Abstract: The role of knots in proteins is still unclear. In this short review, we summarize the current knowledge about structural and dynamical differences between knotted and unknotted proteins. We show how the topological difference helps to distinguish the physical properties, characterize the biological activity or identify the biological function of knotted proteins. This knowledge can be used to correctly annotate protein family members and to identify new members.

Keywords: knotted proteins, methyltransferases, TrmD, Trm5, TrmJ, Trm7

DOI: https://doi.org/10.17466/tq2016/20.4/h

1. Introduction

Since 2010, every year has brought about a hundred new knotted protein structures deposited in the PDB [1, 2]. Many proteins have a knotted region located far from the terminal ends of a structure, therefore forming a deep knot, as shown in Figure 1. This implies that the knotted core is a stable region which cannot be easily untied by thermal fluctuations. However, the reason why such complex structures are formed is still not clear. Moreover, it has been shown that the knotting pattern is strictly conserved between several protein families,
despite their large sequence divergence [3]. More than 90% of knotted proteins are enzymes the active sites of which are located in the vicinity of the knotted core. It has been also shown that folding of a knotted protein is slower and less efficient than that of an unknotted one [4, 5]. All of the above suggests that the entanglement provides advantages and some important functions to proteins and living cells [3, 6]. Searching and understanding those advantages is one of the current challenges of the contemporary biophysics.

Currently, two main approaches are explored. In the first one, researchers focus on comparison of similar proteins that differ in topology, where one has a knot and the other is its unknotted (trivial) counterpart. Such comparison will help in discovering what advantages a protein gains with the presence of a knot. According to our knowledge, it is only one such pair that is known and it consists of two broadly studied proteins: aspartate carbamoyltransferase (ATCase a knotted protein) and ornithine carbamoyltransferase (OTCase). Based on molecular dynamics simulations with the structure-based model, it has been shown that the knotted protein (PDB ID: 1yh1) is mechanically more resistant than the unknotted one (PDB ID: 1c9y) [8]. The bioinformatics approach has been used to identify new members of knotted and unknotted proteins inside this carbamoyltransferase family [9]. It has been shown that the knotted protein can be identified by the presence of a proline-rich loop and an extended 120’s loop. The second approach to studying the role of the knot is based on the analysis of proteins that are analogous in function but are structurally distinct and have a different topology. Methyltransferases make an excellent set for such analysis because of their diversity and numerous examples of both knotted and unknotted proteins [10–12]. The most
prominent pair of this set, which has been broadly studied, is a pair of tRNA (guanine(37)-N(1))-methyltransferases, containing knotted TrmD, and unknotted Trm5 (Figure 2) [13–16]. The reason behind the popularity of this pair lies in their potential use in the drug design. Knotted TrmD proteins occur in Bacteria, whereas unknotted Trm5 proteins occur in Archaea and Eukaryota (human included) [17]. Both proteins maintain the same important function: methylation of N1 atom of guanine 37 in tRNA [18, 19]. Guanine 37 (G37) is adjacent to the anticodon loop and its modification positively influences correct codon-anticodon pairings, therefore it helps in the translation process. The lack of the methyl group in G37 interferes with the growth of the cell and leads to its death [20]. Another pair of analogous proteins that have different topologies are TrmJ and Trm7. These proteins are responsible for methylation of the ribose moiety of nucleosides 32 in tRNA [21]. In this review we report discovered differences between pairs of analogous enzymes, their significance in the drug design based on both experimental and theoretical (all-atom explicit solvent simulations) methods [22] and we also present a few new findings.

**Figure 2.** Cartoon representations of examined proteins; on the left the figure shows crystal structures of knotted methyltransferases: TrmD in a complex with tRNA (PDB ID: 4yvi [7]) and TrmJ (PDB ID: 4xbo [23]); colored in pink and violet, respectively; on the right unknotted analogs of previously mentioned enzymes are shown: the crystal structure of Trm5 in a complex with tRNA (PDB ID: 2zzm [24]) colored in dark green and Trm7 colored in green; due to the lack of the crystal structure of Trm7, its model was obtained via homology modeling [25, 26]; nucleic acids are colored in orange
2. Knotted and unknotted tRNA (guanine(37)-N(1))-methyltransferases

The methylation of N1 of guanine 37 in the tRNA molecule is performed by two distinct proteins: TrmD and Trm5 (Figure 2). TrmD is a member of class IV of methyltransferases that is characterized by the presence of the trefoil knot which is the simplest non-trivial knot. Currently, TrmD is assigned to the SpouT family all the members of which have an alpha/beta knot topology according to the CATH classification. In TrmD from Haemophilus influenzae the core of the knot is located from Y86 to E130 and is part of the N-terminal domain (Figure 1)[1]. The ligand (SAM – S-adenosylmethionine) binds in the bent conformation, characteristic for knotted proteins, into the crevice created by the knotted region. Trm5 binds the same ligand but in different (open) conformation and to a differently built active site. Trm5 is a monomeric protein from class I of methyltransferases which possess a trivial Rossmann Fold structure (Figure 2). It is worth mentioning that historically TrmD and Trm5 were assigned to the same – Rossmann Fold – family.

2.1. Catalytic cycle

TrmD and Trm5 are distant unrelated proteins that differ in structure and topology [17]. Moreover, since they come from different domains of life, they also possess different mechanisms for performing the same function. TrmD is a homodimeric protein which functions in a specific way: only half of the active sites are reactive at the time since only one tRNA molecule binds to the complex. The rate limiting step in the TrmD catalysis is the methyl transfer rate, whereas in monomeric Trm5 it is the release of one of the reaction products – SAH (S-adenosylhomocysteine – demethylated substrate). Kinetic analysis shows that TrmD has a higher affinity for tRNA than Trm5 but lower for SAM [16]. This suggests that the substrates interact differently with enzymes.

2.2. Ligand-protein interactions

Our results [22] show that TrmD and Trm5 maintain the ligand-protein binding differently. Besides, the structurally distinct active sites (one knotted, the other unknotted), there are also different ligand-protein interactions between them. In the Trm5 protein, the active site is created by mostly nonpolar amino acids (Figure 3 (a)). This leads to the stabilization of the ligand by hydrophobic interactions rather than polar ones. In TrmD the situation is quite opposite – the ligand is bound by at least 7 hydrogen bonds (Figure 3 (b)).

2.3. Stability and dynamics

Recently, based on extensive all-atom explicit solvent simulations, it has been shown that knots are not significantly more rigid than their unknotted counterpart [22]. Amino acids within the active site of TrmD and Trm5 show comparable fluctuations (Figure 4) and both active sites are the least fluctuating regions. Even though the knot is thought to be a rigid structure that participates in the stabilization of the protein, it behaves in a similar fashion as the unknotted
Figure 3. Ligand-protein interactions: (a) contributions of polar and nonpolar amino acids in the active site of Trm5 and TrmD; (b) occurrence of particular hydrogen bonds between ligand (SAM) and the protein in molecular dynamics simulations of TrmD and Trm5 with tRNAs; the amino acids colored red are acidic, violet are polar, black are hydrophobic and the only basic amino acid here is blue.

Figure 4. Root Mean Square Fluctuation (RMSF) of C-alpha carbons of amino acids of TrmD and Trm5 from molecular dynamics simulations; the red area shows the TrmD active site (from amino acid 81 to amino acid 156), the green area shows the Trm5 active site (from amino acid 197 to amino acid 265) [22].

active site. These results show that the knot is not a strictly rigid structure, despite the fact that this region is the least fluctuating part of the enzyme.

An analysis of the protein’s dynamics with Principal Component Analysis (PCA) showed that unknotted protein could be described by fewer principal components than a knotted protein to achieve 80% of the total variance in the system. This indicates that a knotted protein has more complex dynamics, contrary to the Trm5 motions. The catalytically active motions of the knot have been shown to be exceptionally diffusive: its first principal component describes only 16% of the fluctuations, whereas the first principal component of the active site of Trm5 describes 39.5%. It all shows that knot is not a rigid but a complex, dynamic structure capable of participation in a signaling pathway [22].
2.4. Implications for drug design

Since the TrmD and Trm5 proteins are coming from different domains of life, the opportunity arises for creation of a selective inhibitor for new antimicrobial drugs which will interact only with TrmD. Such a drug would impair the growth of the bacteria and, at the same time, would not be harmful to the eukaryotic organisms. The potency of that scenario has been already acknowledged: TrmD has been indicated as an important drug target [27]. Due to the significance of the issue, it is essential to better understand the differences between TrmD and Trm5 in order to design an efficient inhibitor.

3. Knotted and unknotted tRNA (cytidine(32)-2'-O)-methyltransferases

Another pair of analogous proteins with a different topology is TrmJ (knotted) and Trm7 (unknotted) shown in Figure 2. The TrmJ protein is a member of class IV methyltransferases with the core of the knot located from G78 to Y130 [23] in the N-terminal domain (Figure 5). This enzyme was first recognized as citidine(32)/uridine(32)-2'-O tRNA methyltransferase [21], but latest studies have confirmed that it is also responsible for adenosine(32)-2'-O in the tRNA methylation and for the oxidative stress response [28]. Based on the enzyme sensitivity to the folding correctness of tRNA[29], bacterial TrmJ is classified as full-length tRNA-dependent, whereas its homolog from Archaea requires only the acceptor stem and the anticodon loop to perform methylation [10]. Trm7, an analogous protein to TrmJ from Eucaryota, not related to TrmJ, is a member of class I methyltransferases [30]. Trm7 modifies nucleotides at position 32 and 34 in the tRNA anticodon loop by adding a methyl group to 2'-O-ribose in the presence of TRM732 and TRM734, respectively [31]. One of the known homologous proteins of Trm7 is the human FtsJ enzyme. Those proteins have 39% sequence similarity, therefore, FtsJ was used as a template during structure homology modeling [25, 26]. The tertiary structure of Trm7 has been obtained for 189 amino acids (out of 310).

![Figure 5. Cartoon representation of tRNA (cytidine(32)/uridine(32)-2'-O)-methyltransferase – TrmJ (PDB ID: 4xbo [32]): The left panel shows a protein chain with a rainbow colored knotted region; The right panel shows a schematic representation of the trefoil knot identified in this protein. In this protein the knot core forms a part of the active site similarly as for the TrmD protein](image-url)
3.1. Implications for drug design

The competitive inhibitors of TrmJ could be used as new antibacterial drugs for dangerous human pathogens such as *Pseudomonas aeruginosa* [28]. It has been shown that patients with cystic fibrosis (CF) or chronic obstructive pulmonary disease (COPD) often suffer from infections caused by this bacterium [33]. In 13% of the cases (which have occurred in the United States), the infection is caused by a Multi-drug resistant (MDR) bacterium and every year about 400 people die because of it [34]. Hence, it is important to create new and better antibiotics against these MDR bacteria. An extensive analysis of protein-ligand interactions and binding site flexibility will be important and necessary to achieve this objective.

4. Conclusions

A theoretical approach based on molecular dynamics simulations, structural and sequential characterization of analogous proteins with a different topology leads to a better understanding of the role of the knot in the protein structure. For now, it is known that the knot has dynamic capabilities necessary for being an active part in the enzyme function. Such studies are also an important step towards designing novel drugs against bacteria and life-threatening diseases. Our results show that the corresponding proteins, knotted and unknotted, are sufficiently different and it is possible to design a selective drug based on structural and topological differences. However, there is still much to uncover for the TrmJ-Trm7 pair which needs further and more detailed studies, similar to those made for TrmD-Trm5.

To fully understand the role of the knot, a more comprehensive survey should be performed, based on a set of analogous pairs of proteins with a different topology. According to our knowledge, there are no such pairs that could be studied. Nevertheless, there are cases of knotted methyltransferases whose function has not been established yet, e.g. a probable methyltransferase from *Haemophilus influenzae* (PDB ID: 3ilk [35]). It may be a good material for further studies. It is easier to find an unknotted analogue, because knotted proteins are not as frequent as unknotted ones. Another possibility to develop this field is to find analogues of one of the broadly studied proteins, such as the knotted methyltransferase TrmL. TrmL is an extensively studied enzyme, in terms of its structure [36], mechanical resistance [37], folding and unfolding [38–40] and the tRNA recognition mechanism [32], which is why it is desirable to find a similar unknotted protein. A comparison of those would significantly bring us closer to understanding the role of the knot in the protein structure. Moreover, this information can be used to correctly annotate protein family members and to identify new members.

Acknowledgements

Scientific work financed from the budget for science in the years 2016–2019 (Grant #0003/ID3/2016/64 Ideas Plus to JS) and by the National Science Centre (#2012/07/E/NZ1/01900).
References

[16] Bystrom A S, van Gabain A and Bjork G R 1989 *Journal of Molecular Biology* 208 (4) 575
[19] Pieper U et al. 2014 *Nucleic Acids Research* 42 (D1) ???
[31] Guy M P *et al.* 2012 *RNA* **18** (10) 1921
[33] Spagnuolo L *et al.* 2016 *Scientific Reports* **6** 36924
[34] CDC 2013 *Antibiotic resistance threats in the united state* 69
[36] Lim K *et al.* 2003 *Proteins: Structure, Function and Bioinformatics* **51** (1) 56
[38] Sułkowska J I, Sułkowski P, Szymczak P and Cieplak M 2010 *Journal of the American Chemical Society* **132** (40) 13954