



OPEN ACCESS

EDITED BY

Piero Andrea Temussi,
University of Naples Federico II, Italy

REVIEWED BY

Adam Liwo,
University of Gdansk, Poland
Peter Virnau,
Johannes Gutenberg University Mainz,
Germany

*CORRESPONDENCE

Joanna I. Sulkowska,
✉ jsulkowska@cent.uw.edu.pl

RECEIVED 16 May 2023

ACCEPTED 04 October 2023

PUBLISHED 06 June 2024

CITATION

Perlinska AP, Nguyen ML, Pilla SP,
Staszor E, Lewandowska I, Bernat A,
Purta E, Augustyniak R, Bujnicki JM and
Sulkowska JI (2024), Are there double
knots in proteins? Prediction and *in vitro*
verification based on TrmD-Tm1570
fusion from *C. nitroreducens*.
Front. Mol. Biosci. 10:1223830.
doi: 10.3389/fmolb.2023.1223830

COPYRIGHT

© 2024 Perlinska, Nguyen, Pilla, Staszor,
Lewandowska, Bernat, Purta,
Augustyniak, Bujnicki and Sulkowska. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Are there double knots in proteins? Prediction and *in vitro* verification based on TrmD-Tm1570 fusion from *C. nitroreducens*

Agata P. Perlinska¹, Mai Lan Nguyen^{1,2}, Smita P. Pilla¹,
Emilia Staszor^{1,3}, Iwona Lewandowska¹, Agata Bernat⁴,
Elżbieta Purta⁴, Rafal Augustyniak³, Janusz M. Bujnicki⁴ and
Joanna I. Sulkowska^{1*}

¹Centre of New Technologies, University of Warsaw, Warsaw, Poland, ²Polish-Japanese Academy of Information Technology, Warsaw, Poland, ³Faculty of Chemistry, University of Warsaw, Warsaw, Poland, ⁴Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

We have been aware of the existence of knotted proteins for over 30 years—but it is hard to predict what is the most complicated knot that can be formed in proteins. Here, we show new and the most complex knotted topologies recorded to date—double trefoil knots ($3_1\#3_1$). We found five domain arrangements (architectures) that result in a doubly knotted structure in almost a thousand proteins. The double knot topology is found in knotted membrane proteins from the CaCA family, that function as ion transporters, in the group of carbonic anhydrases that catalyze the hydration of carbon dioxide, and in the proteins from the SPOUT superfamily that gathers 3_1 knotted methyltransferases with the active site-forming knot. For each family, we predict the presence of a double knot using AlphaFold and RoseTTaFold structure prediction. In the case of the TrmD-Tm1570 protein, which is a member of SPOUT superfamily, we show that it folds *in vitro* and is biologically active. Our results show that this protein forms a homodimeric structure and retains the ability to modify tRNA, which is the function of the single-domain TrmD protein. However, how the protein folds and is degraded remains unknown.

KEYWORDS

methyltransferase, composite knot, SPOUT, domain, evolution

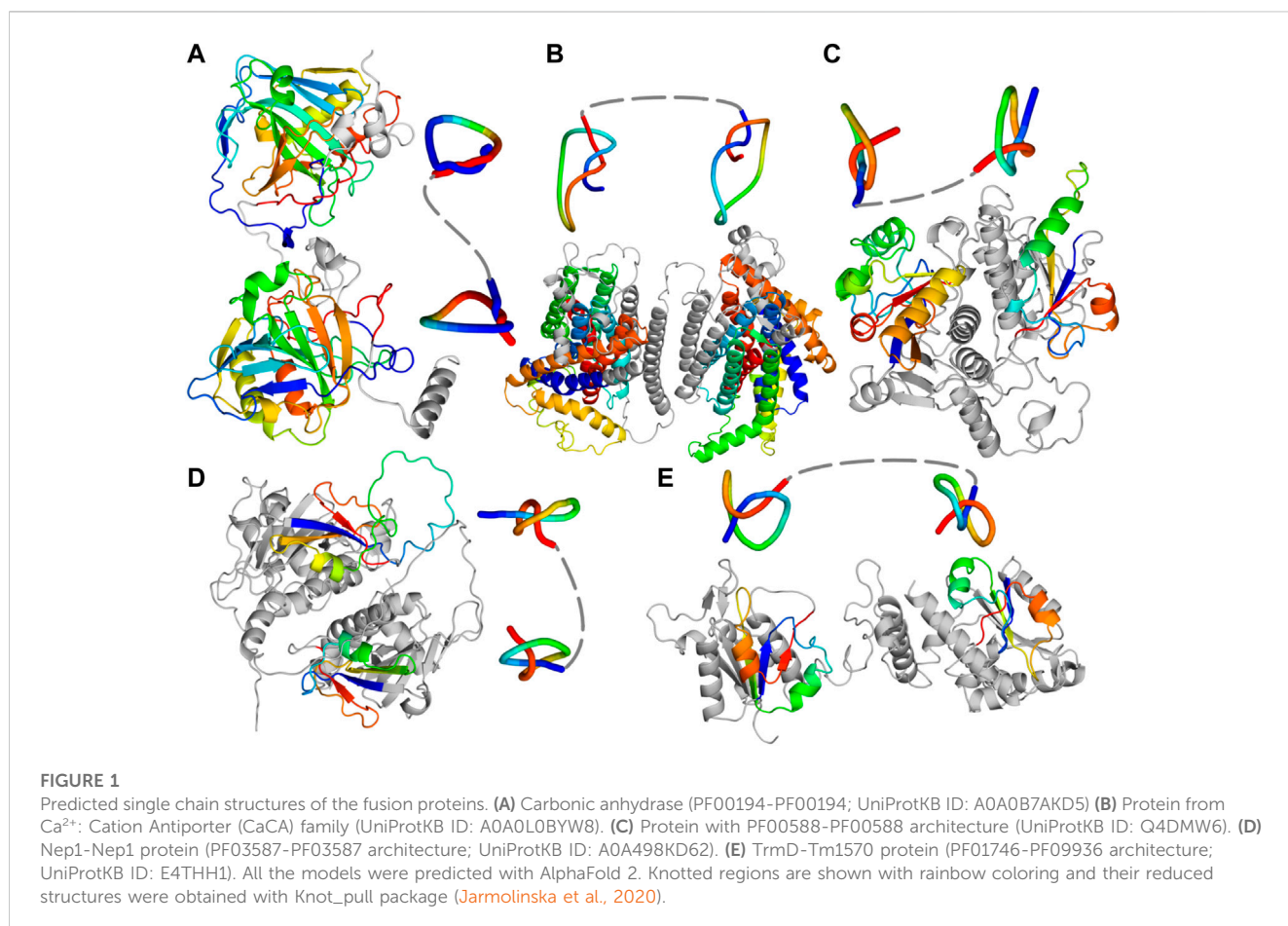
1 Introduction

The presence of knots in proteins has been known for almost 30 years, but only a few types of knots were identified over the years: 3_1 , 4_1 , 5_2 , and 6_1 (Eriksson et al., 1988; Taylor, 2000; Wagner et al., 2005; Das et al., 2006; Schmidberger et al., 2008; Bölinger et al., 2010; Thiruselvam et al., 2017; Sulkowska, 2020). The most commonly encountered knot is the trefoil (3_1), which is found in more than 85% of knotted structures. However, the number of knotted proteins is not high (185) which makes up about 0.3% of all the proteins with structures deposited in the PDB database (Jamroz et al., 2015). This raises at least two fundamental questions. First, why is the number of knotted proteins so low, even though it is

TABLE 1 Proteins with composite knots.

Superfamily	Fusion architecture (Pfam ID)	No. proteins	Example protein (UniProtKB ID)	Knot type
Carbonic anhydrase	PF00194-PF00194	686	A0A0B7AKD5	$3_1\#3_1$
CaCA	PF01699-PF01699-PF01699-PF01699	24	A0A17919N9	$3_1\#3_1$
SPOUT	PF00588-PF00588	125	Q4DMW6	$3_1\#3_1$
SPOUT (Nep1-Nep1)	PF03587-PF03587	34	A0A498KD62	$3_1\#3_1$
SPOUT (TrmD-Tm1570)	PF01746-PF09936	66	E4THH1	$3_1\#3_1$

The majority of the proteins with composite knots represent the SPOUT superfamily, although in most cases the recognition of specific proteins behind that architecture was not possible. Reported Pfam IDs are of entangled domains.



expected from polymer physics to be much higher (Lua and Grosberg 2006)? It could be due to the complicated folding process needed to form the knot, including threading, which is energetically costly (Yeates et al., 2007), or maybe entanglement is disadvantageous in the process of protein degradation. However, since knotted domains frequently contain binding sites for substrates (Sulkowska, 2020), their presence is essential for certain proteins, such as methyltransferases (MTs) (Tkaczuk et al., 2007; Christian et al., 2016; Perlinska et al., 2020b). The second question is why proteins can only form simple types of knots? Why not more complex knots with more crossings or multiple separate knots since multi-domain proteins do exist? Both questions are not trivial to answer. Herein, we will try to solve the second one.

In this paper, we expand our knowledge of protein structure complexity and possible folds by presenting newly identified families of proteins with two separate knots (a so-called composite knot; Table 1). For each family, we determined the structures of doubly knotted proteins using AlphaFold 2 and RoseTTaFold (Figure 1), and in the case of one family (TrmD-Tm1570), we performed a more in-depth *in silico* and *in vitro* study based on *Calditerrivibrio nitroreducens*. All of the proteins with composite knots that we found have two 3_1 knots [based on AlphaKnot database (Niemyska et al., 2022)], which is expected by trefoil's prevalence in the knotted proteins world (Dabrowski-Tumanski et al., 2019). In our search, we have not identified any other structures involving two sequential knots, other than two trefoils. However, just as the single-knotted

proteins, the composite knots are also not commonly found in a proteome, but the scale is different—a well-studied knotted MT called tRNA methyltransferase D (TrmD) is a universal protein for all Bacteria (Ito et al., 2015; Zhong et al., 2019), but the presence of proteins with composite knots is limited to specific organisms, such as *Desulfovibrio vulgaris* or *Oleidesulfovibrio alaskensis*.

Thus we showed that proteins can adopt more complex structures. So, the rarity of knotted proteins might be connected with their complex topology and folding pathway. The knots in proteins (with determined 3D structure) identified to date are exclusively of twist type (Dabrowski-Tumanski et al., 2019), which means that the rate-limiting threading during protein folding occurs only once (Taylor, 2007; Yeates et al., 2007; Bölinger et al., 2010; Sulkowska, 2020). With structures containing two knots (regardless of their type), the threading must happen at least twice, which is a substantial energetic cost for the cell. Therefore, doubly knotted proteins may provide additional benefits over single-knotted proteins to be advantageous for the cell (Dabrowski-Tumanski et al., 2016b; Sulkowska, 2020). This advantage might be connected to the reason for many other proteins fusing together—these proteins are interacting and forming functional complexes while remaining separate (Chia and Kolatkar, 2004; Pasek et al., 2006; Buljan et al., 2010; Marsh et al., 2013). Thus, when their genes are nearby or are fused, they can be expressed together (both in time and place), which is energetically favorable.

Herein, having discussed the general characteristics of four of the five doubly knot families, we focus on one family of doubly knotted proteins—a TrmD-Tm1570 fusion, which joins two domains with MT activity. Both of the domains function also as single proteins (Tkaczuk et al., 2007; Kim et al., 2009; Perlinska et al., 2020a). To better understand the structure, evolution and biological function of this fusion on doubly knotted TrmD-Tm1570 from *C. nitroreducens* (CnTrmD-Tm1570), we conducted an *in vitro* study of this protein's activity and an extensive bioinformatics analysis of sequences and structures of its related proteins. Our results show that the CnTrmD-Tm1570 protein is a functional homodimeric protein that methylates tRNA, and thus the CnTrmD domain is fully operational (Christian et al., 2016). However, the function of the second domain (CnTm1570) is not clear except that it is similar to 2'O ribose-modifying knotted MTs (Kim et al., 2009). Moreover, how the double knotted proteins fold, and whether they can be easily degraded are going to be new open fundamental questions.

2 Materials and methods

2.1 Search for double knotted proteins

First, we obtained information about domains found in entangled protein structures from the KnotProt database (Dabrowski-Tumanski et al., 2019). Next, we calculated the minimal length of each domain, which was based on the shortest knotted region of the domain. This step was important for finding proteins long enough to be doubly knotted. The implicated domains were classified into three groups based on their contribution in creating a knot, slipknot, or both. We then retrieved all the sequences present in the Pfam database (Mistry et al., 2021),

which consisted of at least two entangled functional regions, and created a list of proteins divided by their domain architecture. We selected the proteins with the highest possibility of containing a composite knot. From each domain architecture type, we chose a few sequences (also regarding their length). We considered representative proteins to form a composite knot when their closest homolog with a known 3D structure contained at least one entanglement. Therefore, we identified the most similar sequences within the AlphaFold Protein Structure Database using the Basic Local Alignment Search Tool (BLAST) in ChimeraX 1.3 (Pettersen et al., 2021). The best hit was checked based on its e-value and presence in the entanglements. If the e-value was significant ($e\text{-value} < 10^{-3}$) and the homolog turned out to be knotted, the examined protein was modeled in AlphaFold 2. It is worth noting that all found homologs had e-values between $1e-07$ and 0.0 (with most being between $1e-30$ and 0.0). Obtained structures were additionally analyzed using the AlphaKnot server.

2.2 Dataset preparation and quantification of dimer interfaces

The PDB (Berman et al., 2000) was scanned for entries representing TrmD as well as Tm1570 in different source organisms with a resolution better than 4.0 \AA . We identified the pairwise interactions and quantified the interface area (B) using the following equation:

$$B = \text{SASA}_1 + \text{SASA}_2 - \text{SASA}_{12}. \quad (1)$$

Here, the first two terms represent the solvent accessible surface area (SASA) of the molecules in isolated form and the last term represents the SASA of their binary complex. SASA values were calculated using the program NACCESS (Hubbard and Thornton, 1993), which implements the Lee and Richards algorithm (Lee and Richards, 1971).

2.3 Sequence analysis

The sequences of fusion proteins were retrieved from the UniProt Knowledge base (UniProtKB) (Consortium, 2021), by searching for both protein families of TrmD (Pfam id: PF01746) and Tm1570 (PF09936) proteins. Then the sequences were clustered using the CD-HIT suite (Huang et al., 2010) (three runs, with sequence identity cutoff ranges from 0.9 to 0.7). Multiple sequence alignment was done using Clustal Omega (Madeira et al., 2022) and the conservation of residues was analyzed using UGENE (Okonechnikov et al., 2012).

2.4 Molecular docking

The tRNA from *Haemophilus influenzae* (HiTrmD) (PDB id: 4YVI) was extracted and docked to the TrmD domain of the fusion protein with known constraint (tRNA binding motif) using the HDock server (Yan et al., 2017). The $C\alpha$ root mean square deviation (RMSD) to the PDB structure 4YVI was used to filter the fusion-tRNA complexes, and the best superposed complex was chosen for further study. Following that, four ligands

(S-adenosylmethionine; SAM) were docked to the respective knotted domains of TrmD and Tm1570 in the fusion-tRNA protein-RNA complex using the GLIDE program in the Schrödinger software (Maestro 12.5). The fusion-tRNA structure was prepared using the Protein Preparation Wizard. Prior to docking to Tm1570 we added a water molecule, which based on Tm1570 crystals, is important for a proper pose of SAM. We also rotated the side chain of Asn409 accordingly to Tm1570 crystal structures and minimized the protein, in order for the residue not to block the binding site. Docking to Tm1570 was performed with H-bond constraints defined on the amine group of Ala357 and the backbone oxygens of Ile402 and Asn409.

2.5 Structure prediction with AlphaFold and RoseTTaFold cross-validation

Structures were predicted with our locally installed newest AlphaFold 2 version (2.1.0). Proteins with potential composite knots based on the AlphaFold 2 calculations were validated on the RoseTTaFold server (Baek et al., 2021) using default parameters. Since RoseTTaFold does not accept sequences with more than 1,200 residues, the larger proteins were modeled within this range. All the modelled structures were deposited in the Github repository (https://github.com/ilbsm/Double_knots_structures/tree/main).

2.6 Experimental procedures

Codon-optimized gene coding for TrmD-Tm1570 was synthesized (GenArt, Thermo Scientific) and inserted into the pET28b expression vector. Transformed *E. coli* BL21 (DE3) RIL cells were grown at 37°C until OD₆₀₀ reached 0.8. After induction with 1 mM IPTG the temperature was decreased to 18°C and bacteria were harvested 18 h later. The cell pellet was resuspended in a lysis buffer containing 6 M urea, 20 mM Hepes, 300 mM NaCl, 10 mM imidazole and 0.01% NaN₃. The protein was purified in a HisTrap HP 5 mL column (GE Healthcare). After elution, the protein was dialyzed for at least 5 h against 2 L of buffer (50 mM Hepes, 300 mM NaCl, 20% glycerol, 0.01% NaN₃, 2 mM DTT, pH 7.4) in each concentration of urea containing 2 M, 1 M, 0.8 M, 0.6 M, 0 M (5 different buffers). The final purification step involved a preparative Superdex 75 column with the running buffer containing 50 mM Hepes, 300 mM NaCl, 10% glycerol, 1 mM TCEP, pH 7.4. The pure protein was flash-frozen in liquid nitrogen and stored at -80°C. Activity assays were performed with the MTase-Glo kit (Promega) according to the manufacturer's instructions. The luminescence was performed on a Synergy H1 plate reader (Biotek). Further details are given in SI Materials and Methods.

3 Results

Our search for composite knots in proteins is based on two approaches, with direct and indirect use of available protein structures. The indirect approach is focused on the protein

sequence and domain annotation. We use the KnotProt database (Dabrowski-Tumanski et al., 2019) and the information it contains about knotted proteins along with the location of the knot in their structures. We associate the presence of a knot with the domain in which it is located. With this data, we searched for multi-domain proteins with more than one entangled domain. For this purpose, we use the Pfam database that contains protein domains grouped into families and superfamilies. This resulted in finding five domain architectures that could possess double knots (Table 1).

In our second approach, we analyze the available protein structures directly. Thanks to the ever-growing number of predicted models with the machine learning methods like AlphaFold (Jumper et al., 2021; Varadi et al., 2022) and our AlphaKnot database that analyses their topology (Niemyska et al., 2022), we were able to find proteins with double knots based on their 3D structure. In order to minimize the probability of finding artifacts, we used only models predicted with high confidence (pLDDT score). We did not encounter any other fusion architectures but found seven additional proteins for the PF00588-PF00588 architecture, that we already found with the first method (Table 1).

Finally, we used locally installed AlphaFold 2 (original model; 2.1.0) to predict the structures of selected proteins (at least two members from each family). Moreover, for additional verification, we used RoseTTaFold to predict structures for the same protein sequences and compare their topologies. The example doubly knotted protein from each of the architectures is presented in Figure 1 and Table 1. The other proteins with predicted doubly knotted structures are shown in the Supplementary Table S1).

All of the architectures we found represent composite $3_1\#3_1$ knots, which is expected since the trefoil (3_1) knot is the most common knot type found in proteins. As anticipated, the majority of the architectures are formed by domains from the SPOUT superfamily since it is the biggest group of deeply knotted 3_1 proteins (Sulkowska, 2020). Moreover, we found composite knots also within membrane proteins, namely, the ion transporters.

3.1 Carbonic anhydrases

A substantial number of composite knots is found within carbonic anhydrases (Sayre et al., 2011; Dzubielia, 2013). This is a well-known group of knotted proteins with many structures resolved experimentally. Most of these enzymes have a single domain (PF00194) that contains either a deep or a shallow 3_1 knot (human carbonic anhydrase IX and II, respectively). The fusion of two such domains results in a double 3_1 knotted protein, with a knot encompassing most of the structure (Figure 1A). We found this architecture (PF00194-PF00194) in 686 protein sequences (131 AlphaFold models with $3_1\#3_1$ can be found in the AlphaKnot database. The difference in the number may be due to the fact that several sequences may be too short to form double knot). Importantly, carbonic anhydrases can form disulfide bridges. These bridges can be formed both within a single chain, thereby creating a lasso that stabilizes a shallow knot (Dabrowski-Tumanski et al. 2016a; Niewieczerał and Sulkowska 2019; Niemyska et al., 2020). They can also be formed between the

monomers forming a dimer (Alterio et al. 2009; De Simone and Supuran 2010).

3.2 The Ca²⁺: Cation Antiporter (CaCA) family—transmembrane protein

This family of membrane transporters already contains known knotted proteins, such as Vacuolar cation/proton exchanger (CAX) with a 3₁ knot (Jarmolinska et al., 2019). This protein has two PF01699 domains that together form a knot. Here, we found proteins with four such domains which indicate that they could possess two knots within their structure. This architecture (PF01699-PF01699-PF01699-PF01699) is present in 24 proteins (Table 1) and none of them has a resolved structure. Figure 1B shows a structure we predicted with AlphaFold (RoseTTaFold also predicted two knots in this protein). Other proteins from this group with the potential to be doubly knotted are listed in the supplement (Supplementary Table S1).

3.3 The SPOUT family

Apart from the TrmD-Tm1570 proteins that we describe in more detail below, the SPOUT family contains at least two different architectures also forming a composite 3₁#3₁ knot (Table 1). Most probably they are a result of gene duplication—the gene coding the knot-containing domain fused with its duplicate resulting in a double knotted protein. Individual SPOUT families (Pfam ID) shown in Table 1 represent the MT domain, which functions as a single protein in many organisms (Hou et al., 2017). The common structural feature among all of the structures we predicted is the fact that the duplicated domains are interacting *via* the vast interface (Figures 1C, D). Furthermore, the arrangement of the domains is identical to that formed by single-domain proteins in their homodimeric complex (Supplementary Figure S1).

3.3.1 PF00588-PF00588—the largest group of doubly knotted proteins

The PF00588 is a large protein family with over 35,000 sequences of MTs modifying 2′O of ribose in either tRNA or rRNA. The PF00588-PF00588 fusion proteins form the biggest group of the doubly knotted proteins we found. Interestingly, we encountered proteins with this architecture using domain annotations in Pfam (118 hits) and structure searches in AlphaKnot (seven hits). Since these seven proteins do not have both domains annotated in the Pfam database, we used HHpred to obtain them (Supplementary Table S2).

Both domains of this architecture contain a compact 3₁ knot, which is characteristic of the SPOUT superfamily. All seven models look like a fused dimeric complex of a knotted MT (Figure 1C; Supplementary Figure S1) with the domains arranged in a perpendicular fashion, for example, in TrmH (from the same family) or Tm1570 (from PF09936 family) crystal structures (Kim et al., 2009).

3.3.2 PF03587-PF03587—Nep1-Nep1 fusion

There are 34 proteins with the double knot topology and PF03587-PF03587 domain architecture. The PF03587 family

groups Ribosomal RNA small subunit MTs Nep1 (Nep1) that methylate pseudouridine at position 1,189 (Psi1189) in 18S rRNA. Most of these proteins function as homodimeric single-domain proteins. The predicted structure of Nep1-Nep1 fusion we found (Figure 1D) resembles the crystal structure of Nep1 dimeric complex (Supplementary Figure S1)—similar as in the case of PF00588-PF00588 proteins.

3.3.3 PF01746-PF09936—TrmD-Tm1570 fusion

The next group of doubly knotted proteins is formed by the PF01746-PF09936 architecture with both families found within SPOUT knotted MTs. There are a couple of crucial differences between these proteins and the other SPOUT fusions we discussed above: 1) the evolutionary mechanism does not involve gene duplication, 2) in a single chain the two domains interact using a minimal interface, 3) TrmD-Tm1570 is a homodimer, whereas the aforementioned fusion proteins can function as monomers. All of these points are discussed in detail in the sections below.

There are 66 proteins with the PF01746-PF09936 annotation in Pfam and all are present in Bacteria. Within the PF01746 family, there is a well-studied TrmD protein (Ito et al., 2015) that modifies the N1 position of G37 in tRNA. It is a universal bacterial protein that in a vast majority of organisms is a single-domain protein. In the PF01746-PF09936 architecture, TrmD is fused with Tm1570.

The Tm1570 protein belongs to the SAM-dependent RNA MT family (Pfam ID: PF09936). It contains only 299 proteins and one of them, namely, Tm1570 from *Thermotoga maritima*, has been crystallized (PDB ID: 3dcm). Based on the structure similarity search we performed with Dali (Holm and Laakso, 2016) using the crystal structure, the protein is most similar to TrmJ MTs that modify cytidine 32 at the 2′O position in tRNA (Purta et al. 2006) (Supplementary Table S3). Therefore, it is highly probable that proteins with the PF09936 domain bind tRNA and perform methylation of ribose 2′O. Within the whole PF09936 family, two types of domain architecture can be found: a single-domain protein (like Tm1570) or a two-domain protein, which is the TrmD-Tm1570 fusion.

By analyzing the location of the genes coding the domains of both families (Pfam ID: PF01746 and PF09936) in Bacteria, we found four different ways in which the genes are co-located in the genomes: 1) two single genes in different parts of the genome; 2) two single genes adjacent on the genome; 3) two single but overlapping genes; 4) one fused gene. Figure 2 shows the arrangement in some example organisms. This might represent the evolutionary pathway that led to the creation of the fusion, which started as two separate genes. Even though we did not find evidence supporting the interaction between single-domain TrmD and Tm1570, we hypothesize that the proteins can form a complex and it was more advantageous for the cell to express them, and in the end to fuse them, together.

Next, to verify the presence of the two knots in these proteins, we predicted their structures from nine different organisms (see Supplementary Figure S2) using AlphaFold 2. None of the nine formed a compact single chain structure as is the case in the other families of double knotted proteins from SPOUT superfamily we analyzed (e.g., Nep1-Nep1; Figure 1B). Instead, the nine form an open conformation with the domains scarcely interacting with each other (Figure 1E). This behavior is expected since the two proteins in their single-domain forms are dimerizing in different ways: TrmD in an antiparallel fashion and Tm1570 in a perpendicular fashion

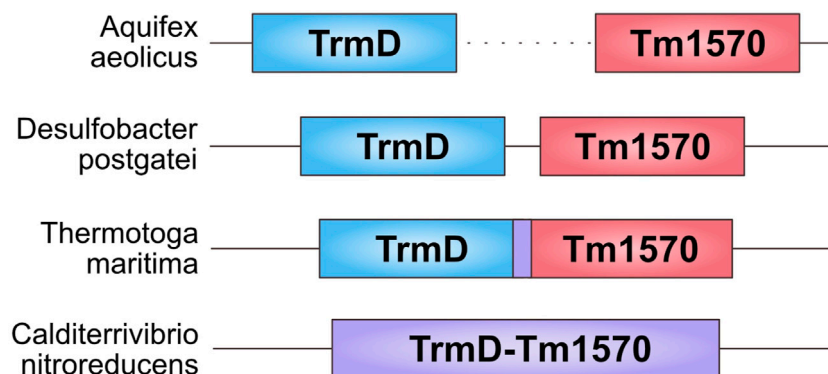


FIGURE 2

Arrangement of TrmD and Tm1570 genes in different organisms. From the top: *Aquifex aeolicus*—the genes are separated by thousands of nucleotides; *Desulfobacter postgatei*—the genes are 11 nucleotides apart; *Thermotoga maritima*—the genes are overlapping by 7 nucleotides; *Calditerrivibrio nitroreducens*—the genes are fused.

(Figure 3). This suggests that TrmD-Tm1570 functions as a dimer and the domains interact with their counterparts from the other chain.

3.4 TrmD-Tm1570 from *Calditerrivibrio nitroreducens*

3.4.1 Homodimeric structure

From the set of double knotted proteins we found, we investigated further one from the TrmD-Tm1570 family from *Calditerrivibrio nitroreducens* (*Cn*TrmD-Tm1570). We used this protein to characterize the structural basis for substrate recognition and the overall structural landscape of double knotted proteins.

Our theoretical and experimental analysis shows that the *Cn*TrmD-Tm1570 protein functions as a homodimer, unlike other double knotted proteins we study here. Therefore, we used AlphaFold Multimer to model the structure and obtained a compact homodimeric complex (Figure 3) with the main interchain interactions present between the corresponding domains. Both of the domains interact in a standard fashion (antiparallel for TrmD and perpendicular for Tm1570) with their counterparts from the other chain. To further verify the model we analyzed the dimeric interface of the fusion and compared it with the ones that are created in single-domain TrmDs and Tm1570s. We analyzed all TrmD and Tm1570 structures available in the PDB along with their homologs (Supplementary Table S4) and found key residues that are crucial for the dimeric interface in the single-domain proteins (the sequence similarity between different TrmD proteins is low, although, the structures can be superimposed very well). These amino acids are conserved in both the TrmD and Tm1570 domains of the fusion (Figure 4; Supplementary Tables S5, S6).

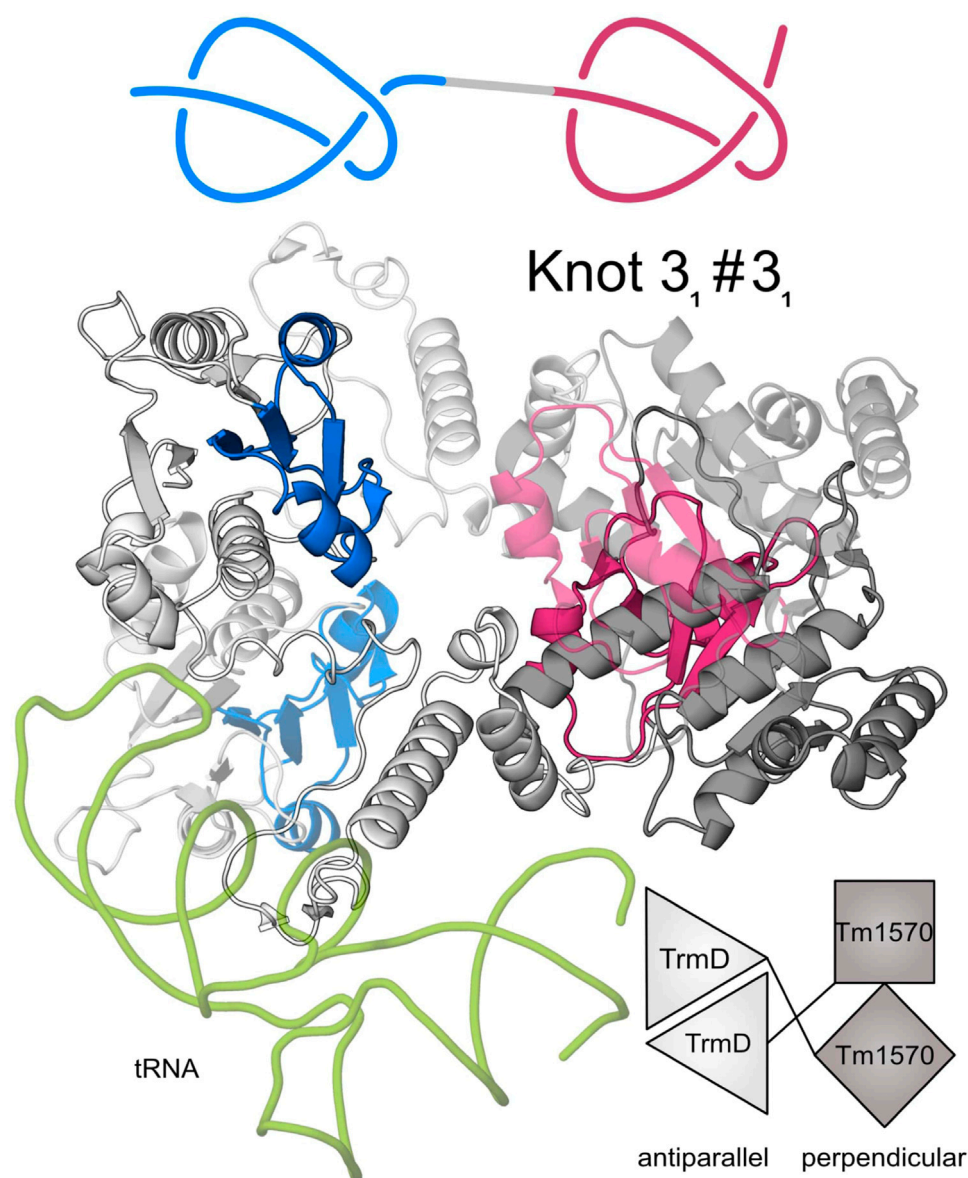
Details concerning the mechanism of function, including conformations of the residues in the active site as well as the steps of chemical catalysis, are well understood for TrmD (single-domain proteins) but not for Tm1570. Herein, based on our analysis of structure and sequence, we found that deep trefoil knots provide the binding sites for SAM in both the TrmD and Tm1570 domains of the fusion. Moreover, they are both structurally and sequentially similar to their counterparts from single-domain proteins (Figure 4).

Thus the ligand binding modes should be similar in both domains. More precisely, the binding site-forming knot in *Haemophilus influenzae* TrmD (*Hi*TrmD) consists of three loops: the cover (Ser88-Gly91), the wall (Gly113-Ile118), and the bottom loop (Ser132-Gly140) (Jaroensuk et al., 2019). The residues of the corresponding three loops in the fusion protein are conserved (the cover loop: Asp87-Gly90, the wall: Gly112-Ile117, and the bottom: Ser131-Gly139). The residues Pro88, Gly90, Arg113, Glu115, Gly116, Ser131, Gly133 and Asp134 are strictly conserved and are involved in the TrmD dimeric interactions (Supplementary Table S6). We used this information to obtain the complex of TrmD-Tm1570 with SAM *via* molecular docking. The resulting structure has four SAM molecules (one ligand in one of the four available active sites), each adopting a conformation that is proper and characteristic for the SPOUT superfamily (Perlinska et al., 2020b) (Supplementary Figure S3).

Finally, based on the known binding mode of tRNA to *Hi*TrmD, we established how the nucleic acid may interact with the *Cn*TrmD-Tm1570 dimer. We found that residues involved in *Hi*TrmD-tRNA interactions are also involved in the fusion (Figure 4). The SGHH motif (residues 198–201 in *Hi*TrmD) that interacts with G37 in the substrate tRNAs for the methylation process (Ito et al., 2015) is present in SGNH form in the *C. nitroreducens* fusion. All of this information strongly suggests that the predicted model correctly demonstrates how the active complex of TrmD-Tm1570 is constructed.

3.4.2 TrmD-Tm1570 is an active tRNA methyltransferase

In order to experimentally assess the activity of the fusion TrmD-Tm1570 from *C. nitroreducens*, we expressed the His-tagged protein in *E. coli* cells. However, the full-length protein, as well as its individual truncated domains encompassing residues 1–240 (TrmD) and 241–433 (Tm1570) (see Figure 5A), was found in the insoluble fractions even when growing bacteria at low temperatures. Nevertheless, we managed to perform purification under denaturing conditions and obtained large quantities of structured and functional proteins after the final refolding step. Size-exclusion chromatography suggests that all fragments form stable dimers given the retention volumes on the

**FIGURE 3**

Predicted structure of CnTrmD-Tm1570 fusion protein based on AlphaFold and docking. This homodimeric (second chain is transparent) complex binds tRNA (green) with its TrmD domains. A single chain of this protein contains two 3_1 knots (marked in blue and red) (Niemyska et al. 2022). TrmD domain (light grey) interacts with its counterpart from the second chain in an antiparallel fashion, whereas Tm1570 (dark grey) in a perpendicular fashion. Details about the modeled complex are in the [Methods](#) section.

preparative Superdex 75 column (GE Healthcare; [Supplementary Figure S4](#)): 56 mL for the fusion TrmD-Tm1570 (49.6 kDa monomer), 63 mL for the TrmD (27.5 kDa monomer), and 65 mL for the Tm1570 domain (22.1 kDa monomer). It is noteworthy that the single-domain *Ec*TrmD (28.4 kDa monomer), which was previously characterized as a dimer (Elkins et al., 2003) and was now expressed for the sake of reference, eluted from the same column at 63 mL. Moreover, we were able to obtain small quantities of the soluble Tm1570 protein without unfolding procedures and this protein migrated through the gel filtration column as Tm1570 subjected to *in vitro* renaturation.

We tested the MT activity of the recombinant TrmD-Tm1570 using the commercially available MTase-Glo kit (Promega), which

allows one to monitor the S-Adenosyl-L-homocysteine (SAH) concentration build-up during the enzymatic reaction. First, we used tRNA Leu (CAG) from *E. coli* (see [Supplementary Figure S5](#) for exact sequence) as the acceptor of the methyl moiety. It turned out that CnTrmD-Tm1570 fusion protein was able to efficiently modify this cloverleaf structure in the presence of SAM. However, as seen in [Figure 5B](#), the same reaction was almost two times faster when the native *C. nitroreducens* substrate, homologous to *E. coli* tRNA Leu (CAG), was used (see [Supplementary Figure S5](#)). Although the key nucleotides, including the G37G38 motif, are conserved in both tRNA sequences, not surprisingly TrmD-Tm1570 shows preference towards its native substrate, originating from the same organism as the protein. To verify the importance of the aforementioned G37G38 motif for the

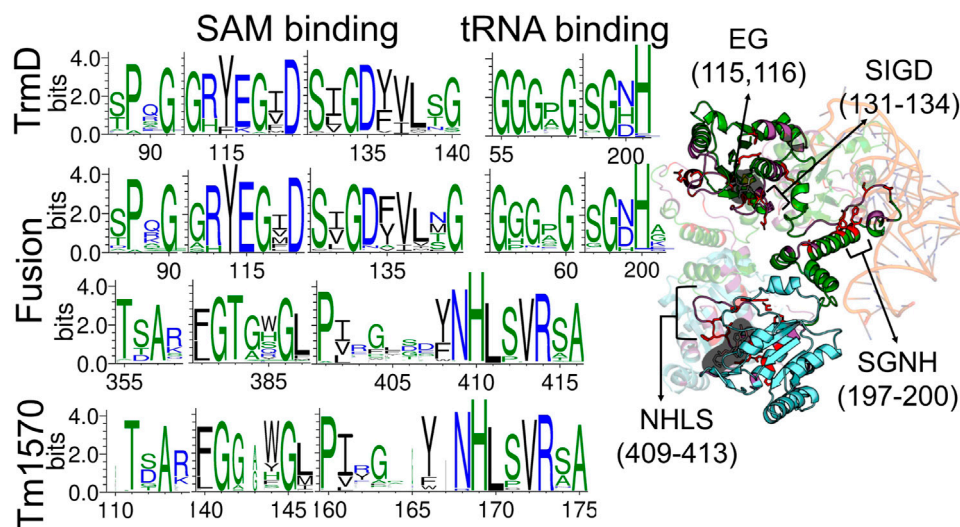


FIGURE 4 Comparison of sequence motifs in TrmD and Tm1570 with the fusion protein. WebLogo3 was used to construct the conserved residue logos at the SAM and tRNA binding sites. Cartoon representation shows the TrmD-Tm1570 protein (TrmD domain in green, Tm1570 in blue) with selected motifs marked on the structure.

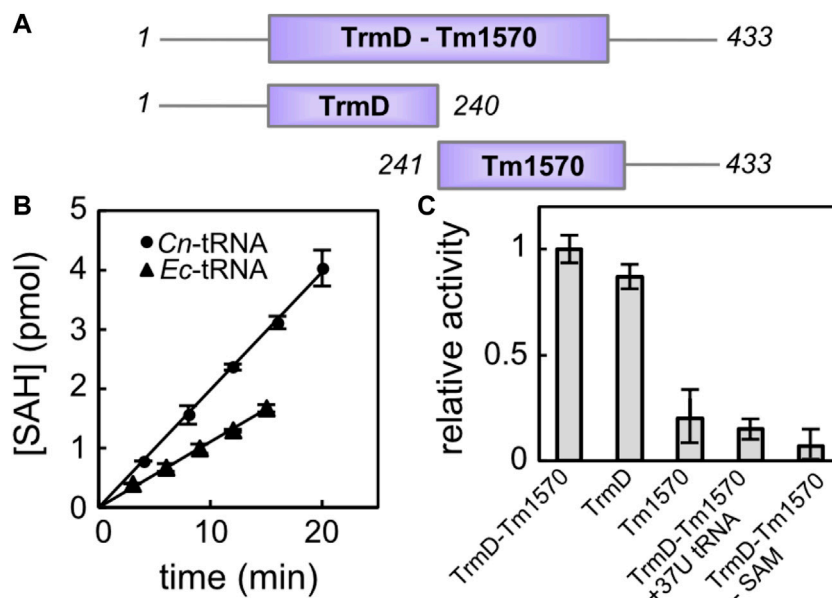


FIGURE 5 Activity of CnTrmD-Tm1570. (A) Protein constructs used for the activity assessment. (B) Time-course of the reaction catalyzed by fusion TrmD-Tm1570 (50 nM) in the presence of SAM (30 μM) towards the 8 μM tRNA Leu (CAG) substrate (*E. coli* tRNA black triangles, *C. nitroreducens* tRNA black circles). (C) Relative activities of different TrmD-Tm1570 constructs. Data obtained for the mutated tRNA substrate and negative control are also included. Reaction conditions are the same as in (B).

reaction catalyzed by TrmD-Tm1570 we also prepared a mutant of *C. nitroreducens* tRNA Leu (CAG), where the guanosine at position 37 was replaced by thymine at the level of the DNA template. As expected for this substrate, the double knotted enzyme showed no activity, comparable to the negative control lacking SAM in the reaction mixture. The activity towards the mutated substrate is higher (at least twice) than control (Figure 5C).

Since we wanted to examine the contribution of both domains to the overall activity of the fusion TrmD-Tm1570 protein, we followed the reaction catalyzed by truncated versions of the full-length protein. With respect to the native *C. nitroreducens* tRNA Leu (CAG) substrate, the TrmD domain retained 86% ± 4% activity of the fusion protein, while Tm1570 retained only 18% ± 9% (see Figure 5C). These data clearly demonstrate that the enzymatic activity of the fusion TrmD-

Tm1570 protein towards tRNA Leu (CAG) substrate is governed by the TrmD domain. However, the Tm1570 domain may contribute to substrate binding and increase the stability of the complex. On the other hand, small activity of the isolated Tm1570, comparable to the negative control which had SAM omitted from the reaction mixture, may suggest that this protein, whose physiological function remains unknown, is a tRNA methyltransferase but with different substrate specificity. It is also possible that Tm1570 regains its full activity in the presence of a specific ligand or under yet-to-be-discovered conditions.

4 Conclusion

Answering the question from the title: yes, proteins with double knots do exist. Here, for the first time we found and analyzed in-depth such proteins. All of them have two 3_1 knots and come from three different protein superfamilies (five distinct architectures). They are either transmembrane ion transporters (from the Ca^{2+} : Cation Antiporter family), carbonic anhydrases, or methyltransferases (from the SPOUT family). Within the SPOUT group, there are three architectures in proteins with either two duplicated domains (PF00588-PF00588 and PF03587-PF03587) or two separate domains (PF01746-PF09936—TrmD-Tm1570). We found that these two groups differ in terms of structure organization—the duplicated and fused domains form dimer-like single chain structures, whereas the TrmD-Tm1570 proteins need two chains to resemble a functional SPOUT MT (which are mostly dimers).

Using both theoretical and experimental approaches we studied in detail TrmD-Tm1570—a fusion between TrmD and Tm1570 proteins found in 66 organisms. Based on *C. nitroreducens* we established that the protein is a homodimer capable of binding four ligands (S-adenosylmethionine)—one in each knotted binding site, and a single tRNA molecule (based on the similarities with *Hi*TrmD). Moreover, the *Cn*TrmD-Tm1570 is folding and functioning *in vitro*—it methylates tRNA using its TrmD domain. Unfortunately, we were not able to determine the function of the Tm1570 domain—which is unknown for single-domain proteins as well. However, based on structural similarities, we believe that it also acts as a tRNA MT, probably by modifying 2' O in ribose. During the preparation of this manuscript, we were able to solve experimentally the structure of *Cn*TrmD-Tm1570 protein and deposit it in the PDB (PDB id: 8b1n) (da Silva et al., 2023). Our model and the crystal structure are similar, in particular, both of them show “open” domain organization that enables homodimer formation.

Herein, we have identified proteins with knotted $3_1\#3_1$ structure, composed of two sequential trefoils. A natural question is whether more complicated composite knots also exist in nature. In our search we have not identified any other more complicated sequential pairs (involving other knots than two trefoils). However, another hypothetically possible (albeit quite unlikely) structure involves one trefoil knot formed within another trefoil. Identification of such a structure would require much more sophisticated search; if such a structure exists, it would very likely have even more complex functional properties. Another method would need to be employed to find such a structure, because the methods used in this research are specific to the case of sequential trefoils.

Note that our research was conducted based on the sequences of all known knotted proteins (domains) (Dabrowski-Tumanski et al., 2019), including predicted knotted proteins (Perlinska et al., 2023) based on the results of AlphaFold 2 till the end of June 2022. During the preparation of this manuscript, a paper by Brems et al. (2022) was published showing the composite knots in AlphaFold structures of SPOUT methyltransferases and carbonic anhydrases. Moreover, given that new primary knots are being identified in structures predicted by AlphaFold Niemyska et al. (2022), we predict that different types of composite knots exist in nature, and their identification is an important task for future research.

Finally, let us comment on folding of proteins with $3_1\#3_1$ knots. Based on known folding pathways of proteins from SPOUT family one could imagine knotting the N-terminal chain directly on the ribosome (Chwastyk and Cieplak, 2015; Dabrowski-Tumanski et al., 2018; Baiesi et al., 2019), while the C-terminal knot could follow a well known slipknot pathway (Wallin et al., 2007; Sulkowska et al., 2009) or other mechanisms suggested for proteins with a deep knot based on numerical simulations (Potestio et al., 2010; Li et al., 2012). Another possibility is to use flipping mechanism observed in numerical simulations for proteins with, e.g., 6_1 knot (Bölinger et al., 2010). This mechanism was later developed to a topological descriptor of knot folding by Flapan et al. (2019). One could extend this scenario to explain folding of newly identified proteins with no-twist types of knots, such as 6_3 recently predicted based on AlphaFold approach (Perlinska et al., 2022). However, it is not clear how to use it directly to $3_1\#3_1$ knots; moreover, in the case of all doubly knotted proteins listed in this work, we do not see loops responsible for flipping. Reconstructing a folding pathway of composite knots is also an important task for future research.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

AP and JS: investigation and writing—original manuscript. AP and MN: formal analysis, methodology, validation, visualization, and writing—review and editing. SP: formal analysis, investigation. ES, IL, AB: conducting experimental investigation. AB, EP, and JB: editing of the revised manuscript. RA: conceptualization, data curation and writing—original manuscript. JB and EP: conceptualization of experimental part. JS: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, supervision, validation, visualization, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Science Centre #UMO-2018/31/B/NZ1/04016, 2020/39/I/NZ1/03582 and 2021/43/I/NZ1/03341 to JS.

Acknowledgments

This research was carried out with the support of the Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw under computational allocation no GS82-12 and COST EUTOPIA action.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alterio, V., Hilvo, M., Di Fiore, A., Supuran, C. T., Pan, P., Parkkila, S., et al. (2009). Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase ix. *Proc. Natl. Acad. Sci.* 106, 16233–16238. doi:10.1073/pnas.0908301106
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Baiesi, M., Orlandini, E., Seno, F., and Trovato, A. (2019). Sequence and structural patterns detected in entangled proteins reveal the importance of co-translational folding. *Sci. Rep.* 9, 8426–8512. doi:10.1038/s41598-019-44928-3
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bölinger, D., Sulkowska, J. I., Hsu, H.-P., Mirny, L. A., Kardar, M., Onuchic, J. N., et al. (2010). A Stevedore's protein knot. *PLoS Comput. Biol.* 6, e1000731. doi:10.1371/journal.pcbi.1000731
- Brems, M. A., Runkel, R., Yeates, T. O., and Virnau, P. (2022). AlphaFold predicts the most complex protein knot and composite protein knots. *Protein Sci.* 31, e4380. doi:10.1002/pro.4380
- Buljan, M., Frankish, A., and Bateman, A. (2010). Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol.* 11, R74–R15. doi:10.1186/gb-2010-11-7-r74
- Chia, J.-M., and Kolatkar, P. R. (2004). Implications for domain fusion protein-protein interactions based on structural information. *BMC Bioinforma.* 5, 161–167. doi:10.1186/1471-2105-5-161
- Christian, T., Sakaguchi, R., Perlinska, A. P., Lahoud, G., Ito, T., Taylor, E. A., et al. (2016). Methyl transfer by substrate signaling from a knotted protein fold. *Nat. Struct. Mol. Biol.* 23, 941–948. doi:10.1038/nsmb.3282
- Chwastyk, M., and Cieplak, M. (2015). Cotranslational folding of deeply knotted proteins. *J. Phys. Condens. Matter* 27, 354105. doi:10.1088/0953-8984/27/35/354105
- Consortium, T. U. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi:10.1093/nar/gkaa1100
- da Silva, F. B., Lewandowska, I., Kluzka, A., Niewieczeral, S., Augustyniak, R., and Sulkowska, J. I. (2023). First crystal structure of double knotted protein TrmD-Tm1570 – inside from degradation perspective. bioRxiv. Available at: <https://www.biorxiv.org/content/10.1101/2023.03.13.532328v1>.
- Dabrowski-Tumanski, P., Niemyska, W., Pasznik, P., and Sulkowska, J. I. (2016a). Lassoprot: server to analyze biopolymers with lassos. *Nucleic acids Res.* 44, W383–W389. doi:10.1093/nar/gkw308
- Dabrowski-Tumanski, P., Piejko, M., Niewieczeral, S., Stasiak, A., and Sulkowska, J. I. (2018). Protein knotting by active threading of nascent polypeptide chain exiting from the ribosome exit channel. *J. Phys. Chem. B* 122, 11616–11625. doi:10.1021/acs.jpcc.8b07634
- Dabrowski-Tumanski, P., Rubach, P., Goundaroulis, D., Dorier, J., Sulkowski, P., Millett, K. C., et al. (2019). KnotProt 2.0: a database of proteins with knots and other entangled structures. *Nucleic Acids Res.* 47, D367–D375. doi:10.1093/nar/gky1140
- Dabrowski-Tumanski, P., Stasiak, A., and Sulkowska, J. I. (2016b). In search of functional advantages of knots in proteins. *PLoS One* 11, e0165986. doi:10.1371/journal.pone.0165986
- Das, C., Hoang, Q. Q., Kreinbring, C. A., Luchansky, S. J., Meray, R. K., Ray, S. S., et al. (2006). Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase uch-11. *Proc. Natl. Acad. Sci.* 103, 4675–4680. doi:10.1073/pnas.0510403103
- De Simone, G., and Supuran, C. T. (2010). Carbonic anhydrase ix: biochemical and crystallographic characterization of a novel antitumor target. *Biochimica Biophysica Acta (BBA)-Proteins Proteomics* 1804, 404–409. doi:10.1016/j.bbapap.2009.07.027
- Dzubiella, J. (2013). Tightening and Untying the Knot in Human Carbonic Anhydrase III. *J. Phys. Chem. Lett.* 4 (11), 1829–1833. doi:10.1021/jz400748b
- Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M., et al. (2003). Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J. Mol. Biol.* 333, 931–949. doi:10.1016/j.jmb.2003.09.011
- Eriksson, A. E., Jones, T. A., and Liljas, A. (1988). Refined structure of human carbonic anhydrase ii at 2.0 Å resolution. *Proteins Struct. Funct. Bioinforma.* 4, 274–282. doi:10.1002/prot.340040406
- Flapan, E., He, A., and Wong, H. (2019). Topological descriptions of protein folding. *Proc. Natl. Acad. Sci.* 116, 9360–9369. doi:10.1073/pnas.1808312116
- Holm, L., and Laakso, L. M. (2016). Dali server update. *Nucleic Acids Res.* 44, W351–W355. doi:10.1093/nar/gkw357
- Hou, Y.-M., Matsubara, R., Takase, R., Masuda, I., and Sulkowska, J. I. (2017). TrmD: a methyl transferase for tRNA methylation with m1G37. *Enzym.* 41, 89–115. doi:10.1016/bs.enz.2017.03.003
- Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi:10.1093/bioinformatics/btq003
- Hubbard, S. J., and Thornton, J. M. (1993). *Naccess. Computer program*. London: Department of Biochemistry and Molecular Biology, University College.
- Ito, T., Masuda, I., Yoshida, K.-i., Goto-Ito, S., Sekine, S.-i., Suh, S. W., et al. (2015). Structural basis for methyl-donor-dependent and sequence-specific binding to tRNA substrates by knotted methyltransferase TrmD. *Proc. Natl. Acad. Sci.* 112, E4197–E4205. doi:10.1073/pnas.1422981112
- Jamroz, M., Niemyska, W., Rawdon, E. J., Stasiak, A., Millett, K. C., Sulkowski, P., et al. (2015). KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.* 43, D306–D314. doi:10.1093/nar/gku1059
- Jarmolinska, A. I., Gambin, A., and Sulkowska, J. I. (2020). Knot_pull—python package for biopolymer smoothing and knot detection. *Bioinformatics* 36, 953–955. doi:10.1093/bioinformatics/btz644
- Jarmolinska, A. I., Perlinska, A. P., Runkel, R., Trefz, B., Ginn, H. M., Virnau, P., et al. (2019). Proteins' knotty problems. *J. Mol. Biol.* 431, 244–257. doi:10.1016/j.jmb.2018.10.012
- Jaroensuk, J., Wong, Y. H., Zhong, W., Liew, C. W., Maenpuen, S., Sahili, A. E., et al. (2019). Crystal structure and catalytic mechanism of the essential m1G37 tRNA methyltransferase TrmD from *Pseudomonas aeruginosa*. *RNA* 25, 1481–1496. doi:10.1261/rna.066746.118
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kim, D. J., Kim, H. S., Lee, S. J., and Suh, S. W. (2009). Crystal structure of *Thermotoga maritima* SPOUT superfamily RNA methyltransferase Tm1570 in complex with S-adenosyl-L-methionine. *Proteins Struct. Funct. Bioinforma.* 74, 245–249. doi:10.1002/prot.22249
- Lee, B., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400. doi:10.1016/0022-2836(71)90324-x

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1223830/full#supplementary-material>

- Li, W., Terakawa, T., Wang, W., and Takada, S. (2012). Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc. Natl. Acad. Sci.* 109, 17789–17794. doi:10.1073/pnas.1201807109
- Lua, R. C., and Grosberg, A. Y. (2006). Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput. Biol.* 2, e45. doi:10.1371/journal.pcbi.0020045
- Madeira, F., Pearce, M., Tivey, A., Basutkar, P., Lee, J., Edbali, O., et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. doi:10.1093/nar/gkac240
- Marsh, J. A., Hernández, H., Hall, Z., Ahnert, S. E., Perica, T., Robinson, C. V., et al. (2013). Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* 153, 461–470. doi:10.1016/j.cell.2013.02.044
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913
- Niemyska, W., Millett, K. C., and Sulkowska, J. I. (2020). Gln: a method to reveal unique properties of lasso type topology in proteins. *Sci. Rep.* 10, 15186. doi:10.1038/s41598-020-71874-2
- Niemyska, W., Rubach, P., Gren, B. A., Nguyen, M. L., Garstka, W., Bruno da Silva, F., et al. (2022). AlphaKnot: server to analyze entanglement in structures predicted by AlphaFold methods. *Nucleic Acids Res.* 50, W44–W50. doi:10.1093/nar/gkac388
- Niewiczzerzał, S., and Sulkowska, J. I. (2019). Supercoiling in a protein increases its stability. *Phys. Rev. Lett.* 123, 138102. doi:10.1103/PhysRevLett.123.138102
- Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi:10.1093/bioinformatics/bts091
- Pasek, S., Risler, J.-L., and Brézellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22, 1418–1423. doi:10.1093/bioinformatics/btl135
- Perlinska, A. P., Kalek, M., Christian, T., Hou, Y.-M., and Sulkowska, J. I. (2020a). Mg²⁺-Dependent methyl transfer by a knotted protein: a molecular dynamics simulation and quantum mechanics study. *ACS Catal.* 10, 8058–8068. doi:10.1021/acscatal.0c00059
- Perlinska, A. P., Niemyska, W. H., Gren, B. A., Bukowicki, M., Nowakowski, S., Rubach, P., et al. (2023). AlphaFold predicts novel human proteins with knots. *Protein Sci.* 32, e4631. doi:10.1002/pro.4631
- Perlinska, A. P., Niemyska, W. H., Gren, B. A., Rubach, P., and Sulkowska, J. I. (2022). New 63 knot and other knots in human proteome from AlphaFold predictions. bioRxiv. Available at: <https://www.biorxiv.org/content/10.1101/2021.12.30.474018v1>.
- Perlinska, A. P., Stasiulewicz, A., Nawrocka, E. K., Kazimierczuk, K., Setny, P., and Sulkowska, J. I. (2020b). Restriction of S-adenosylmethionine conformational freedom by knotted protein binding sites. *PLoS Comput. Biol.* 16, e1007904. doi:10.1371/journal.pcbi.1007904
- Petersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi:10.1002/pro.3943
- Potestio, R., Micheletti, C., and Orland, H. (2010). Knotted vs. unknotted proteins: evidence of knot-promoting loops. *PLoS Comput. Biol.* 6, e1000864. doi:10.1371/journal.pcbi.1000864
- Purta, E., Van Vliet, F., Tkaczuk, K. L., Dunin-Horkawicz, S., Mori, H., Droogmans, L., et al. (2006). The yfhQ gene of *Escherichia coli* encodes a tRNA: cm32/Um32 methyltransferase. *BMC Mol. Biol.* 7, 23–13. doi:10.1186/1471-2199-7-23
- Sayre, T. C., Lee, T. M., King, N. P., and Yeates, T. O. (2011). Protein stabilization in a highly knotted protein polymer. *Protein Eng. Des. Sel.: PEDS* 24 (8), 627–630. doi:10.1093/protein/gzr024
- Schmidberger, J. W., Wilce, J. A., Weightman, A. J., Whisstock, J. C., and Wilce, M. C. (2008). The crystal structure of dehi reveals a new α -haloacid dehalogenase fold and active-site mechanism. *J. Mol. Biol.* 378, 284–294. doi:10.1016/j.jmb.2008.02.035
- Sulkowska, J. I., Sulkowski, P., and Onuchic, J. (2009). Dodging the crisis of folding proteins with knots. *Proc. Natl. Acad. Sci.* 106, 3119–3124. doi:10.1073/pnas.0811147106
- Sulkowska, J. I. (2020). On folding of entangled proteins: knots, lassos, links and theta-curves. *Curr. Opin. Struct. Biol.* 1, 131–141. doi:10.1016/j.sbi.2020.01.007
- Taylor, W. R. (2000). A deeply knotted protein structure and how it might fold. *Nature* 406, 916–919. doi:10.1038/35022623
- Taylor, W. R. (2007). Protein knots and fold complexity: some new twists. *Comput. Biol. Chem.* 31, 151–162. doi:10.1016/j.compbiolchem.2007.03.002
- Thiruseviam, V., Kumarevel, T., Karthe, P., Kuramitsu, S., Yokoyama, S., and Ponnuswamy, M. N. (2017). Crystal structure analysis of a hypothetical protein (mj0366) from *Methanocaldococcus jannaschii* revealed a novel topological arrangement of the knot fold. *Biochem. Biophysical Res. Commun.* 482, 264–269. doi:10.1016/j.bbrc.2016.11.052
- Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M. (2007). Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinforma.* 8, 73–31. doi:10.1186/1471-2105-8-73
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi:10.1093/nar/gkab1061
- Wagner, J. R., Brunzelle, J. S., Forest, K. T., and Vierstra, R. D. (2005). A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 438, 325–331. doi:10.1038/nature04118
- Wallin, S., Zeldovich, K. B., and Shakhnovich, E. I. (2007). The folding mechanics of a knotted protein. *J. Mol. Biol.* 368, 884–893. doi:10.1016/j.jmb.2007.02.035
- Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S.-Y. (2017). HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45, W365–W373. doi:10.1093/nar/gkx407
- Yeates, T. O., Norcross, T. S., and King, N. P. (2007). Knotted and topologically complex proteins as models for studying folding and stability. *Curr. Opin. Chem. Biol.* 11, 595–603. doi:10.1016/j.cbpa.2007.10.002
- Zhong, W., Koay, A., Ngo, A., Li, Y., Nah, Q., Wong, Y. H., et al. (2019). Targeting the bacterial epitranscriptome for antibiotic development: discovery of novel tRNA-(N1G37) methyltransferase (TrmD) inhibitors. *ACS Infect. Dis.* 5, 326–335. doi:10.1021/acscinfecdis.8b00275