverse the mitochondrial envelope in an extended state, FEBS Lett. 275 (1–2) (1990) 190–194.
[70] J.C. Young, N.J. Hoogenraad, F.U. Hartl, Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70, Cell 112 (1) (2003) 41–50.
[71] C. Knox, E. Sass, W. Neupert, O. Pines, Import into mitochondria, folding and retrograde movement of fumarase in yeast, J. Biol. Chem. 273 (40) (1998) 25587–25593.
[72] J.I. Sułkowska, P. Sułkowski, P. Szymczak, M. Cieplak, Tightening of knots in proteins, Phys. Rev. Lett. 100 (5) (2008), 058106.
[73] J. Dzubiella, Sequence-specific size, structure, and stability of tight protein knots, Biophys. J. 96 (3) (2009) 831–839.
[74] K.N. Truscott, P. Kovermann, A. Geissler, A. Merlin, M. Meijer, A.J. Driessen, J. Rassow, N. Pfanner, R. Wagner, A presequence-and voltage-sensitive channel of the mitochondrial preprotein translocase formed by Tim23, Nat. Struct. Mol. Biol. 8 (12) (2001) 1074–1082.
[75] K. Hill, K. Model, M.T. Ryan, K. Dietmeier, F. Martin, R. Wagner, N. Pfanner, Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins, Nature 395 (6701) (1998) 516–521.
[76] M.L. Parker, S. Houston, C. Wetherell, C.E. Cameron, M.J. Boulanger, The structure of *Treponema pallidum* Tp0624 reveals a modular assembly of divergently functionalized and previously uncharacterized domains, PLoS One 11 (11) (2016), e0166274.
[77] N.C. Lim, S.E. Jackson, Mechanistic insights into the folding of knotted proteins in vitro and in vivo, J. Mol. Biol. 427 (2) (2015) 248–258.
[78] T. Nishizawa, S. Kita, A.D. Maturana, N. Furuya, K. Hirata, G. Kasuya, S. Ogasawara, N. Dohmae, T. Iwamoto, R. Ishitani, et al., Structural basis for the counter-transport mechanism of a H^+/Ca^{2+} exchanger, Science 341 (6142) (2013) 168–172.
[79] A.B. Waight, B.P. Pedersen, A. Schlessinger, M. Bonomi, B.H. Chau, Z. Roe-Zurz, A.J. Risenmay, A. Sali, R.M. Stroud, Structural basis for alternating access of a eukaryotic calcium/proton exchanger, Nature 499 (7456) (2013) 107–110.
[80] J. Liao, F. Marinelli, C. Lee, Y. Huang, J.D. Faraldo-Gómez, Y. Jiang, Mechanism of extracellular ion exchange and binding-site occlusion in a sodium/calcium exchanger, Nat. Struct. Mol. Biol. 23 (6) (2016) 590–599.
[81] N.C.H. Lim, S.E. Jackson, Molecular knots in biology and chemistry, J. Phys. Condens. Matter 27 (35) (2015).
[82] J.K. Pittman, K.D. Hirschi, Phylogenetic analysis and protein structure modelling identifies distinct Ca^{2+}/cation antiporters and conservation of gene family structure within arabidopsis and rice species, Rice 9 (1) (2016) 3.
[83] E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, Acta Crystallogr. D Biol. Crystallogr. 60 (12) (2004) 2256–2268.
[84] T. Shigaki, I. Rees, L. Nakhleh, K. Hirschi, Identification of three distinct phylogenetic groups of CAX cation/proton antiporters, J. Mol. Evol. 63 (6) (2006) 815–825.
[85] W. Hu, C. Wang, J. Liang, T. Zhang, Z. Hu, Z. Wang, W. Lan, F. Li, H. Wu, J. Ding, G. Wu, Z. Deng, C. Cao, Structural insights into DndE from *Escherichia coli* b7a involved in DNA phosphorothioation modification, Cell Res. 22 (7) (2012) 1203.
[86] K. Gupta, A.A. Watson, T. Baptista, E. Scheer, A.L. Chambers, C. Koehler, J. Zou, I. Obong-Ebong, E. Kandiah, A. Temblador, et al., Architecture of taf11/taf13/tbp complex suggests novel regulation properties of general transcription factor TFIID, elife 6, 2017.
[87] K. Koniaris, M. Muthukumar, Self-entanglement in ring polymers, J. Chem. Phys. 95 (4) (1991) 2873–2881.
[88] P. Virnau, Detection and visualization of physical knots in macromolecules, Phys. Procedia 6 (2010) 117–125.
[89] K.C. Millett, E.J. Rawdon, A. Stasiak, J.I. Sułkowska, Identifying knots in proteins, Biochem. Soc. Trans. 41 (2) (2013) 533–537.
[90] P. Freyd, D. Yetter, J. Hoste, W.R. Lickorish, K. Millett, A. Ocneanu, A new polynomial invariant of knots and links, Bull. Am. Math. Soc. 12 (2) (1985) 239–246.
[91] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: a synthesis, Proteins: Struct. Funct. Bioinf. 21 (3) (1995) 167–195.