# PAPER • OPEN ACCESS

# Molecular knots in biology and chemistry

To cite this article: Nicole C H Lim and Sophie E Jackson 2015 J. Phys.: Condens. Matter 27 354101

View the article online for updates and enhancements.

# You may also like

- <u>Statics and dynamics of DNA knotting</u> Enzo Orlandini
- <u>A SURVEY OF MOLECULAR</u> <u>HYDROGEN IN THE CRAB NEBULA</u> E. D. Loh, J. A. Baldwin, Z. K. Curtis et al.

 Prediction of the optimal set of contacts to fold the smallest knotted protein
 P Dabrowski-Tumanski, A I Jarmolinska and J I Sulkowska 8

#### doi:10.1088/0953-8984/27/35/354101

# Molecular knots in biology and chemistry

# Nicole C H Lim<sup>1,2</sup> and Sophie E Jackson<sup>1</sup>

<sup>1</sup> Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
 <sup>2</sup> Faculty of Sciences, Universiti Brunei Darussalam, Gadong BE 1410, Brunei Darussalam

E-mail: chnl2@cam.ac.uk and sej13@cam.ac.uk

Received 8 April 2015, revised 12 May 2015 Accepted for publication 18 May 2015 Published 20 August 2015



### Abstract

Knots and entanglements are ubiquitous. Beyond their aesthetic appeal, these fascinating topological entities can be either useful or cumbersome. In recent decades, the importance and prevalence of molecular knots have been increasingly recognised by scientists from different disciplines. In this review, we provide an overview on the various molecular knots found in naturally occurring biological systems (DNA, RNA and proteins), and those created by synthetic chemists. We discuss the current knowledge in these fields, including recent developments in experimental and, in some cases, computational studies which are beginning to shed light into the complex interplay between the structure, formation and properties of these topologically intricate molecules.

Keywords: DNA knots, RNA pseudoknots, protein knots, protein slipknots, synthetic knots

(Some figures may appear in colour only in the online journal)

# 1. Introduction

Knots and entanglements are common topological features observed not only in the macroscopic world, but also at the molecular level (figure 1). In everyday life, they can be found in various useful applications, from applying surgical sutures to tying shoelaces. However, in some cases, knots can be a nuisance, for example, they can form spontaneously in electrical cables, headphones and garden pipes. They can also lead to undesirable outcomes such as the obstruction of blood circulation to the fetus when tight knots form in the umbilical cord during human pregnancy [1].

Recently, the importance and prevalence of knots at a molecular level have become truly apparent and this has attracted increasing interest from scientists in different fields. In nature, molecular knots (including slipknots and pseudo-knots) are found throughout biology and exist in three major classes of biopolymers: DNA, RNA and proteins [7–15]. Although it is still unclear as to whether these complex topologies are evolutionary advantageous, most natural knots are thought to play a significant role in the structural, dynamic and/or functional properties of the biological systems they

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

are associated with. In addition, molecular knots are increasingly becoming targets of chemical synthesis [16, 17]. Understanding how knots form at a molecular level as well as how the properties of knotted molecular structures differ from unknotted ones is vital.

This review highlights some of the molecular knotted structures discovered in biology and chemistry. It focuses on the structural and mechanistic studies into which and how knots are formed, and summarises the recent developments made towards understanding their properties and potential functions. The review begins with a brief introduction to the classification and detection of knots, followed by an overview of knotted DNA, RNA pseudoknots, protein knots and slipknots, as well as synthetic molecular knots.

## 2. Classification and detection of knots

Concepts from the mathematical field of knot theory have been applied in almost all branches of science, providing tools essential for the detection and classification of different knotted structures. Mathematically, a knot (sometimes termed as a 'true knot') is defined as a topological state of a closed loop that is impossible to untie without being spliced [18]. Technically, this means that knots cannot be defined in open chains. However, many knots such as those found in biological



**Figure 1.** Examples of macroscopic (a)–(d) and molecular (e)–(g) knots. (a) Surgical suture knots used to close a wound [2]. (b) Tying a shoelace knot [3]. (c) Knots formed in entangled earphones. (d) A tight knot formed in an umbilical cord [4]. (e) Electron micrograph of a knotted DNA; figure taken with permission from reference [5]. (f) Ribbon diagram of a stevedore ( $6_1$ ) knotted  $\alpha$ -haloacid dehalogenase protein, PDB code: 4N2X. Inset: simplified view of the protein chain showing the knot. (g) Chemical structure of a synthesised organic trefoil knot. Inset: schematic representation of the knotted structure [6].

systems are open chains. In the case of a simple linear string, one considers it knotted if it does not disentangle itself after being pulled at both ends. This idea is usually applied to open chains and is analogous to their ends being unambiguously connected with a loop to produce a corresponding closed curve.

Detecting knots in topologically complex systems is often not straightforward and requires mathematical methods to both detect and classify the knot type. To identify knotted structures, various algorithms can be employed. One of the simplest knot detection algorithms, known as the Alexander polynomial, can detect and classify a knot according to the minimum number of crossings in a projection of the chain onto a plane [18]. Each knot type is labelled in accordance with the Alexander-Briggs notation, where the first number is the crossing number (usually a measure of knot complexity) and the subscripted index number denotes the knot's order amongst all knots with that crossing number. A simple ring with zero crossings is referred to as the unknot  $(0_1)$  or the trivial knot whilst the simplest, non-trivial knot type is the trefoil knot  $(3_1)$  with three crossings. Other common knot types include the figure-of-eight knot  $(4_1)$  that has four crossings, two knots with five crossings  $(5_1, 5_2)$  and three knots with six crossings  $(6_1, 6_2, 6_3)$  (figure 2). In addition to the Alexander polynomial, the Jones and HOMFLY polynomials are more advanced algorithms that can discriminate between increasingly complex knot types. Further details of these polynomials are provided elsewhere [18–21].

It is important to note that amongst these knot polynomials, the HOMFLY polynomial is a powerful method for detecting the chirality of knots. However, even HOMFLY



Figure 2. Common knot types with up to six crossings denoted by the Alexander–Briggs notation. Knots were generated using KnotPlot (http://knotplot.com/).

can not characterise chirality in all cases [22, 23]. Most knots are not equivalent to their mirror images and they are usually known as *chiral* knots. The simplest chiral knot is the trefoil knot ( $3_1$ ), which comes in a left and a right-handed form, as shown in figure 3. In contrast, *achiral* knots are knots that can be converted to (or are indistinguishable from) their mirror images. Examples include the trivial ( $0_1$ ) and figure-of-eight ( $4_1$ ) knots. In knot theory, knots can also be classified as either *torus* or *twist* knots. Torus knots are a family of knots that can be drawn as closed curves on the surface of a torus (equivalent to a holed-doughnut) and include the  $3_1$ ,  $5_1$ ,  $7_1$  knots, etc. Twist knots, on the other hand, are knots that can be formed by linking together the ends of a repeatedly twisted, closed loop and comprise the  $4_1$ ,  $5_2$ ,  $6_1$  knots, etc.

Although the polynomials are useful for analysing simpler knots, they cannot differentiate knots with projections of many



Figure 3. The two distinct chiral trefoil knots; left and right-handed trefoil knots are illustrated on the left and right, respectively. Knots were generated using KnotPlot (http://knotplot.com/).

crossings or detect knots in extensively knotted systems, as these tend to be computationally challenging. In order to solve this problem, an alternative smoothing algorithm, sometimes referred to as the KMT reduction, was developed such that complex knotted structures are simplified by omitting regions of the chain unnecessary for maintaining the knot [24, 25]. This method produces highly reduced configurations of the original chain and, thus allows efficient computation of the polynomials. In the case of protein structures, this reduction algorithm is very useful for depicting the knotted chain in a simplified manner so that knots can be detected directly and easily visualised [26, 27]. Additionally, as proposed by Taylor, the method can also simultaneously pinpoint the location and depth of the knotted core by calculating the smallest number of residues that can be removed from each side before the structure becomes unknotted [26]. 'Shallow' knots tend to disappear when a few amino acids are deleted from each terminus whilst 'deep' knots remain until a significant amount of the chain (more than 20 amino acid residues on either side of the knotted core) have been removed. However, depending on how the chain is reduced, this method can result in the classification of different knot types. Millett and co-workers have introduced a relatively simple, unbiased method known as the uniform closure method, in which the free ends of a linear open chain are connected to random, uniformly chosen points on a large sphere surrounding the chain [28]. The procedure is repeated many times and a spectrum of knots is obtained, in which the knot type that is dominant is labelled as the knot type of the chain.

# 3. DNA

DNA (deoxyribonucleic acid) is a molecule that encodes the genetic information required for the development and functioning of all living organisms and many viruses. It is not only used as a template for replication but it is also involved in RNA synthesis, which, in some cases, leads on to protein synthesis. Based on the Watson–Crick model, DNA consists of two complementary polynucleotide chains that are intertwined around each other, forming a right-handed double helix [29] (figure 4(a)). DNA can exist as a linear or a closed circular form and is typically tightly packaged. As a result of the structure and metabolism of the double helix, DNA molecules can form three topological states: knotted, catenated or supercoiled (figure 4(b)). In this section, we briefly discuss knots in naturally occurring DNA, mainly focussing on the knotting mechanism and its biological consequences.

#### 3.1. Knots in DNA: structure and formation

A DNA knot is defined as the self-entanglement of a single DNA molecule, therefore this excludes catenane structures that are formed by more than one chain (figure 5(a)). In 1976, Liu and co-workers first discovered that single-stranded DNA chains in bacteriophages could knot when treated with Escherichia coli omega protein, a type I topoisomerase [30]. This was subsequently followed by the discovery of knots in double-stranded DNA chains in 1980 when a supercoiled plasmid was incubated with excess amounts of type II topoisomerase from bacteriophage T4 [31]. Since then, various knotted structures formed in nicked, circular duplex DNA molecules by E. coli topoisomerase I have been identified in vitro, ranging from simple trefoil knots to more complex higher order and composite knots (figure 5(b)) [9]. With the use of electron microscopy imaging and agarose gel electrophoresis, Dean and co-workers characterised these topologically different knotted DNA structures in detail [9].

In the last three decades, an increasing number of studies of DNA knots have been undertaken [32-35]. As discussed above, knots in DNA can form in vitro when DNA strands are cut and re-joined with the help of topoisomerases. DNA topoisomerases control the topology of DNA by introducing transient breaks in DNA strands then re-ligating them to different ends [36, 37]. They are classified into two types: type I or type II. Type I topoisomerases mediate the passage of a single strand of duplex DNA through a nick in the complementary strand. In contrast, type II topoisomerases introduce a transient double-stranded break in one segment of the DNA, allowing a second segment of duplex DNA to pass through before the strands are chemically ligated. A variety of knotted DNA products can also form when recombinases act on supercoiled circular DNA substrates (an example is shown in figure 5(c) [38–40]. Recombinases are involved in changing the topology of DNA by a complex process called site-specific recombination [41]. In this case, they mediate genome rearrangement such that a DNA segment is inserted, excised or inverted in accordance with the appropriate recombination sites [41].

DNA knots can also arise *in vivo* during replication and transcription, as these processes require the action of topoisomerases to release accumulated torsional stress in the DNA [42]. In partially replicated bacterial plasmids with two origins of replication in head-to-head orientation, it has been observed that topoisomerases induce knot formation within replication bubbles that are helically wound (figure 5(d)) [35]. Olavarrieta and co-workers have also shown that complex knotting of the duplex DNA in small pBR322-derived plasmids can be initiated by a head-on collision of replication and transcription, resulting in plasmid instability in *E. coli* (figure 5(e)) [43]. Recently, the Schvartzman group has suggested that if the progression of the replication forks in DNA synthesis



**Figure 4.** (a) Double-helical structure of a DNA molecule, PDB code: 3BSE. Cartoon representation generated using Pymol (www.pymol. org/). (b) Different topological forms of a DNA molecule, formation of which is catalysed by type II topoisomerases: (i) supercoiled, (ii) catenated and (iii) knotted. A single line represents a double strand of DNA.

is impaired, sister duplexes can become loosely intertwined and this can lead to the introduction of knots by the action of topoisomerase IV (Topo IV) [44]. It should be noted, however, that these observations are made on small bacterial plasmids and whether they are applicable to large bacterial or eukaryotic chromosomes is still uncertain.

Several studies have also previously reported that linear viral genomic DNA can cyclise and form knots upon extraction from P4 bacteriophages (figure 5(f)) [31, 45]. Furthermore, it was found that the probability of DNA knotting was enhanced in intact P4 deletion mutants [46] and tailless P4 phages [47]. In a series of experiments, Arsuaga and co-workers showed that most viral DNA molecules (>95%) are highly knotted due to the tight confinement and writhe bias of their packing geometry within the phage capsid (figure 5(g)) [7, 33]. Writhe is the amount a piece of DNA is deformed to form coils as a result of torsional stress, which leads to the phenomenon of DNA supercoiling. Although the specific mechanism of knot formation is still unclear, characterisation of the complex knot spectrum of bacteriophage P4 genome by high-resolution gel electrophoresis revealed that chiral and torus knots were favoured by confinement over achiral and twist ones [7]. Results from recent simulations also showed that there was a preference for chiral knots, albeit no significant bias of torus over twist knots was found [48]. As yet, it remains to be seen what factors actually determine viral genome organisation in terms of its knot types and distribution.

#### 3.2. Biological consequences of DNA knotting

How does DNA knotting affect its biological activity within cells? As discussed above, several processes such as DNA compaction, topoisomerisation, site-specific recombination, replication and transcription can result in the formation of DNA knots in cells. However, the presence of knots in DNA has potentially detrimental effects in several cellular processes such as transcription and replication [50–52] and, if unresolved, can lead to mutational defects in the genome or even cell death. To overcome these problems, cells express and produce essential, ubiquitous enzymes called topoisomerases,

which can remove knots promptly and efficiently [53, 54]. Contrary to this, it has to be noted that these enzymes also play a role in creating DNA knots. As a result of their presence and dual-functionality, cells have evolved and taken advantage of the topologically constrained nature of their DNA. Lopez and co-workers demonstrated that Topo IV in bacteria can not only form knots in DNA during replication but it is also responsible in unknotting them later on so that DNA can get correctly segregated to every daughter cell [44].

In the case of bacteriophages, recent simulations have revealed that the organization and topology of packaged DNA in capsids are important in how fast the DNA gets ejected into an infected bacterial cell [55]. Marenduzzo and co-workers observed that ordered DNA spools in the capsid, favoured by DNA cholesteric interactions, were ejected at a faster rate than disordered, entangled DNA [55]. It was also shown that torus knots exited the capsid more easily than twist knots, which can halt the ejection process.

#### 3.3. Summary

DNA is an extremely long biological polymer, and it is no surprise therefore that linear and circular, single- and doublestranded DNA molecules are all known to form a wide range of knotted structures from simple trefoil  $(3_1)$  knots to more complex knots such as those with nine crossings. Whereas there are examples of DNA forming both chiral and achiral knots as well as torus and twist knots, there is some evidence, at least in the context of highly packaged viral genomic DNA, that there is a preference for chiral and torus knots. In many cases, it is well established that DNA becomes knotted as a direct result of biological processes such as recombination, replication and transcription. In these cases, knotting is problematic and, consequently, numerous enzymes exist (topoisomerases) which catalyse the unknotting of a DNA chain through a 'cut and paste' mechanism in which the DNA is first cut, then moved/ rotated and subsequently religated. Effectively, this breaks the chain into small segments and rearranges them to eliminate the knot. The biological consequences of not removing the knot can be severe, e.g. cell death. In contrast, there may also



**Figure 5.** (a) Schematic diagram of a trefoil knot, 3<sub>1</sub>, in double-stranded DNA generated using KnotPlot (http://knotplot.com/). (b) Left panel: agarose gel electrophoresis of knotted DNA plasmids, where the mobility increases with the number of knot crossings, reflecting more compact species. Lane 1: unknotted DNA; Lanes 2–7: knotted DNA species. I and II indicate the mobilities of nicked circular and linear DNA, respectively. Right panel: number of crossings in knotted DNA based on electron micrographs of DNA gel bands. Adapted with permission from [9]. (c) Illustration of a site-specific recombination reaction, where arrows indicate the recombination sites. Reprinted from [40], with permission from Elsevier. (d) Schematic representation of the topological consequences of two actively transcribing genes with the origins of replication in convergent orientation. Reprinted from [43], with permission from Elsevier. (e) Schematic diagram of the topological conformation caused by the head-on collision of transcription and replication. Reprinted from [43], with permission from Elsevier. (g) Left panel: knotted DNA from bacteriophage P4 capsids separated by agarose gel electrophorosis. Middle panel: magnified portion, highlighting knot populations of low crossing number. Right panel: Knot populations and subpopulations contain three to nine crossings (labelled 3–9) and six or more crossings (labelled 6'–9'), respectively [7]. Copyright (2005) The National Academy of Sciences, USA.

be benefits of knotting, such as the case of highly packaged viral genomes. Here, knotting may aid in the tight packing and it can also affect the rate at which the genomic DNA is ejected from its viral carrier/storage compartment, the capsid.

DNA can also form a range of other topologically complex states including catenane structures such as Hopf and higherorder links. For a comprehensive overview on the various topological forms of DNA, interested readers are directed towards the following references [8, 37, 56–58].

In addition to all of the studies discussed above on knotting in naturally occurring DNA, there is also considerable literature on knotting in synthetic single-stranded DNA. In particular, Seeman and co-workers have been able to rationally design and build synthetic forms of DNA with a range of knot types and links. A detailed discussion of this work is out of the scope of this review, however, a summary of the different types of structures that have been synthesised is given in table 3, and interested readers are directed to the references provided in the Table.

# 4. RNA

RNA (ribonucleic acid) is a single-stranded, linear polymer made up of four different types of nucleotides that are linked together by phosphodiester bonds. With the help of complementary base pairing and other types of hydrogen bonds between nucleotides in the same chain, RNA molecules can fold into various complex 3D structures and thus achieve diverse biological functions within cells; from mediating the transfer of genetic information from DNA into protein, to catalysis [59, 60]. In addition to these, many viruses have RNA as their genetic material.



**Figure 6.** Formation of an H-type RNA pseudoknot. (a) Linear organisation of the base-pairing elements (indicated with dashed lines) within an H-type RNA pseudoknot. (b) Formation of an initial hairpin prior to pseudoknotting. Bases from the loop are paired to bases outside the hairpin, as indicated with dashed lines. (c) A classical, H-type pseudoknot motif. (d) A ribbon representation of the acceptor arm pseudoknot structure of the 3' end of the turnip yellow mosaic virus genomic RNA is shown based on the NMR structure, PDB code: 1A60. Loops L1 (pink) and L3 (cyan) cross the deep and shallow groove of the helix, respectively. S1 is purple and S2 is blue. L2 is not present in the example shown.

Among the most common RNA structures is the pseudoknot motif, which was first discovered in turnip yellow mosaic virus (TYMV) in 1982 [61]. Although pseudoknots are not true topological knots, they fold into complex 3D conformations where there are a number of topological crossings of the chain. Here, we describe the main structural features of RNA pseudoknots and discuss how they have been intimately linked to the biological properties of naturally occurring RNAs.

#### 4.1. Pseudoknot structures

A pseudoknot is generally defined as an RNA structure that consists of at least two helical segments linked together by single-stranded regions or loops [62]. Although pseudoknots can possess several distinct folding topologies, the best characterised to date is the so-called H (hairpin)-type or classical pseudoknot. As illustrated in figure 6, this is the simplest type of pseudoknot structure that results from the base pairing of a single-stranded segment of RNA in the loop of a hairpin to a complementary sequence outside the loop region. It comprises of two base-paired stem segments (S1 and S2) and, depending on the number of loop bases involved in the pseudoknotting interaction, two or three single-stranded connecting loops (L1, L2 and L3) [63]. However, in most classical pseudoknots (>85% [64]), L2 is missing and thus S1 and S2 can coaxially stack on top of each other to form a quasi-continuous helix. Figure 6(d) depicts this arrangement in the H-type pseudoknot structure of the 3'-terminus of the TYMV RNA, where L1 spans S2 and crosses the deep groove of the helix whilst L3 spans stem S1 and crosses the minor groove. In addition to coaxial stacking, pseudoknots can also be further stabilised by hydrogen bonds formed between single-stranded loop regions and the adjacent stem segments. As the connecting loops and

stems can vary in length, and the interactions between them can differ, RNA pseudoknots represent a structurally diverse group. Hence, it comes as no surprise that these structures are associated with various vital roles in biology. These include forming functional domains within ribozymes [65] and telomerase [66] as well as inducing ribosomal frameshifting in many viruses [10, 67] and regulating translation [68].

#### 4.2. Functional roles of the pseudoknot motif

The RNA pseudoknot is a ubiquitous folding topology that has been identified in almost all organisms [14]. Below, we describe well-characterised examples of pseudoknots involved in catalysis, ribosomal frameshifting and translational regulation, highlighting how the structures are related to their function. In most cases, it has also been shown that the function of pseudoknots is associated with their position along the RNA sequence [63, 69, 70]. For example, pseudoknots located at the core of the tertiary fold of RNAs tend to be crucial in catalysis whilst those found at the 5' end of mRNAs are typically involved in translational control. In addition, in non-coding regions (NCRs) of viral RNAs, pseudoknots play a role in the regulation of initiation of protein synthesis and in template recognition by viral replicases.

4.2.1. Catalytically active pseudoknots. Catalytic RNAs, or ribozymes, are RNA molecules that can catalyse specific biochemical reactions. It has been shown that most ribozymes fold into similar 3D structures that are essential for their function [71]. As a model to understand the mechanism of catalytic RNAs, extensive studies have been done on the hepatitis delta virus (HDV) ribozyme, the fastest known naturally occurring self-cleaving ribozyme [72-74]. HDV is a satellite RNA virus of hepatitis B virus, which together can cause severe infection in humans [75]. The host RNA polymerase II replicates the circular genome of HDV through a double rolling-circle mechanism, producing long RNA transcripts that must be cleaved for viral replication. The processing of the HDV RNA is performed by the self-cleaving HDV ribozyme encoded in the RNA [76]. As illustrated in figure 7(a), the HDV ribozyme has a characteristic 'nested' double pseudoknot that not only forms the active site necessary for the specificity of substrate binding and catalysis but also stabilises the overall RNA structure [77]. This pseudoknot motif has also been discovered in other small self-cleaving ribozymes, particularly in the core of glmS ribozymes in many Gram-positive bacteria [78, 79] and mammalian cytoplasmic polyadenylation element-binding protein 3 ribozymes [80]. As a result, these RNAs are able to achieve an overall complex and stable conformation.

Eukaryotic chromosomes possess telomere ends that protect themselves from loss of genetic material due to successive DNA replication events [81]. Maintenance of the telomeres is performed by the ribonucleoprotein telomerase, an RNAdependent DNA polymerase made up of a specialised reverse transcriptase and a telomerase RNA (TR) [82, 83]. Although telomerase activity is essential for highly proliferative cells such as stem cells, it is also known to be elevated in ~90% of cancer cells [84, 85] and may play a role in aging [86].



**Figure 7.** Sequences and structures of RNA pseudoknots. Loops and stems are colour-coded in reference to figure 6, where L1 is pink, L3 is cyan, S1 is purple and S2 is blue. (a) Hepatitis delta virus ribozyme, PDB code: 1DRZ. For simplicity, only the largest of the two pseudoknots is shown colour-coded. In this example, L2 exists and is shown in red. The grey loop is the U1A RNA binding domain, which is used to aid crystallisation of the ribozyme. (b) Human telomerase, PDB code: 1YMO. (c) Mouse mammary tumour virus, PDB code: 1RNK. (d) Simian retrovirus 1, PDB code: 1E95. (e) The base of domain (dom) III of the Hepatitis C virus internal ribosome entry site, PDB code: 3T4B, where a double pseudoknot (PK1 and PK2) structure surrounding a four-way helical junction is shown. In PK1, L2 (red) and a third base-paired stem, SII/J (orange) exists, in addition to L1, L3, S1 and S2. PK2 is formed between the IIe tetraloop (green) and the main helix of dom III (yellow).

TRs not only provide the template for DNA synthesis but also contain a highly conserved classic H-type pseudoknot within the core domain, which is needed for telomerase assembly and activity [87–90]. Figure 7(b) shows a structure of the human TR pseudoknot, where triple nucleotide interactions U—A-U between L1 and S2 in the deep groove form a triple helix important for telomerase repeat addition processivity [66]. Studies have also shown that the conformational switch that exists between the pseudoknot and a less stable hairpin might be crucial for telomerase activity [91, 92]. Mutations in the TR pseudoknot have also been associated with inherited human disorders such as aplastic anemia and autosomal dyskeratosis congenital [86, 93, 94].

4.2.2. Ribosomal frameshift-inducing pseudoknots. Besides catalysis, RNA pseudoknots are also commonly involved in inducing ribosomes to move into alternative reading frames, a process known as frameshifting. RNA viruses, in particular, exploit the programmed -1 ribosomal frameshifting (-1 PRF) mechanism to regulate gene expression, which enables a single mRNA to get translated into two proteins at a defined ratio [95]. Importantly, this translational mechanism is known to be essential for the replication and proliferation of all retroviruses. Frameshift signals encoded in mRNAs consist of two essential elements: a heptanucleotide 'slippery' sequence X XXY YYZ and a downstream RNA structural element, typically a pseudoknot [96, 97]. It was discovered that even though the slip-site alone can increase frameshifting efficiency by 1%, it is the pseudoknot that is responsible in significantly

stimulating the frameshift event, in some cases, by up to 30–50% [10, 98]. As such, pseudoknot structures in the coding regions associated with frameshifting are potential targets for the development of antiviral therapeutics.

The actual molecular mechanism as to how pseudoknots promote efficient -1 frameshifting still remains unclear. It has been suggested that the downstream pseudoknot structure causes the ribosome to pause on the 'slippery' sequence and forces it to shift back one nucleotide and continue mRNA translation in the -1 reading frame [99]. Studies have shown that this could be due to the unusual topology of the pseudoknot, which makes it resistant to unwinding by the ribosome's helicase activity [100–102].

The first -1 PRF stimulatory RNA element extensively studied in terms of its structure and function was the mouse mammary tumor virus (MMTV) frameshift-inducing pseudoknot [103]. Figure 7(c) shows the NMR structure of the MMTV pseudoknot, which has a characteristic unpaired adenine intercalated between two helical stems rich in guanine/ cytosine. Consequently, this induces a pronounced bend of approximately 60° between the two helices, thus preventing them from being coaxially stacked. Through mutational analysis, structural and functional studies have revealed that the wedged nucleotide and subsequent bending between the helical stems strongly correlate with efficient frameshifting [104]. However, this does not seem to be the case for the simian retrovirus 1 (SRV-1) pseudoknot, where the S1 and S2 helices are coaxially stacked as a result of the base pairing between the adenine nucleotide found in between S1 and

S2 with the last uridine nucleotide in L3 (figure 7(d)) [105]. Instead, subsequent structural studies revealed that favourable interactions between L3 and S1 in the helical junction might be responsible for the frameshifting efficiency in SRV-1.

4.2.3. Pseudoknots involved in translational regulation. Pseudoknot structures have also been shown to regulate translation in viruses and bacteria. In the case of the hepatitis C virus (HCV), its genomic RNA consists of an internal ribosome entry site (IRES) in the 5' untranslated region, where the ribosome is recruited and translation initiated [106, 107]. The HCV IRES is made up of three main structural domains that adopt a tertiary conformation [106, 108]. The core domain of the HCV IRES consists of a four-way helical junction at the base of domain III, where a double pseudoknot is formed (figure 7(e)). The structural integrity of this domain has been found to be essential in positioning the mRNA start codon correctly on the 40S ribosomal subunit during translation initiation [109]. As the pseudoknot domain is highly conserved and is crucial for viral translation, it represents a potential target for HCV therapeutics. Pseudoknots have also been found in the 3' NCR of many viral positive-strand genomic RNA, where they are associated with translational control, replication and genome packaging. Further details of the structurefunction relationship of these 3'-NCR pseudoknots can be found elsewhere [69, 110].

A domain in the bacterial transfer-messenger RNA (tmRNA) has also been shown to consist of four pseudoknots [111]. tmRNAs remarkably possess dual tRNA- and mRNA-like structural and functional properties. They recognise and recycle stalled ribosomes, add a short proteolysis-inducing tag to incomplete growing polypeptide chains and assist degradation of the aberrant mRNAs lacking a stop codon [112]. Although the actual roles of each pseudoknot is still unclear, collectively, they have been suggested to aid in the folding of tmRNA, slow down tmRNA degradation and serve as binding sites for proteins that assist the functioning of tmRNA [68].

# 4.3. Computational prediction of RNA pseudoknots

The function of an RNA molecule can often be inferred from its 3D structure. Since RNA structures are hierarchical and the structural determination of their 3D conformation using experimental methods is difficult, RNA secondary structure prediction is important in elucidating the potential structures and therefore, functions of RNAs. A number of different approaches to RNA pseudoknot structure prediction have been developed over the last decade. These are described below.

Most pseudoknot-free structure prediction programs are based on determining a minimum free-energy (MFE) conformation from the primary nucleotide sequence. However, the prediction of RNA pseudoknots is computationally complex as the search for a MFE structure, in these cases, has been shown to be a Non-deterministic Polynomial-time (NP)-complete problem with respect to sequence length [113]. Dynamic programming (DP)-based methods, which use free energy minimization, can only predict limited classes of pseudoknots. For example, in the case of PKNOTS, the algorithm accurately predicts structures for RNA sequences of length up to 100 bases [114]. Other programs that also use the DP-method include NUPACK [115] and pknotsRG [116]. These approaches, however, are effective only for short sequences, as computation time can increase as the third to sixth power of sequence length, depending on the algorithm used [114, 115, 117].

To overcome this issue, heuristic prediction methods such as FlexStem [118] and HotKnots V2.0 [119] have been developed. Although the predicted structure is not necessarily the MFE, such approaches can handle a wider class of pseudoknots and longer sequences. In another case, the IPknot method, developed by Sato and co-workers, can predict pseudoknotted structures from sequences up to 1000 bases with increased speed and accuracy [120]. Based on integer programming (IP), this method breaks down the pseudoknotted structure into pseudoknot-free substructures and approximates a basepairing probability distribution that considers pseudoknots. In addition, it can also use multiple aligned sequences to predict a consensus pseudoknotted structure [120].

Another algorithm that can predict the MFE RNA pseudoknot structure is TT2NE, which is based on classifying RNA structures according to their genus [121]. Although it can only predict structures for sequences up to 200 bases, it has been shown that the quality of predictions is significantly improved when compared to other state-of-the-art algorithms [121]. Based on the same concept, the same group recently developed McGenus, a Monte Carlo algorithm [122]. Here, the method stochastically searches the MFE structures from sequences of up to 1000 bases. More recently, Jabbari and co-workers have developed an iterative-based method called Iterative HFold, which uses a pseudoknot-free structure to predict pseudoknotted structures rather than a sequence as input [123].

Pseudoknotted structure prediction programs are a valuable resource; examples of some of these recent programs and webservers are listed in table 1. Further details of currently available pseudoknot structure prediction programs can be found elsewhere [124–126]. In general, most of the approaches have been developed with the aim of predicting pseudoknotted structures with increased speed and accuracy. However, it remains clear that these algorithms are still restricted by the lack of understanding of pseudoknot thermodynamics and the capacity to cope with pseudoknots containing stem regions with bulged residues or non-Watson-Crick pairs. In addition, steric constraints and the contribution of entropy to the free energy are often ignored, as there is limited information on the full 3D geometry of pseudoknots. Environmental factors such as ions, solvent, protein and other RNAs are also important in the structure and function of RNA; and ideally these also need to be accurately incorporated into the predictions.

# 4.4. Summary

In contrast to DNA, naturally occurring RNA, strictly speaking, does not form knotted structures. However, it frequently adopts structurally complex conformations in which there are a number of topological crossings of its chain. These structures are known as pseudoknots and are widespread in

Table 1.	Examples	of RNA	pseudoknot	prediction	programmes.
----------	----------	--------	------------	------------	-------------

Programme	Year	Task	URL
Pseudobase [127]	1999	Pseudoknot database	http://pseudobaseplusplus.utep.edu
HotKnots [128]	2005	Pseudoknot prediction from short sequences	www.cs.ubc.ca/labs/beta/Software/HotKnots/
PseudoViewer [129]	2006	Webserver for pseudoknot visualisation	http://pseudoviewer.inha.ac.kr
PknotsRG [116]	2007	MFE pseudoknot prediction from short sequences	http://bibiserv.techfak.uni-bielefeld.de/pknotsrg/welcome.html
McQFold [130]	2008	Pseudoknot prediction by Markov-chain Monte-Carlo (MCMC) sampling	www.cs.uni-frankfurt.de/~metzler/McQFold
ProbKnot [131]	2010	Fast prediction of pseudoknots of any topology	http://rna.urmc.rochester.edu/RNAstructure.html
HotKnots V2.0 [119]	2010	Pseudoknot prediction from short sequences	www.cs.ubc.ca/labs/beta/Software/HotKnots/
IPknot [120]	2011	Pseudoknot prediction from single or aligned sequence(s) with <1000 bases	www.ncrna.org/software/ipknot/ http://rna.naist.jp/ipknot/
TT2NE [121]	2011	Pseudoknot prediction from short sequences ( $\leq 200$ bases)	http://ipht.cea.fr/rna/tt2ne.php
McGenus [122]	2012	Pseudoknot prediction from sequences $\leq 1000$ bases	http://ipht.cea.fr/rna/mcgenus.php
Iterative HFold [123]	2014	Pseudoknot prediction based on an inputted pseudoknot-free structure	www.csubc.ca/~hjabbari/software.php

terms of the different classes of RNA in which they are found. They vary in the length and presence/absence of loop regions and therefore represent a structurally diverse group. It is perhaps, therefore unsurprising that the pseudoknot structure is associated with a range of different biological processes, including catalysis, ribosomal frameshifting and regulation of translation. Although it is not completely understood how their structure results in their specific activities, it is clear that the pseudoknot structure is stable (although it can be in equilibrium with other conformations such as hairpins), may be particularly stable with respect to unwinding by helicases, or degradation. Prediction of the structure of pseudoknots in RNA has rapidly developed over recent years, and, although it is still challenging for very long sequences, a number of different approaches can be used which are increasing in speed and accuracy. Interested readers are directed towards the following references for a more detailed discussion of all of these topics [10, 11, 14, 63, 68, 69]. It is interesting to note that RNA sequences have been designed to form a synthetic trefoil knot [132], see Discussion for further details.

# 5. Proteins

Proteins are linear biopolymers composed of different aminoacid residues covalently linked together by peptide bonds. They play a crucial role in almost all biological processes including cell signalling, catalysing metabolic reactions and structural support. In order to perform their function, most proteins have to fold to a compact 3D structure (native state), which is ultimately dictated by its unique amino-acid sequence.

Many thousands of proteins with a diverse array of structures and functions are known. Due to their structural variation and complexity, proteins have been shown to possess a wide range of intricate topological features (figure 8). Intermolecular non-covalent interactions can lead to interlocked, oligomeric rings of protein subunits, where the two rings form a Hopf link and therefore become inseparable (figure 8(a)) [133]. In other cases, covalent bonding such as disulphide bonds or metal-side chain interactions can also result in covalent links or knots formed either during or after folding. Figure 8(b) illustrates a Hopf link structure formed as a result of intra-molecular disulphide bonds within each subunit of a dimeric protein [134]. In addition, the recently discovered pierced lasso bundle (PLB) topology is an example of a knot-like motif where the disulphide bond creates a covalent loop through which part of the polypeptide chain is threaded (figure 8(c)) [135]. 'Cysteine knots' can form when a disulphide bond between two segments of a polypeptide chain pass through a ring formed by two other disulphide bonds and their connecting backbone segments (figure 8(d)). Examples include the cyclotide family of naturally occurring plant-based miniproteins and the superfamily of growth factors and toxins [136–138]. In all of these cases, the link or knot is created by a covalent bond or oligomeric structure.

Complex topologies such as linking or knotting can also be manifested within the protein backbone chain itself. Figure 8(e) illustrates an example of a class of proteins that possess a knotted topological feature in their structures formed by the path of the polypeptide backbone alone [13, 15, 139]. In another case, protein slipknot structures also arise when a protein chain forms a knot but then folds back upon



**Figure 8.** Different types of topologically complex protein structures. In each panel, the protein structure produced using Pymol (www. pymol.org/) is shown on the left, with a simplified representation of the topology of the system on the right. (a) The crystal structure of bovine mitochondrial peroxiredoxin III forms a Hopf link, PDB code: 1ZYE. In the simplified representation, the blue and red filled circles represent a single chain subunit which associate together to form a higher-order oligomeric ring structure. (b) *P. aerophilum* dimeric citrate synthase is topologically linked by two intramolecular disulphide bonds (black bars), PDB code: 2IBP. Each protein chain is coloured separately, in this case, blue or teal. (c) A pierced lasso bundle topology of the native structure of leptin, where a disulphide bridge (black bars) creates a covalent loop through which part of the polypeptide chain is threaded, PDB code: 1AX8. (d) The crystal structure of nerve growth factor contains a cysteine knot motif defined by three disulphide bonds (black bars), PDB code: 1BET. (e) The polypeptide backbone chain of *E. coli* methyltransferase YbeA contains a trefoil knot (3<sub>1</sub>), PDB code: 1NS5. (f) The crystal structure of human phosphatase has a slipknotted topology, PDB code: 1EW2. For (c)–(f), both structures and reduced representations are coloured from blue (N-terminus) to red (C-terminus). Cysteine residues in (b)–(d) are represented as sticks and lines in the structure and simplified representation, respectively.

itself to completely untie the knot, thus rendering the structure unknotted when considered in its entirety (figure 8(f)) [140–142]. This section of the review focuses on the structure, function and, in particular, the folding of these types of knotted and slipknotted proteins. Proteins that have knots formed by covalent bonds such as disulphides are not discussed here and readers who are interested in these structures are directed to other publications on these systems [136, 137, 143–146].

### 5.1. Knotted and slipknotted proteins

For a long time, it was thought that it was highly unlikely, if not impossible, for a polypeptide chain to 'knot' itself to form a functional folded protein. This was, in part, due to the fact that, at that time, no examples of deeply knotted proteins were identified within the protein data bank (PDB) [147]. In this study, a very shallow knot was discovered in carbonic anhydrase by Mansfield [147]. One of the challenges in the search for protein knots was the difficulty in determining whether a knot is present within a complex structure. Thus, for many years, knots in protein structures went undetected. As various computational and mathematical tools were developed to detect and identify knots, it became clear that topologically knotted protein structures do exist, even some with extremely deep knots [24, 26, 148, 149]. Now there are a few web-servers that have simplified the task of knot identification in proteins and can determine quickly whether a structure contains a knot and, if so, what type [150, 151]. In addition, the recent KnotProt database (http://knotprot.cent.uw.edu.pl/) created by Sulkowska and co-workers classifies knotted proteins and represents their knotting complexity (knot type and depth of knot) as a 'knotting fingerprint' in the form of a matrix diagram [142, 152, 153]. Matrix diagrams, which are an excellent method for visualising knots and slipknots in proteins, were originally used in the analysis of slipknots in proteins by the Yeates group [140].

To date, over 750 knotted proteins have been discovered within the PDB, equivalent to approximately 1% of all entries [152]. A current list of examples of these structures is provided in table 2. It is worth noting that the KnotProt database is updated regularly [152]. Over the years, a growing number of knotted proteins have been observed in all three domains of life [15, 142, 154]. These include structures that contain a trefoil (3<sub>1</sub>), figure-of-eight (4<sub>1</sub>), Gordian (5<sub>2</sub>) and stevedore (6<sub>1</sub>) knot with three, four, five and six projected crossings of the polypeptide backbone, respectively (figure 9).

**Table 2.** Examples of knotted and slipknotted proteins. For each fold, the PDB code for the structure of the protein or a typical protein in the family is given. + and - indicates right and left-handed knots and slipknots, respectively.

Protein family or Protein	PDB code	Knot type
RNA methyltransferase ( $\alpha/\beta$ knot)	1NS5	$3_1 + knot$
Carbonic anhydrase	1LUG	$3_1 + knot$
SAM synthetase	1FUG	$3_1 + knot$
Transcarbamylase fold	1JS1	$3_1 + knot$
Sodium/calcium exchanger membrane protein	3V5S	$3_1 + knot$
Zinc-finger fold	2K0A	$3_1 - knot$
Ribbon-helix-helix superfamily	2EFV	$3_1 - knot$
Artificially knotted protein	3MLG	$3_1 - knot$
Class II ketol acid reductoisomerase	1YVE	4 <sub>1</sub> knot
Chromophore binding domain of phytochrome	2O9C	4 <sub>1</sub> knot
Ubiquitin C-terminal hydrolases (UCHs)	2ETL	$5_2 - knot$
$\alpha$ -haloacid dehalogenase I	3BJX	$6_1 + knot$
Alkaline phosphatase	1ALK	$3_1$ + slipknot
Thymidine kinase	1P6X	$3_1$ + slipknot
Glutamate symport protein	2NWL	$3_1$ + slipknot
Sulfatase	4TN0	$3_1$ + slipknot
STIV B116	2J85	$3_1$ + slipknot
Apoptosis inducing factor	1GV4	3 <sub>1</sub> – slipknot
Sodium:neurotransmitter symporter family	2A65	$3_1 + \& 4_1$ slipknot
Betaine/Carnitine/Choline Transporter (BCCT) family	4AIN	$3_1 + \& 4_1$ slipknot



**Figure 9.** Structures of knotted proteins that contain the four different types of knots  $(3_1, 4_1, 5_2, 6_1)$  in the polypeptide backbone. (a) YbeA, a trefoil-knotted  $(3_1)$  methyltransferase from *E. coli*, PDB code: 1NS5. (b) *E. coli* class II ketol-acid reductoisomerase, containing the figure-of-eight  $(4_1)$  knot, PDB code: 1YRL. (c) Human ubiquitin carboxy-terminal hydrolase L1 (UCH-L1), containing a knot with five projected crossings  $(5_2)$ , PDB code: 2ETL. (d)  $\alpha$ -haloacid dehalogenase containing a stevedore  $(6_1)$  knot, PDB code: 4N2X. Top panel: ribbon diagrams of the polypeptide chains produced using Pymol (www.pymol.org/). Lower panel: simplified view of the protein chain showing the knot, generated using KnotPlot (http://knotplot.com/). Both structures and reduced representations are coloured from blue (N-terminus) to red (C-terminus).

Trefoil knots are the most prevalent and simplest type of knot discovered in proteins. The first protein trefoil knot to be identified was that found in carbonic anhydrase—a family of proteins involved in catalysing the reaction of carbon dioxide to hydrogen carbonate and H<sup>+</sup> [147]. This trefoil, however, is rather shallow as the C-terminus extends through a wide loop by only a few residues. A few years after Mansfield's 1994 study, a much deeper trefoil knot was detected in *E. coli* 

S-adenosylmethionine synthetase, an enzyme that catalyses the reaction between methionine and ATP [155, 156]. By far, the largest and most well-studied family of deeply knotted proteins is the trefoil  $\alpha/\beta$  knot fold—a class of methyltransferases (MTases) which are members of the SpoU family [157, 158]. These knotted proteins share common structural features and it is highly likely that all are MTases that catalyse the transfer of the methyl group of S-adenosyl methionine (AdoMet) to carbon, nitrogen or oxygen atoms in DNA, RNA, proteins and other small molecules [159]. In solution, all form dimers with the knotted region comprising part of the AdoMet binding site and forming a large part of the dimer interface [157, 160–163]. Trefoil knots have also been found in two homologues of N-succinylornithine transcarbamylase; the AOTCase from *X. campestris* catalyses the reaction from N-acetylornithine and carbamyl phosphate to acetylcitrulline [164], and SOTCase from *B. fragilis* promotes the carbamylation of N-succinylornithine [165]. Besides being found in enzymes, trefoil knots have also been identified in Rds3p, a eukaryotic metal-binding protein essential for pre-mRNA splicing [166] and more recently, in the family of sodium/calcium exchanger membrane proteins [152].

More complex knots have also been identified in proteins that catalyse various enzymatic reactions. A deeply embedded, figure-of-eight protein knot has been found in plant ketol-acid reductoisomerases, which are involved in the biosynthesis of branched-chain amino acids [167, 168]. In addition, a Gordian knot has been identified in the family of mammalian ubiquitin carboxyl-terminal hydrolases (UCHs); the proteins are deubiquitinating enzymes that catalyse the cleavage of the isopeptide bond formed between ubiquitin and lysine side chains of protein and other adducts, and thus are involved in the ubiquitinproteasome system [169–171]. The most complex protein knot known to date is the 61 stevedore knot discovered in DehI, a  $\alpha$ -haloacid dehalogenase that catalyses the removal of halides from organic haloacids [154]. Apart from these enzymes, it has been shown that the figure-of-eight knot also exists in the chromophore-binding domain of a red/far-red photoreceptor phytochrome from bacterium *D. radiodurans* [172, 173].

Slipknotted structures have also been found in a number of proteins (figure 8(f)) [140]. They cannot be identified using the standard methods for knot detection in proteins as, in these cases, the knot becomes undone when the chain is pulled at both termini. As such, it comes as no surprise that these structures had been overlooked until relatively recently. In 2007, Yeates and co-workers first discovered a number of protein slipknots by using an approach based on the fact that slipknots become real knots at some point when the polypeptide chains are shortened [140]. At present, over 450 protein slipknots have been identified [152] and a list of examples of these structures is listed in table 2. It is worth noting that the KnotProt database is the first, and currently only, database that provides details on slipknotted structures [152].

Alkaline phosphatase is the largest family of proteins that contain deep slipknots [15, 140, 152]. In the case of *E. coli* alkaline phosphatase, 30 residues have to be deleted from the C-terminus before a knotted conformation results. Similar to that of knotted proteins, many of the protein slipknots discovered to date are also found in other enzymes such as thymidine kinases and sulfatases [15, 140, 152]. Interestingly, slipknots have also been found in transmembrane proteins that span the entire cell membrane to which they are permanently embedded [15, 140, 152]. Examples include the families of sodium:neurotransmitter transporters, betaine/carnitine/choline transporters (BCCT) and proton:glutamate transporters [142].

Further details of knotted and slipknotted protein structures can be found in other recent reviews [12, 13, 15, 174] and the KnotProt server [152]. It should be noted that the KnotProt database also provides extensive key information about the biological functions of proteins with knots and slipknots [152].

## 5.2. Potential roles and implications of the knot and slipknot

Topologically knotted proteins have been found to be conserved across different families [142], suggesting that the knot itself may be advantageous and important to the function of the protein. It has been speculated that a knotted topology could play a key role in increasing catalytic activity or ligand binding affinity (potentially by decreasing dynamics) or enhancing stability (thermodynamic, kinetic and mechanical) of a protein. As yet, relatively little is known about the functional advantages, if any, of these complex knotted structures over their unknotted counterparts. However, various experimental and computational studies have been undertaken to address this question.

Many reports have shown that the knotted regions of knotted proteins play crucial roles in enzymatic activities and ligand binding. As discussed in section 5.1., it has been observed that the knotted regions of the proteins in the  $\alpha/\beta$ knotted SpoU MTase family comprise part of the active site to which the ligand binds (two examples of  $\alpha/\beta$  knot MTases are illustrated in figure 10(a) [159–162]. In the case of the N-succinylornithine transcarbamylase, Virnau and co-workers have demonstrated through a computational study that the presence of the knot in the knotted homologue AOTCase may structurally modify its active site and subsequently, may alter its enzymatic activity (in terms of substrate specificity) compared to its unknotted homologue OTCase (figure 10 (b)) [149]. In addition, structural studies of the D. radiodurans phytochrome revealed that the deeply embedded knot in the chromophore-binding domain is in contact with the chromophore [172, 173]. A recent study on the conservation of knotting fingerprints in UCHs also showed that there was a correlation between the locations of active site residues and points characterising its knotted topology (i.e. the knotted core) [142]. Despite these examples, there is still little direct experimental evidence that a knotted structure can influence the activity of a protein.

The question of whether knots have any effect on the conformational dynamics of proteins has also been raised. In the phytochrome protein, it has been noted that the figure-ofeight knot sits where increased rigidity could be important in driving conformational changes that occur when light energy is absorbed by the chromophore [172, 175]. Recent computational approaches using simple lattice models have shown a narrow and less extended native basin for a 5<sub>2</sub>-knotted structure relative to a similar but unknotted one, suggesting enhanced rigidity [176]. However, experimental studies by Andersson *et al*, which measured <sup>15</sup>N spin relaxation parameters using NMR experiments for the 5<sub>2</sub>-knotted UCH-L1, reported no significant differences between the relaxation properties of the knotted protein relative to unknotted proteins of a similar size



**Figure 10.** Examples highlighting the potential roles of knots and slipknots. (a) Dimeric structures of the  $\alpha/\beta$ -knot MTases YibK, PDB code: 1MXI (left) and YbeA, PDB code: 1NS5 (right), coloured to show the knotting loop in cyan and the knotted chain in red. S-adenosyl homocysteine, an MTase co-factor, is shown as a stick model. (b) Structures of the knotted section (residues 171–278) of AOTCase with the reaction product N-acetylcitrulline and interacting side chains represented as sticks, PDB code: 3KZK (left), and corresponding (unknotted) section (residues 189–286) in OTCase with the inhibitor L-norvaline (analogous to its L-ornithine ligand) and interacting side chains shown as sticks, PDB code: 1C9Y (right). The knot containsf a rigid proline-rich loop (residues 178–185, coloured red) through which the chain is threaded. (c) Left panel: engineered knotted and unknotted ('superficially knotted') polymers using two different protein constructs. Right panel: first derivative melting curves obtained for the knotted and unknotted polymers. Adapted from [179], by permission of Oxford University Press. (d) Structures of transmembrane proteins LeuT(Aa), PDB code: 2A65 and Glt(Ph), PDB code: 2NWL, where the slipknot loop is coloured cyan and the slipknotted chain in red. Helices are represented as cylinders to ease visualisation. All structures are produced using Pymol (www.pymol.org/).

[177]. Thus, it remains to be clearly established, particularly experimentally, whether knotted structures can influence the conformational dynamics of a protein.

Much research effort has been undertaken to address the question of whether a knot can provide additional thermodynamic, kinetic or mechanical stability to a protein structure. Sulkowska *et al* performed coarse-grained simulations of the thermal and mechanical unfolding of the knotted (AOTCase) and unknotted (OTCase) variants of the transcarbamylase-like proteins as well as a synthetic construct of the knotted parent protein rewired so as to remove the knot [178]. In this case, the knotted structure was found to have longer unfolding times than the other two unknotted proteins, which were attributed to topological and geometrical frustration [178]. In an attempt to investigate the potential thermal stabilities of knotted proteins in an experimental study, Yeates and coworkers engineered a knotted and an unknotted ('superficially knotted') polymer [179]. They showed that the knotted chain had a higher thermal stability than the unknotted one (figure 10(c)), although it is important to note that the unfolding in both cases was not fully reversible and therefore only apparent melting temperatures were reported. However, computational studies using Monte Carlo simulations of a simple lattice model using Gō-like potentials showed that a trefoil knot did not have any effect on the thermodynamic stability of a simple protein structure [180]. Instead, it was found that the knot enhances kinetic stability as the knotted protein unfolds at a distinctively slower rate than its unknotted counterpart [180]. Further studies by the same group demonstrated that a more topologically complex protein knot, the  $5_2$  knot, clearly enhanced the protein's kinetic stability in comparison to that of a protein containing a  $3_1$  knot [176].

The resistance of knotted proteins to mechanical unfolding has been examined by atomic force microscopy (AFM). The first system to be studied was the shallow trefoil-knotted carbonic anhydrase B. In this particular case, an extremely high resistance to unfolding was observed when the protein was pulled from its termini in contrast to a considerably lower resistance when the molecule was pulled from other positions resulting in the untying of the knot [181, 182]. Although these initial studies suggested a dramatic effect of a knot on mechanical stability, the results have not been observed in AFM studies of other knotted systems [175]. In the case of carbonic anhydrase B, recent simulations have shed light on the possible reasons for its remarkable mechanostability [183]. These studies revealed that after an initial, rather limited unfolding event, the knot is wrapped around an inner  $\beta$ -sheet structure in the core of the protein. Thus, the knot is tightened but effectively locally captured by a structural obstacle in the chain. This is aided by the stabilising effects of a zinc ion, which coordinates to the region that becomes entangled by the knot. The simulations explain why in the AFM experiments, the contour length observed is so much smaller than that expected for a fully stretched polypeptide chain containing a tightened knot. In an interesting extension of their initial work, Ikai and co-workers made a tandem repeat of carbonic anhydrase B. Combining AFM with biochemical measurements of activity and binding, they were able to establish that the C-terminal knotted region was essential for activity [184].

The mechanical stability of the  $4_1$ -knotted phytochrome protein has also been investigated by Bornschögl and coworkers using AFM [175]. In this case, however, they did not observe any enhanced resistance when the knot was tightened as the extension force for unfolding (73 pN) was within the range found for other unknotted proteins. It appears that whether a knot contributes to mechanical stability or not, may depend upon a number of factors including other aspects of the protein's structure and potentially pulling speed/force etc. Several computational studies have suggested that knotting might increase a knotted protein's mechanical stability, thus making it more resistant to cellular translocation and degradation pathways [149, 178, 185, 186]. Again, whether knotting confers any advantageous stabilising effect to a knotted protein over its unknotted counterpart is still inconclusive and thus remains to be tested with more experimental and computational studies.

The significant number of protein slipknots that have now been identified has also posed the question of whether such topologies have any functional or structural role in the protein. In the case of the homodimeric E. coli alkaline phosphatase, Yeates and co-workers engineered cysteine residues at various positions in the protruding loop of the slipknot such that inter-molecular disulphide bonding between the two subunits resulted in a knotted system [140]. Using thermal denaturation, the results showed that the knotted mutants were more thermally stable than either the wild-type or other control mutants. This suggested that the slipknot in the structure may play a role in the enzyme's thermostability [140]. It is also worth noting that the slipknotted B116-like protein is found in a virus that infects thermophilic Sulfolobus archaebacteria [140]. In another study, knotting fingerprint analyses of transmembrane transporting channels from five different families of proteins showed that the slipknotted topology is conserved. This has led to speculations that the slipknot loop, which straps together several transmembrane  $\alpha$ -helices, may stabilise their location inside the membrane during their transporter and symporter action [142] (see figure 10(d) for examples of the structures of two slipknotted transmembrane proteins).

# 5.3. Experimental and computational insights into how knotted and slipknotted proteins fold

The study of how proteins achieve their unique 3D conformation (native state) has been the focus of many researchers in the field of protein folding. For many decades, extensive folding studies focussed on small, monomeric proteins and thus mechanisms of how they fold are now relatively well established [187–191]. These include the framework, nucleation-condensation and hydrophobic collapse mechanisms, which can be viewed as points on a spectrum of a unified mechanism [187, 188]. Current folding theories have shown that small, monomeric proteins, which fold efficiently and rapidly, can achieve their low-energy native configuration from an ensemble of denatured polypeptide chains in a highly cooperative manner and traverse relatively smooth, funneled energy landscapes [192, 193]. However, it is still unclear how these concepts and mechanisms are applicable to larger proteins with more complex topologies including the classes of knotted and slipknotted proteins. Not only do such proteins have to avoid kinetic traps but they also have to overcome significant topological barriers during folding. This section summarises recent developments made towards understanding the mechanisms involved in the formation of these types of complex structures.

5.3.1 Experimental studies on knotted proteins. Although the elucidation of how knotted proteins fold using experimental approaches remains challenging, in recent years, some significant progress has been made. Most of the experimental folding studies on knotted proteins have focussed on the trefoil-knotted  $\alpha/\beta$  MTases, YibK from *H. influenzae* and YbeA from *E. coli* [194–201]. Both proteins are homodimers, which bind to the co-factors AdoMet and S-adenosyl homocysteine (AdoHcy) and contain a trefoil knot at the C-terminus in which at least 40 residues pass through a similarly sized loop (figure 10(a)) [160, 202]. Extensive biophysical techniques have been employed to probe the knotting and folding mechanisms of purified, recombinantly expressed YibK and YbeA. Both unfold reversibly *in vitro* upon addition of chemical denaturant with a concomitant loss of secondary and tertiary structure [195, 198]. Kinetic studies demonstrated that YibK and YbeA fold similarly via sequential mechanisms that involved one or more monomeric intermediate states and a slow rate-limiting dimerization step [196, 198].

To probe chain knotting events during the folding of YibK and YbeA, Mallam and co-workers constructed a set of knotted fusion proteins in which A. fulgidus This, a stable 91-residue protein, was fused to the N-, C- or both termini of both MTases [201]. This was used as a 'molecular plug' in an attempt to disrupt threading events or to prevent the chain from knotting altogether. Remarkably, these experiments established that both proteins can withstand the fusion of additional domains to both their N- and C-termini and are able to fold to native or native-like states capable of binding cofactor. The fusion proteins created in this study represent some of the most deeply knotted proteins known, the C-terminal fusions requiring some 140 or more residues to pass through a loop to form the knotted native state. Surprisingly, all the fusion proteins showed unfolding and refolding kinetics very similar to the parent MTase giving the first hint that the polypeptide chain might remain knotted even in a highly unstructured chemically denatured state. This was subsequently shown to be the case through in vitro folding experiments on circularized variants of YibK and YbeA, Mallam and co-workers discovered that the denatured ensembles, even in high concentrations of chemical denaturant under which conditions there was little or no secondary or tertiary structure, contained kinetically trapped knotted polypeptide chains [194]. It was then concluded that all the previous in vitro folding experiments on these recombinantly expressed and chemically denatured proteins actually probed refolding from an unfolded but knotted denatured state to a knotted and folded native structure. This unexpected result suggests that there are interactions in the denatured state that kinetically stabilize the knot. Although far-UV CD measurements indicate that there is no significant secondary structure present in the denatured state, recent backbone NMR assignments and chemical shifts of urea-denatured YbeA, show that, in fact, some residual secondary structure still remains under these conditions [203]. The fact that the knot can persist in the denatured state over a long period of time was also confirmed by another group who shared that equilibrium unfolding and refolding transitions of a structurally homologous MTase displayed apparent hysteresis [204]. This behaviour was speculated to be consistent with the uncoupling of the unfolding and untying events of the knotted protein [204]. Recently, single-molecule fluorescence resonance energy transfer (FRET) experiments were performed to characterise the denatured state of TrmD, another trefoil-knotted MTase [205]. Results suggested that the knot was not only retained under denaturing conditions (similar to that of YibK and YbeA) but also slid towards the C-terminus of the polypeptide chain during the unfolding process [205].

Up until recently, there have been no experimental studies into how the knot is first formed from an unknotted linear polypeptide chain. However, with the use of a coupled in vitro transcription-translation system and kinetic pulse-proteolysis experiments, Mallam and Jackson were able to specifically probe folding of nascent chains of YibK and YbeA after they were first synthesised by the ribosome (figure 11(a)) [199]. The results showed that the nascent chains could fold correctly to their trefoil-knotted structure, albeit very slowly. Moreover, a significant lag period between chain synthesis and emergence of a proteolytically stable native state was observed. The results were consistent with the protein knotting and folding from an initially unknotted nascent chain, thus demonstrating that a process associated with the knotting step is rate limiting. Additionally, the GroEL-GroES chaperonin was found to have a dramatic effect on the folding rate of the newly translated polypeptide chains, thus establishing that chaperonins are likely to be important in the post-translational folding of these bacterial knotted proteins in vivo.

Very recently, we have investigated the knotting and folding behaviour of the nascent chains of the different Nand C-terminal This fusions of YibK and YbeA with the use of the coupled in vitro transcription-translation system and kinetic pulse-proteolysis experiments [206]. The results demonstrated that these multi-domain proteins with extremely deep knots can be synthesized in vitro and spontaneously knot without the help of any molecular chaperones, albeit very slowly. In addition, it was concluded that the C-terminus of these proteins is critical to the threading of the polypeptide chain to form the knot, thus providing the first experimental insight as to the mechanism of knotting for this class of bacterial knotted MTase. Further experiments with the GroEL-GroES chaperonin demonstrated that it actively assists the folding of knotted proteins by a mechanism that may involve the unfolding of kinetically trapped unknotted and misfolded intermediates (figure 11(b)). These key observations provide not only vital information into the complex folding pathway of trefoil-knotted proteins but also further insights into how topologically knotted proteins have withstood evolutionary pressures and achieve efficient folding in vivo.

In 2010, the Yeates group engineered an artificially trefoilknotted protein by covalently linking together two monomers intertwined in the dimeric structure of HP0242 from *H. pylori* [207]. An *in vitro* experimental characterisation of this designed knotted protein and an unknotted monomeric variant of the HP0242 dimer was undertaken. Results showed that, although the knotted variant was more stable than the unknotted one, it folded at a considerably slower rate (approximately 20-fold), indicating that knotting, or some event associated with it, is likely rate-limiting.

AFM has also been used to study the mechanical unfolding of the shallow trefoil-knotted carbonic anhydrase B. In this case, the polypeptide chain was found to extend to a distance much shorter than its theoretical stretching length, indicating that the knotted structure is tightened but retained [182, 208]. Similarly, AFM mechanical unfolding experiments on the figure-of-eight knot in the chromophore-binding domain of the phytochrome also resulted in a tightened knot



**Figure 11.** Experimental characterisation of the folding of the trefoil-knotted methyltransferases, YibK and YbeA. (a) A schematic representation of the folding and knotting pathways that have been experimentally observed. (b) A schematic diagram illustrating a possible active mechanism for the bacterial GroEL-GroES chaperonin action on the folding of bacterial trefoil-knotted methyltransferase. D, denatured; I, intermediate; N, native.

of approximately 17 residues [175]. Although these experiments do not necessarily provide extensive information on the folding pathways of these proteins, they were critical in demonstrating that the knots were present in the structure and in determining the minimum length of polypeptide chain required for knotting.

In addition to the trefoil-knotted proteins described in detail above, the other family of knotted proteins for which there has been any substantial experimental characterisation of their folding pathways are the  $5_2$ -knotted UCHs [177, 209]. The unfolding of two human UCHs- UCH-L1, a neuronal form of the enzyme, and UCH-L3, ubiquitously expressed in many cell types, have been determined and, in both cases, the in vitro unfolding/refolding with chemical denaturants was shown to be fully reversible [177, 209]. In the case of UCH-L3, equilibrium unfolding data were fitted to a simple two-state model [209] whilst that for UCH-L1 were consistent with a three-state model in which an intermediate state is populated [177]. Using NMR hydrogen-deuterium exchange (HDX) experiments, the intermediate state was characterised indirectly and it was found that the central  $\beta$ -sheet core of the protein remains structured whilst many of the surrounding  $\alpha$ -helices have unfolded [177]. Although a more complete analysis of the folding pathway of UCH-L1 has yet to be published, the folding is similar to UCH-L3, such that, both have multiple unfolding and refolding phases that indicate parallel pathways and the population of at least two, metastable intermediate states (Luo et al unpublished results).

5.3.2. Computational studies on knotted proteins. Many computational studies have shed considerable light on the folding of knotted proteins. Coarse-grained simulations have been excellent at revealing the possible mechanism(s) and generic features of how knotted proteins fold [210, 211]. Wallin *et al* performed the first such simulation using a  $C_{\alpha}$  model representation of YibK and, similar to experimental studies, observed two parallel folding pathways [210]. They also concluded that specific, non-native interactions involving residues in the C-terminal region of the chain were needed for the protein to knot and fold successfully. In contrast, Sulkowska and co-workers showed that native interactions alone are sufficient for simulating the folding of YibK and YbeA using a coarsegrained structure-based model, although the number of successful trajectories was only 1-2% [211]. These simulations also illustrated that partial unfolding (backtracking) events were needed because the order in which native contacts are formed is critical for the correct folding of the knotted structure and that folding frequently occurred through a slipknotted intermediate (figure 12(a)). Importantly, in the same study, simulations of a rewired, unknotted variant established that there are significant topological barriers in the folding of the knotted structure [211]. Using a similar model, initial results from recent kinetic unfolding simulations of a structurally homologous MTase revealed that unfolding of the protein to a fully unfolded, unknotted state occurs in a stepwise process [204]. In addition, the simulations showed that unknotting of the chain is slow compared to the initial unfolding [199].



**Figure 12.** Computational simulations of the folding pathways of knotted proteins. (a) Structure-based model used to simulate the folding of trefoil-knotted MTase where the folding route that leads to the native knotted conformation occurs through an intermediate 'slipknot' configuration. Incorrect configurations have to use a 'backtracking' mechanism in order to escape kinetic traps which act as topological barriers. Adapted from [211]. (b) Snapshots taken from the folding simulation of the 6<sub>1</sub>-knotted protein, DehI. Copyright 2010 Bölinger *et al* [154]. (c) An all-atom structure-based molecular dynamics simulation of the folding pathway of MJ0366. The protein forms a loop with the correct chirality (I), from which it follows two routes to the native state (N): a 'plugging' or 'slipknotting' route. T is an example of how the protein may be kinetically trapped and thus unable to proceed to N. Adapted from [141]. (d) Schematic representations of pulling a trefoil-knotted protein in different points (indicated by the circles) and their resulting final conformations.

Similar computational approaches were also employed in the folding simulations of the  $6_1$ -knot in DehI [154]. Although the probability of successful folds was low, the study revealed that the complex knotted structure can be formed by a simple tying process. In this case, two unknotted loops, a small loop and a larger loop (which includes a proline-rich unstructured region) are aligned and a knot can be formed by two alternative routes (figure 12(b)) [154]. In the first route, the C-terminus is threaded through the smaller loop (S-loop) via a slipknot conformation before the larger loop (B-loop) flips over the smaller loop. In the other route, the order of the two steps is reversed.

In contrast to very small proteins with simple architectures (which generally have fast unfolding and folding rates), allatom molecular dynamics (MD) simulations have not been extensively applied to knotted systems, as they are frequently too large for such atomistic approaches to be used. However, it has been possible to use this method in a few cases on small, shallow knotted proteins, such as for MJ0366 from *M. jannaschii*, one of the smallest trefoil-knotted protein discovered to date [141]. Data from a thermodynamic analysis of the unfolding/folding revealed that the system is three-state, and an intermediate is first formed by twisting of a loop, followed by a rate-limiting step associated with the threading of the C-terminus through the loop. At temperatures near the folding temperature, two folding mechanisms were observed for the formation of the knotted native structure, whereby threading can occur via (i) a plugging route (the C-terminus goes through the knotting loop first) or (ii) the formation of a slipknot (figure 12(c)) [141]. Interestingly, lowering the temperature of the simulation resulted in mechanistic changes. These include a knotting via threading of the N-terminus and the 'backtracking' of misfolded proteins in topological traps. More recently, simulations on VirC2, a protein that has the same fold as MJ0366 but which possesses a deeper knot, also showed that it has a similar free energy profile, suggesting that topology plays a major role in the folding mechanism [212]. A Gō-like potential in which there is minimal energy frustration was also used to simulate the folding of a truncated mutant of another trefoil-knotted MTase [213]. Results from this study suggested a pathway in which the N-terminal region of the protein folds first and that threading of the C-terminus through the structure to form the knot is a late and rate-limiting step [213].

Molecular dynamics simulations were also used to simulate the high temperature unfolding of YibK [214]. The simulations revealed up to four intermediate states on the free energy landscape consistent with the parallel pathways and multiple intermediates observed in experimental studies. In addition, it was found that the denatured state of YibK only untied at very high simulation temperatures, when the C-terminus threads out of the knotting loop via a slipknot conformation. Other unfolding simulations have also been used to investigate the mechanical stability of knotted proteins and the effect of pulling position, pulling speed and temperature on the unfolding/untying of two other MTases [215]. It was shown that pulling the chain at both termini leads to the tightening of the knot whilst pulling at other positions can result in the unknotting of the chain (figure 12(d)).

Various computational studies have also employed Monte Carlo simulations on lattice models using Go-like potentials to understand the folding mechanism of knotted proteins. In these cases, a potential based on a generic polymer model is used and additional attractive interactions are included for residues that are in contact with each other in the native state. Faisca and co-workers demonstrated that the folding of a model deeply knotted trefoil protein was much slower than a structurally similar but unknotted variant, and that knotting was a late event and concomitant with folding [216]. Using the same model, Soler and Faisca examined the effect of surface tethering on the folding of the system [217]. In this case, it was shown that the mobility of the terminus closest to the knot is critical for successful folding and hindrance results in a decrease in the folding rate and a change in the knotting pathway such that it involves threading of the other terminus. Recently, the same group extended these studies and used the same model to investigate in further detail the effect of knots, knot depth and motif on folding properties of 31-knotted proteins [180]. The results revealed that deeply knotted proteins have a higher probability of retaining their knots in the denatured ensemble, consistent with experimental studies. Furthermore, it was shown that specific native contacts within the trefoil-knotted core are crucial in maintaining the knot in the denatured state, and that threading occurs in the late stages of folding [180]. Most recently, Soler and co-workers extended their studies to investigate the folding mechanism of the more complex  $5_2$ -knot [176]. Similar to the trefoil knots, it was shown that the chain terminus that is closest to the knotted core is important for the threading movement to form the knot and in no cases was a mechanism that involved the initial formation of a 31-knot observed. However, it was discovered that the probability of concomitant knotting and folding of 52-knotted proteins is significantly smaller than that for trefoil knots as threading to form the 52 knot is a particularly late conformational event [176].

Monte Carlo simulations of a  $C_{\alpha}$  model of trefoil-knotted AOTCase showed that non-native contacts between the C-terminus and other regions in the protein are critical to form the knotting loop through which the chain is threaded [218], consistent with the study by Wallin and co-workers [210]. The importance of non-native interactions in promoting the folding of the native knotted topology of AOTCase and MJ036 was also recently highlighted in simulations employing protein models with different structural resolution (coarse-grained or atomistic) and various force fields (from pure native-centric to realistic atomistic ones) [219]. Again, it appears that these contacts were found to be between the C-terminus and a loop, through which the chain is threaded.

5.3.3. Experimental and computational studies on slipknots. Numerous simulation studies have shown that a slipknot may be an important intermediate configuration in the folding of knotted proteins [141, 142, 211, 212] and thus, understanding the mechanisms involved in their formation could offer insights into how deeply knotted proteins fold. Using structure-based coarse-grained simulations, Sulkowska and co-workers investigated the folding of thymidine kinase and found that its slipknotted structure can be achieved by a simple 'flipping' mechanism in which a slipknot loop rotates over the unknotted native core of the protein [211]. The rotation of the loop is most likely assisted by the presence of glycine and proline residues in the hinge regions [211]. However, the low success rate of folding events observed suggests that other factors may be needed to overcome the topological barrier or that the barrier is large. The same group extended these studies and used the same model to analyse the mechanical unfolding of the slipknot in the same protein [220]. Weak stretching forces resulted in the smooth untying of the slipknot whilst a metastable intermediate with a tightened knot was observed at sufficiently large pulling forces. It is worth noting that this behavior of slipknotted structures is different to that observed for uniformly elastic polymers [220]. Recently, He and co-workers used AFM to study experimentally the mechanical unfolding of AFV3-109, a protein which has a relatively simple slipknotted structure [221, 222]. Results showed that the slipknot untied and the polypeptide chain was fully extended when mechanical forces were applied at both termini as expected [221]. In contrast, applying forces at the N-terminus and the threaded loop resulted in the tightening of the slipknot into a trefoil knot involving ~13 amino acid residues [222]. In both cases, the unfolding process was found to proceed via multiple parallel pathways in either a two- or three-state fashion, and is consistent with a kinetic partitioning mechanism for mechanical unfolding [221, 222].

# 5.4. Evolution and conservation

Despite the fact that there are now a considerable number of topologically knotted proteins in the PDB, it is worth noting that most proteins are unknotted. This suggests that evolution has, in general, avoided such structures. However, a recent study by Sulkowska and co-workers has established that, when they do occur, that both knotted and slipknotted topologies are conserved across different families despite very low sequence similarity [142]. Unsurprisingly, the parts of proteins which are strongly conserved are found within the knotted core and potential hinge regions which it has been speculated are important in the threading of the chain to form a knot or slipknot [142].

For some families of proteins, where there are a sizeable number of knotted and unknotted variants, it has been possible to undertake a phylogenetic analysis of the sequences, and thereby identify how knotted structures may have evolved from unknotted ancestors. Potestio and co-workers generated a phylogenetic tree of transcarbamylase-like folds [223]. In this case, it was known that some knotted and unknotted variants had different degrees of sequence identity suggesting pathways where structures and therefore sequences had diverged at different times. For example, the two knotted enzymes AOTCase and SOTCase share only 35% sequence identity [224] whilst the knotted AOTCase has 41% sequence identity with unknotted OTCase [225]. Reconstruction of the phylogenetic tree demonstrated that all the knotted homologues populate a sub-branch of the tree and that they differ from unknotted homologues by the presence of additional loop segments [223]. Thus, it has been suggested that some knotted structures have evolved from unknotted ones by the insertion of a 'knot-promoting' loop, which effectively encompasses another part of the chain thus forming the knot.

Loops have also been implicated in the formation of knotted structures from other studies. Virnau and co-workers used computational approaches to show that the knotted transcarbamylase AOTCase possesses a rather rigid proline-rich loop, which is lacking in the unknotted OTCase (figure 10(b)) [149]. Interestingly, the stevedore knot in  $\alpha$ -haloacid dehalogenase DehI is also partly formed by a large proline-rich loop that links two unknotted regions within the structure [154].

Using a completely different approach, the group of Yeates have also demonstrated another route to knotted structures through the rational design of a novel knotted structure. In this case, a monomeric knotted protein was created by fusion of C- and N-terminal chains of a homodimer that forms a highly entangled but unknotted structure. This study demonstrated that the genetic fusion and tandem repeat of a gene of an unknotted dimeric protein could lead to trefoil-knotted structures [207].

It is clear that, once formed through some evolutionary pathway, knotted and slipknotted protein structures are highly conserved. However, through both experimental and computational studies, we also know that these types of structures have more complex folding pathways than their unknotted counterparts. This suggests that the knotted and slipknotted motifs within protein families may, in some way, be advantageous and important to either the function, or regulation, of the protein.

#### 5.5. Summary

In summary, both experimental and computational studies have made significant progress in establishing some of the key general features of the folding pathways of topologically complex proteins. In contrast to small monomeric proteins with simple folds, it is clear that proteins with topologically knotted or slipknotted structures have much more complex energy landscapes with many intermediate states and parallel pathways. Computational studies have provided insights into the folding process, which may involve formation of a twisted loop followed by threading via an intermediate slipknot configuration, a plugging route or a 'flipping' mechanism, in which the knotting step may be rate-limiting [141, 211, 226]. In addition, it seems that non-native interactions may play a more important role for these types of structures with complex architectures than for the folding of smaller proteins with relatively simple folds [227–229]. Moreover, the formation of transient misfolded species that results in kinetic traps in the free energy landscape of topologically knotted proteins highly likely requires backtracking events and potentially the action of molecular chaperones so that the native structure can be both rapidly and efficiently achieved [199, 206, 211]. Such a 'frustrated' folding energy landscape is in contrast to the relatively smooth folding funnels proposed for smaller, simpler proteins [192, 230].

A number of recent studies have shown that knotted and slipknotted proteins are conserved suggesting that the knot, or slipknot, potentially play a role in the structure, stability, function or regulation of the protein. Despite this finding, it still has to be unambiguously established whether there are any advantageous properties of a knotted structure over an unknotted one. Indeed, whether there are any chemical or physical properties of such structures that are fundamentally different from unknotted ones. Understanding and identifying such properties will potentially provide key insights for future protein engineering applications and therapeutic developments.

# 6. Synthetic molecular knots

Over the past few decades, the importance of knots and entanglements in naturally occurring biological systems has played a significant role in motivating chemists to develop synthetic strategies for creating topologically complex molecules [16, 17]. Although molecular knots and links have proven to be challenging targets for synthesis, the formation of such fascinating structures has acquired much interest, not only for their beauty, but also for the potential of discovering novel properties. Moreover, the knowledge gained from the synthetically engineered knots may shed light on the folding mechanisms and properties of natural topologically complex structures.

The first synthesis of a non-trivial topology goes back to the [2] catenanes (interlocked rings) reported by Wasserman in 1960 [231]. Since then, many higher order molecular links such as Solomon links [232, 233], Borromean rings [234, 235], a Star of David catenane [236] and a range of [n]catenanes [17, 237-240] (*n* denoting the number of interlocked rings) have been successfully synthesised (figure 13). Molecular knots are also increasingly becoming targets for chemical synthesis [16, 17, 241, 242]. Despite the fact that Frisch and Wasserman had first suggested the possibility of using Möbius strips to direct trefoil knot formation in 1961 [243], it was not until 1989 that the first molecular trefoil knot was synthesised [244]. Excluding DNA-based knotted molecules [245, 246] and composite knots [247], to date, only three different types of knots have been synthesised. These are the trefoil knot  $(3_1)$ , figure-of-eight knot  $(4_1)$  and pentafoil knot  $(5_1)$ . This section of the review focuses on the different



**Figure 13.** Examples of synthesised higher order molecular links: (a) a Solomon link [232], (b) a Borromean ring [234], (c) a Star of David catenane [236] and (d) a copper-templated [2]catenane [249]. In each case, an x-ray crystal structure (left) and a schematic representation of the link (right) are shown. X-ray crystal structures in (a) and (c) were reprinted with permission from [232] and [236], respectively.

synthetic approaches used to produce these molecular knotted structures, their mechanism of formation and physical properties. Molecular links are not discussed here and readers who are interested in these structures are directed to the following [17, 238, 241, 248].

# 6.1. Molecular knots: synthetic approaches and mechanism of formation

The synthesis of molecular knots is challenging, as it requires defined pathways and (usually) entropically demanding transition states to achieve a specific knotted structure. Many early experimental efforts (albeit unsuccessful) and proposed synthetic routes towards molecular knots have provided significant insights into the problems of assembling such systems [250, 251]. Over the past two decades, the field of chemical topology has seen various synthetic strategies and approaches being employed for the preparation of different knotted molecules, many of which rely on template effects related to non-covalent interactions identified from supramolecular and coordination chemistry [16, 17, 248]. Here, we discuss these approaches and, in particular, compare the different mechanisms of knot formation using stepwise synthetic approaches to those of 'all-in one' strategies.

6.1.1. Metal template-based synthetic approaches. Using an extension of Sauvage's original strategy for assembling [2] catenanes [239], Dietrich–Buchecker and Sauvage reported the first successful synthesis of a molecular trefoil knot in 1989 [244]. In this case, the end-groups of a dimetallic, double-stranded helicate, composed of two bisphenanthroline ligands and two copper(I) ions, were connected using Williamson ether synthesis. This generated the three crossing points needed for a trefoil knot; however, it was isolated in 3% yield only. A separate study by Dietrich–Buchecker and co-workers later showed that different spacers linking the phenanthroline

groups were critical in determining the yield. In particular, the use of a rigid 1,3-phenylene spacer was found to assist in the stabilisation of the helicate assembly thus resulting in a yield of 29% [252]. However, it was not until the introduction of efficient catalysts for ring-closing olefin metathesis (RCM) that the best yield for a molecular trefoil knot (74%) was achieved (figure 14(a)) [253]. This successful approach was then extended to the preparation of composite knots, details of which can be found in [247]. In another case, the same group used octahedral iron(II) ions reacted with terpyridine-based ligands to template the synthesis of a trefoil knot [254]. The yield achieved, however, was significantly lower (20%), probably because the macrocyclisation was not as effective as that of the previous ligand-metal ion system. Through collaborative work, the groups of von Zelewsky and Sauvage were able to synthesise the first diastereospecific molecular trefoil knot in 74% yield by fusing chiral groups to a 2,2'-bipyridine ligand, thus controlling the stereochemistry of the two copper(I) ions to which the ligands were coordinated [255].

In 2001, Hunter and co-workers reported the synthesis of a stable, 'open-knotted' structure, wherein a single linear tris-bipyridine ligand was coordinated around an octahedral zinc(II) ion [256]. This strategy directly relates to that published by Sokolov in 1973, when he first proposed that a trefoil knot motif could be achieved by arranging three bidentate ligands around an octahedral metal centre to generate the necessary crossings [257]. However, it was not until a decade later that the same group was able to produce the closed trefoil-knotted structure in 68% yield by trapping the acyclic complex through RCM and subsequent removal of the metal template (figure 14(b)) [258].

Active metal template strategies have also played a significant role in the preparation of interlocked compounds [238]. In this case, a metal ion acts simultaneously as a template as well as a catalyst for the synthesis of an entangled structure. In 2011, Leigh and co-workers used this strategy



**Figure 14.** Schematic representations of metal-template based approaches for the synthesis of molecular knots. (a) Synthesis of a phenanthroline molecular trefoil knot in which copper(I) ions are used as templates for the linear helicates to generate the crossings necessary [253], (b) synthesis of a molecular trefoil knot in which an octahedral zinc(II) ion acts as a template for folding and subsequent threading of the ligand [258], (c) active-template synthesis of a molecular trefoil knot [260]. Metal ions are represented as circles.

to synthesise the smallest molecular trefoil knot to date (a 76-atom long closed structure) in a yield of 24% [259]. A tetrahedral copper(I) ion acts as a template to coordinate a single polypyridyl ligand and form the crossing points, while another copper(I) ion binds to the functional end groups of the ligand, threads the loop through its coordination geometry and subsequently catalyses the covalent bond formation to create the trefoil knot motif (figure 14(c)).

Up until recently, the synthesis of molecular knots via a metal-based template strategy has been mainly performed with transition metals. However, recently, with the use of a lanthanide  $(Ln^{3+})$  ion, Leigh and co-workers demonstrated that it can template three 2,6-pyridinedicarboxamide ligands to which subsequent cyclisation by RCM resulted in an 81-atom loop trefoil knot molecule isolated in 58% yield (figure 14(d)) [260].

In these metal-based template approaches, the molecular knots are clearly formed in a stepwise manner, whereby the ligand(s) are first coordinated to metal ion(s). In some cases, this step results in a molecule in which the single ligand assembles around a central metal ion in such a way that there are a number of crossings of regions of the ligand. Alternatively, a number of ligands preassemble around the central ion(s) resulting in crossings of the individual building blocks. In other cases, there is a threading event through a loop created by the initial metal-ligand complex. In all cases, covalent linkage of either the termini of a single ligand or the monomeric units results in a closed knotted structure.

6.1.2. Hydrogen-bond template approaches. Although not as frequently used as the metal-based template strategies, amide–amide hydrogen bonding interactions have also been



**Figure 15.** Schematic representations of self-assembly approaches used in the synthesis of molecular knots. (a) Synthesis of a molecular trefoil knot via amide-amide hydrogen bonding [262]. (b)–(c) NDI-based aqueous disulphide DCL approaches resulting in the synthesis of: (b) a trefoil knot [6], (c) a Solomon link and a figure-of-eight knot [264]. Figures were adapted with permission from [264]. Copyright (2014) American Chemical Society. (d) Synthesis of a molecular pentafoil knot [269].

shown to be important in the synthesis of molecular knots. In 1994, Hunter and co-workers used this approach to produce what they thought was a [2] catenane from the reaction of a diamine and a diacyl chloride [261]. However, several years later, Vögtle and co-workers repeated this one-step synthesis and with the use of x-ray crystallography, discovered that the resultant molecule was, in fact, a trefoil knot [262]. It was then suggested that it was highly likely that the linear diamine, composed of three units of the diamine and two units of diacyl, forms first, then folds into a helical loop which subsequently self-threads its remaining part through the loop. A reaction between the remaining carboxylic acid chloride unit and the terminal amino groups of the open loop then results in the closing of the loop to form the trefoil knot in 20% yield (figure 15(a)) [242, 262]. This synthetic approach highlights the importance of intramolecular hydrogen bonding in the loop for subsequent knot formation.

In 2006, Feigel and co-workers reported the synthesis of a molecular trefoil knot in 21% yield, which also made use of amide-amide hydrogen bonding interactions [263]. In this case, the trefoil knot was formed unexpectedly during the amide coupling reaction of  $3-\alpha$ -aminodeoxycholanic acid with L-valine. Similar to the previous synthesis, this is a one-pot procedure in which no external templating agent was needed to form the knotted architecture.

6.1.3. Dynamic combinatorial library (DCL) approaches. Recently, Sanders and co-workers reported the self-assembly of a trefoil knot from a naphthalenediimide (NDI)-based aqueous disulphide dynamic combinatorial library (DCL) (figure 15(b)) [6]. In brief, the DCL approach allows the molecules themselves to discover different conformations in solution until those, which are thermodynamically the most stable, persist in the mixture once equilibrium is reached. In this study, knot formation was found to occur after an open linear trimer is formed. This then folds into a structure in which the hydrophobic NDI surfaces are buried and the terminal thiols are close in space to allow disulphide bond formation and ring closure. It was concluded that hydrophobic interactions play an important role in driving the folding of the linear molecule into a thermodynamically favoured knotted structure. It is worth noting that the chirality of the building blocks in the DCL resulted in stereoselectivity of the knotted conformation.

The Sanders group have also used different homochiral NDI-based building blocks, resulting in the formation of two thermodynamically favourable species in water, a topologically chiral Solomon link (60% of the library) and a topologically achiral figure-of-eight ( $4_1$ ) knot (18% of the library) (figure 15(c)) [264]. Solomon link formation is not discussed here and can be found in the cited [264]. As with the previous study, it has been suggested that hydrophobic interactions are the driving force needed for a linear open

tetramer to form a thermodynamically stable  $4_1$ -knotted molecule. Interestingly, however, it was found that by using a racemic mixture of the same building blocks, another more stable structure, a topologically achiral *meso*  $4_1$  knot, can be formed (90% of the library). Although the formation of this *meso*  $4_1$  knot, it is significantly slower than that of the homochiral  $4_1$  knot, it is more thermodynamically stable. From this study, it was concluded that chirality and the number of rigid components in the building block can affect the major structure formed.

6.1.4. Other synthetic approaches to molecular knots. Several other synthetic approaches have also been investigated. In 1997, Stoddart and co-workers isolated a trefoil knot in low yield, wherein a double helical precursor chain is formed with the help of  $\pi$ -donor/ $\pi$ -acceptor interactions [265]. In another case, Siegel's group was able to synthesise a trefoil-knotted precursor in which the crossing points were generated by the combination of a covalently bonded organic scaffold and the coordination of copper(I) ions [266, 267]. Glaser couplings were then used to close the open chain, leading to 85% yield. Although the metal template can be removed with KCN treatment, the final removal of the organic template to create a formal trefoil knot has yet to be achieved. More recently, Trabolsi and co-workers reported the self-assembly of a trefoil knot, amongst other products (a [2] catenane and a Solomon link), with the use of a hybrid metal template-based/ dynamic covalent chemistry (DCC) strategy [268]. A pair of chelating imine ligands were combined with zinc(II) ions in a one-pot reaction to form the three structures, and the trefoil knot was precipitated and isolated by filtration.

In 2012, Leigh and co-workers reported the synthesis of the most complex molecular knot created to date, a pentafoil ( $5_1$ ) knot, in 44% yield [269]. In this one-pot synthesis, iron(II) cyclic double helicates, which create the crossover points, are templated about a chloride anion in a cyclic array and the helicate monomers joined by reversible imine bond formation to form a 160-atom loop pentafoil-knotted structure (figure 15(d)).

### 6.2. Properties of molecular knots

Chirality is ubiquitous in chemistry, and knots are often chiral species. If the pure topological enantiomers of such can be obtained from the resolution of racemates, they will have specific optical properties. In many cases, it has been possible to isolate enantiomerically pure species. For example, enantiomers can be separated with the use of chiral HPLC [270, 271]. In another case, Sauvage and von Zelewsky were able to specifically form a single enantiomer by controlling the stereochemistry of the chiral helicate precursor [255]. Sanders and co-workers were also able to stereoselectively synthesise a trefoil knot by constraining the chirality of the building block in the DCL approach [6]. Recently, Leigh and co-workers who reported the synthesis of a lanthanide-templated molecular trefoil knot speculated that its chirality may influence the photophysical properties of the encapsulated lanthanide ion [260].

The study of the conformational properties of intertwined molecules is also of great interest due to their potential applications in the assembly of molecular switches. As molecular knots are increasingly becoming targets of chemical synthesis, it is important to understand what kind of motion is expected from the knotted topology. A study by Sauvage's group compared the dynamics of two different types of molecular trefoil knots formed by the metal-template based approach, in which the phenanthroline units were linked either by oligomethylene or *m*-phenylene spacers. In both cases, the molecular knots which still contained copper(I) ions were found to be generally rigid in solution [272]. However, removal of the metal ions led to rearrangement of the knotted backbone and, in both cases resulted in different dynamic behaviour. Those molecules containing the oligomethylene linkers had significantly greater conformational mobility in solution in comparison to those with *m*-phenylene spacers [244]. This study also showed that the conformational rigidity of partially or fully demetalated molecular knots can be restored again after re-complexation [272].

How do the conformational dynamics of the amide molecular knots formed via hydrogen bond interactions compare to those of the phenanthroline molecular knots? Based on <sup>1</sup>H- and <sup>31</sup>P-NMR spectroscopic measurements, Vögtle and co-workers reported that the amide molecular trefoil knots retain relatively rigid, non-symmetrical structures in DMSO, even though no metal ion is present [242]. However, addition of other solvents to the solution of these knots rapidly resulted in conformational change, and, in some cases, led to increased flexibility or increased rigidity [242]. Such changes in dynamics brought about by change in solution conditions makes these systems interesting for the development of molecular switches. In another case, the organic trefoil knot synthesised using the DCL approach exhibited sharp NMR signals in water demonstrating that the molecule was relatively rigid under these conditions. The signals remained unchanged upon increasing the temperature (from 298 to 358 K) or adding acetonitrile (from 0 to 50%), indicating that the structure is sufficiently stable such that it does not undergo gross conformational change upon a change in conditions [6], in contrast to the trefoil knots synthesised and studied by the Sauvage and Vögtle groups. Such conformational rigidity was also observed in the highly symmetric, achiral figure-of-eight knot synthesised using the DCL approach [264].

## 6.3. Summary

Although challenging, recently, chemists have successfully developed a number of different experimental strategies for the creation of molecular knots. These approaches have been used to synthesise a number of linked species, including Solomon links, Borromean rings, and a Star of David catenane. However, they have also been employed to make true knotted molecules including a  $3_1$ ,  $4_1$  and  $5_1$  knot. The different synthetic strategies can generally be considered as either a template-based method (for example the metal-based templates), or those which use hydrogen bonding or  $\pi - \pi$  interactions to first preassemble the building block(s) in such a way

that covalent linkage of either the termini or of the monomer units results in a knotted structure. Alternatively, DCL-based approaches utilise the fact that a number of building blocks can come together to form chains of different lengths which can then fold to a thermodynamically stable state. In the first case, there need not be any threading event, but preassembly is crucial, whilst in the DCL approach, threading can occur. The properties of the molecular knots created using synthetic strategies are beginning to become established. Whereas, in some cases, the molecules are rigid in the presence of the templating metal ion, they can clearly undergo conformational change and their flexibility can alter when the metal ion is removed. The dynamics of such systems have also been found to vary depending upon environmental conditions. In other cases, such as those knotted molecules created by the DCL approach, which favours thermodynamically stable states, the evidence suggests these are relatively rigid molecules whose structures do not change significantly with environmental conditions.

As more and more new topologically complex structures are created, this raises the issue of whether knotting, linking, etc convey novel or important properties on the molecule. If they do, then it may be possible to exploit them in practical applications such as materials and pharmaceuticals. Readers who are interested in comprehensive discussion of these synthetic approaches are directed to the following reviews [16, 17, 241, 242, 248].

# 7. Discussion

In our macroscopic world, we are all very familiar with knots and other types of entangled objects. In many cases, knotted chains are incredibly useful at joining, securing and stabilising structures. However, in other cases, they are problematic and need undoing for something to function optimally. At the molecular level, this also appears to be the case. For some types of knotted or entangled molecules, the knot or entanglement conveys a particular and advantageous function to the chain, e.g. regulation of frameshifting in RNA. In contrast, knots can sometimes be hugely detrimental and need to be untied not only for optimal function but, in some cases, survival of an organism.

The mathematical field of knot theory has existed since the 19th century, and a great deal is known in terms of the identification and classification of knots, as well as polynomials to describe the different knotted topologies. In contrast, some molecular knots in Nature, the first of which was identified in DNA some three decades ago, were discovered in other biological polymers only recently, e.g. knots and slipknots in proteins and pseudoknots in RNA. These molecular knots remain relatively poorly understood in terms of their formation, properties and function. In addition, it is only recently that chemists have been able to synthesise knotted molecular species and they are only just beginning to study the properties of these types of structures.

In this review, we have attempted to summarise our current understanding of molecular knots in naturally occurring biological polymers, as well as those synthesised using a number of different chemical methods. Table 3 summarises a number of the key properties of the molecular knots that have been identified in the four different classes of molecules considered here: DNA, RNA, proteins and synthetic. This includes the types of knotted or otherwise topologically complex structures, such as slipknots, pseudoknots, linked chains (catenanes etc) that have either been found in Nature or synthesised. Compared to the knotted topologies that can be generated computationally, the molecular knots characterised experimentally are all relatively simple. Knotted DNA has been found with a crossing number up to 9 whilst for proteins, the highest crossing number identified to date remains a stevedore knot with 6 crossing points. Similarly, for synthesised knotted molecules, only a pentafoil knot with 5 crossings has been obtained for a single chain. Of interest is the observation that no truly knotted species have been found for naturally occurring RNA.

Given that naturally occurring knotted forms of DNA and proteins are known, it is curious that there are no naturally occurring knotted forms of RNA. This is particularly so given the fact that synthetic RNA has been designed and shown to adopt a knotted circular form [132]. Recently, Micheletti et al conducted a systematic search for knots in the thousands of RNA structures in the PDB [289]. Although they found three examples of potentially knotted structures, all of these were low resolution. Comparing these with the structures from homologues that had been determined at much higher resolution, and also taking into account parameters that indicate how well the structure fits the experimental cryo-EM data, Micheletti et al concluded that it was unlikely that these knots were real. They speculate on why naturally occurring RNAs do not contain knots and suggest a number of possible causes: that RNA, which has a much smaller number of folds compared to proteins, has evolved to minimise geometrical complexity and therefore potential topological hindrance due to the need (at least for mRNAs) to translocate through the ribosome during translation. They also conjecture that because RNA structures are more modular in nature and that modular growth has led to longer RNAs, that this is incompatible with forming knotted structures. It is also recognised that the kinetics of folding may play a role: for RNA there is strong evidence that secondary structure forms early followed by tertiary structure, whereas, at least for relatively small proteins, we know that secondary and tertiary structure formation is generally concomitant. It is also noted that knotted forms of RNA may yet be found, for example, in the recently discovered eukaryotic circular RNAs [289].

Knots in homopolymers have been investigated in many experimental and computational studies over many years and much is known about how a number of key parameters such as chain length, solvent conditions, etc, affect the types of knots formed in these systems. On the other hand, much less is known about the factors influencing knot formation in heteropolymers, which includes all the molecular knots discussed here. It is clear, that in contrast to homopolymers, which can have high probabilities of knotting, heteropolymers, in general, have a much lower propensity to form knotted structures. In particular, proteins and RNA (discussed above). Although a number of knotted structures have been identified and characterised in proteins, with four different knot types in eleven different protein folds, it is clear that there are considerably fewer knotted proteins than one might expect for polymers of their size. A recent computational study by Virnau and coworkers has addressed why this might be by investigating the influence of sequence on the probability of knotting using a simple lattice model [290]. In this case, a basic hydrophobicpolar (HP) model was used in which there are favourable interactions between non-bonded H monomers. Chains of some 500 monomer units were studied, which had an overall composition of 50%:50% H:P but which varied in sequence. The introduction of just one more parameter into the model greatly decreased the number of knotted states found. Within the sequences assessed, some had a probability of knotting close to zero, whilst others had very high probabilities of knotting. Thus, this elegant study demonstrates that sequence has a very large impact on whether a chain is likely to form knotted species, or not. Consequently, these results suggest that Nature has favoured/selected protein sequences that are unlikely to form knotted structures. One way in which that might happen is to favour sequences where there is high local structural order, known to suppress knot formation. Given this, but also the fact that sequences are known which greatly increase the probability of knotting, this suggests that for the classes of knotted proteins that exist, that there may be some advantageous property of these systems which has allowed them to be selected and conserved.

Recent computational studies have also addressed the effects of how chain stiffness (or persistence length) influences the knottedness in single chain homopolymers [291]. This is more straightforward to study computationally than experimentally. These computational studies revealed non-trivial behaviour in which the extremes (i.e. a highly flexible chain or a rigid knot) both favoured the unknot. In between these extremes, there is an optimum chain stiffness which maximises the probability of knotting for any given chain length. At this point, the chain is semi-flexible in nature allowing loops to form through which other regions of the chain can thread to form the knot. It is interesting to see whether there is any evidence from experimental studies for this. table 3 reports the persistence length of the different biopolymers that can form topologically knotted or other entangled states. In addition, the number of minimum number of monomers within a chain known to form a knotted structure is given. As is expected, the persistence length of proteins, ssRNA and ssDNA (0.7–3 nm) [277, 279] is considerably smaller than for dsDNA or dsRNA (50-72 nm) [276, 278]. For those biopolymers that are relatively flexible (low persistence length), knotted structures have been characterised for minimum chain lengths of 82 amino acids [141], 104 [132] and 80 nt [285] for proteins, ssRNA and ssDNA, respectively. In contrast, biopolymers with much larger persistence lengths, such as dsDNA, form knots with chains that are considerably longer in length (5000 bp long) [273, 274]. It is interesting to note that RNA, which contains considerable secondary structure and has a persistence length of the order of 70 nm for dsRNA, are not known to form any knotted structures. As such, experimental results support the findings of the computational studies and establish that chain stiffness is an important property of a chain in determining whether it is likely to form a knotted structure.

At this stage, it is not possible to say much about whether such a correlation will be found for synthetic knotted systems created by chemists. For those molecules which are formed by a template-based method, the ligands first preassemble around a central metal ion or other template, and then the ligands are covalently linked. In this case, it is unlikely that there will be a correlation as the mechanism of knot formation does not involve loop formation and threading. In contrast, for knotted molecules synthesised using DCL approaches, there is evidence of an initial polymerisation of monomeric units to form a short chain and then threading of that chain to form the knot. Here, one might expect chain stiffness to play a role. However, with a single example of such, it is impossible, as yet, to say the degree to which persistence length influences knot formation.

For DNA, it is very well established that there are enzymes (topoisomerases), which catalyse both knotting and unknotting. The mechanism of action of this family of enzymes is known, and involves cutting of the DNA chain, movement of one part of the chain relative to another, and the ligation of the two ends of the cut chain to form a closed system. Thus, for DNA, no threading events are required for knot formation. However, there is some evidence that DNA can form knots with no cleavage of the chain. It is interesting to note that topoisomerases, known to work on DNA, can also catalyse knotting and unknotting events in synthetic forms of RNA, even though there is no known biological activity associated with this. For one family of knotted proteins, the bacterial methyltransferases, the chaperonin GroEL-GroES has been shown to significantly accelerate knotting and folding. However, the mechanism of action is not yet established and it is not known whether this chaperonin catalyses the folding of other classes of knotted proteins. So, for naturally occurring systems, catalysis of the knotting of the biopolymer is possible.

In a few cases, some of the physical properties of molecular knots have been characterised in some detail. For example, the electrophoretic mobility of unknotted, knotted and catenated forms of DNA, and to a much lesser degree of RNA, have been studied [280, 292]. Increasing the knot complexity, i.e. the average crossing number (ACN), is known to increase the electrophoretic mobility of DNA. This is due to the increased compactness of the molecule as the ACN increases. So clearly, in DNA and RNA, knots result in more compact states. It is unsurprising, therefore, that knotted DNA can be found in situations where it is densely packaged such as in viral DNA capsids. Is this also the case for knotted proteins and synthetic molecules? Comparing the size of families of proteins which have both knotted and unknotted variants, there is no evidence that the knotted structures are more compact or more densely packed. This may simply be due to the fact that, in many cases, the knotted region is associated with only part of the overall protein structure and a large amount of chain is often found in non-knotted regions. For the knotted small molecules

Type of mo	lecule	Knot types	Other types of entangled species	Catalysis of knot-unknot transitions	Mummum number of monomers needed to create knot <sup>a</sup>	Persistence length	Chemical/ physical properties	Mechanism of formation	Function of the knot or biological consequences of knot formation
	Naturally	3.			5000  bp $(dsDNA)^{\text{b}}$	50nm		Threading	Dense packaging in viral
	occurring: dsDNA	6 <sup>1</sup> 4	Hopf links and other catenanes	Topoisomerases	>750  bp $(dsDNA)^{\circ}$	$(dsDNA)^{d}$ 1.5–3 nm	Compact <sup>e</sup>	No need to thread: action of	DNA capsid
DNA	ssDNA	7–9			Not known for ssDNA	(ssDNA) <sup>d</sup>		upoisomerases and biological processes	transcription
	Synthetic: ssDNA <sup>f</sup>	$\frac{3_1}{4_1}$	Polyhedral catenanes, Borromean rings	Topoisomerases	80 nt	1.5-3  nm $(ssDNA)^{d}$	Compact	Not characterised	N/A
RNA	Naturally occurring	None	Pseudoknot	N/A	N/A	72 nm ( <i>dsRNA</i> ) <sup>d</sup>	Not characterised	Not characterised	Linked to various functions: regulation of ribosome frameshifting
	Synthetic <sup>g</sup>	31	N/A	Topo III (E. coli)	104 nt	0.75–1.27 nm (ssRNA) <sup>d</sup>	Not characterised	Not characterised	N/A
Proteins		م <sub>ک</sub> ہ <u>4</u> ع	Hopf links (both formed by covalent bonding and also non-covalent interactions) and slipknots	GroEL-GroES (not essential) for a bacterial knotted MTase, otherwise not known	82 aa (shallow knot in MJ0366) Tightened knots can be as small as 13 aa	0.7 nm	Dynamics Thermodynamic stability Kinetic stability Mechanical resistance	Threading (through slipknot formation or via a plugging mechanism)	Some experimental and computational evidence for enhanced thermodynamic and/or kinetic stability Experimental evidence that degradation is inhibited for $5_2$ -knotted proteins <sup>h</sup> Computational evidence that translocation is inhibited
Synthetic	Template- based approaches	$3_1$ $5_1$	Solomon links, Borromean rings, Star of David catenane, [n] catenanes	None	1 but typically 3 or more <sup>i</sup>	Not known	Not characterised	Pre-assembly and ligation of monomer units	N/A
molecule	DCL approach	$3_1^{-1}$	Solomon link, catenanes <sup>i</sup>	None	1 but typically 3 or more	Not known	Not characterised	Creation of trimer and threading then closure (for the 3 <sub>1</sub> knot)	N/A
<sup>a</sup> This is the r yet be observ dsDNA which	ninimum length tl ed. <sup>b</sup> Size of the s. ι has a length of 2	hat has bee hortest gen 36 nm [275	n observed to be knotted to c omic phage P4 DNA knot fo 1. <sup>d</sup> References for the persist	ate. In some cases, systemat primed in the capsid [273, 274 fence length values: dsDNA 1	ic studies suggest this ]. <sup>c</sup> This value is estir	will be the mini nated from optics	num length possible, d tweezer pulling exp	whilst for other systems, sl eriments on the basis of a t	norter polymer lengths ma ightened trefoil knot in

J. Phys.: Condens. Matter 27 (2015) 354101

that have been synthesized chemically, it is a little difficult to judge. Unknotted variants generally do not exist, however, the cyclised trefoil knotted species synthesised using DCL methods is extremely compact and certainly more compact than the linear trimer from which it is formed [6].

For proteins, there has been considerable speculation on how the properties of knotted species may differ from their unknotted counterparts. This has been of particular interest, given the evidence that knotted and slipknotted proteins are highly conserved. It has been suggested that changes in the dynamics and rigidity of the protein structure (especially close to active sites or binding sites), as a result of the knot, may play an important role in the activity/function of the protein. This was initially proposed at the time when all of the knotted proteins known were enzymes, however, there are now a number of knotted structures where the knotted region is not involved in any catalytic process. In general, there is relatively little evidence to support this hypothesis. Certainly, for the knotted proteins which have been investigated, there is no evidence that the dynamics of these structures is any different from unknotted ones [177].

It has also been suggested that thermodynamic and kinetic stability of knotted proteins may be greater than that for unknotted ones. Some computational studies have found small increases in thermodynamic stability for knotted proteins, however, other very similar studies have not. Although there is some experimental evidence for enhanced thermal stability of knotted structures, this has only been shown for one or two knotted and pseudo-knotted polymers of knotted proteins, and in this case, only apparent stabilities could be reported due to the irreversible nature of the thermal unfolding. Thus, it remains to be unambiguously established whether a knot increases the thermodynamic stability of proteins. Computationally, there is evidence that knots can decrease unfolding rates and, thus, the kinetic stability of the system. However, other elements of structure, such as the addition of stable beta motifs, also had a similar effect. Experimentally, there is no evidence to suggest that knotted proteins have significantly different unfolding rates compared to the range of unfolding rates measured for proteins lacking knots. This may not be a fair comparison, as unfolding rates can vary by orders of magnitude for proteins with the same unknotted topology but different sequences. This raises the real experimental and technical problem. In order to establish the effect of a knot on any physical property of a protein, it is essential to compare the knotted species with an unknotted species that is the same in all other respects other than the knot. So it is necessary that knotted and unknotted variants have the same overall secondary structure and overall packing of secondary structural elements to form a tertiary structure, and that the sequences corresponding to the secondary structural elements are the same, but only differs in the 'wiring', i.e. the order of the secondary structure in the overall sequence. Computationally, rewiring of a knotted structure to form an unknotted one is trivial. Experimentally, it can be achieved in a number of different ways: circular permutation where the original N and C- termini are joined with a linker and the new N- and C- termini of the circular permutant are in positions which remove

N C H Lim and S E Jackson

the knot. Alternatively, it can be achieved by repositioning the regions of DNA corresponding to different elements of secondary structure (in general it is easiest to do this by making a synthetic gene). The Yeates group has taken a different approach by using disulphide binding to create chains of knotted and pseudo-knotted protein domains. Although the Jackson group has tried to create unknotted variants of known, and characterised knotted proteins, using the first two approaches, neither of them were successful and all attempts resulted in protein which aggregated and could not be studied (Jackson, Pina, Werrell, unpublished results). In order for the effect of the knot on thermodynamic and kinetic stability to be fully addressed experimentally, a system is needed where appropriate knotted and unknotted variants can be made, and where the unfolding is fully reversible.

An increased resistance to mechanical unfolding has also been proposed as a possible consequence of having a knotted structure. Despite early experiments on carbonic anhydrase, where results appeared to show a dramatic increase in mechanical stability in order to obtain full unfolding where high forces were used, later computational studies established that at such high forces the knot becomes wrapped tightly around an element of structure. At lower forces, one would expect the protein to unfold (in terms of its secondary and tertiary structure) to a state in which the chain is extended but still contains a tight knot. This has now been established for a number of other knotted structures and the forces required for mechanical unfolding are well within the range found for many other unknotted proteins. At this point, it is worth mentioning terminology, which can be confusing and can be used differently depending upon discipline. For a structural biologist, a protein can be considered unfolded or unstructured if it has lost all stable secondary and tertiary structure, but still contains a knot. For a physicist or mathematician, this may be not be considered a truly unfolded or unstructured state. For those outside this field, it should also be noted that some transient secondary and/or tertiary structure is known in a number of unfolded states of proteins populated under highly denaturing conditions (for unknotted proteins). This is referred to as residual structure in the denatured state.

Recently, the Jackson and Itzhaki groups have studied the resistance of knotted proteins, including the bacterial trefoil knotted methyltransferases and the 5<sub>2</sub>-knotted UCH-L1 to degradation by the bacterial Clp degradation machine. In contrast to the trefoil knotted proteins which are rapidly degraded, UCH-L1 is extremely resistant to degradation (unpublished results). These results are similar to computational studies on translocation of a knotted protein through a pore, which show that, under certain conditions, the knot can tighten and prevent further translocation. Such translocation is necessary for a protein to be pulled inside the catalytic centre at the heart of the degradation machine. This is currently being investigated further.

How do knots form in heteropolymers? For DNA, there is a substantial amount of evidence for how it can knot and unknot. *In vivo*, topoisomerases are likely to play a dominant role and therefore the mechanism involves effectively cutting of the chain, movement of one part of the chain relative to another,

followed by pasting of the two ends of the chain together to form a different topology. However, even for DNA, it has been shown that a standard threading mechanism involving formation of a loop through which another part of the chain passes can also occur. Such threading mechanisms also occur in proteins, and generally the terminus of the polypeptide chain closest to the knot undergoes the threading event. Considerable insight into loop formation and threading has come from computational studies, and a number of different pathways have been found. The formation of a slipknot, created by a  $\beta$ -hairpin like structure forming at the end of the chain, which is then pulled all the way through the loop, is frequently observed in simulations. However, a plugging pathway in which the end of the chain simply threads through the loop without forming any metastable structure has also been detected. With the synthetic knotted molecules, threading has also been shown to occur with a trefoil-knotted species that forms from a linear trimer using the DCL approach. In general, template-based synthetic methods do not require a threading event. However, relatively little is known about the mechanism of formation of these types of knotted structure.

For naturally occurring biopolymers such as DNA, RNA and proteins, one can ask the question as to whether the knot affects not only the physical properties of the system (discussed above), but also whether there is some biological function associated with the knot, or a biological consequence of knot formation. In the case of DNA, a great deal is known about how the mechanisms by which knots are introduced into the DNA chain. This can occur as the result of many cellular processes such as transcription, replication and recombination. In these cases, if not untied, the knots can go on to have severe detrimental consequences on the cell or organism, hence the ubiquitous nature of cellular topoisomerases which can remove knots promptly and efficiently. This suggests that knots in DNA are problematic. Given the principle of microscopic reversibility, topoisomerases not only catalyse the removal of knots from DNA but also introduce knots into the polymer. However, in another case, knots in DNA may be beneficial. For example, they are common in the densely packaged DNA found in viral capsids and can influence the rate of ejection of the DNA from the capsid. In this case, whether the knots just form as a natural consequence of the confined volume in a capsid, or whether they play an important biological role in influencing ejection rates remains to be established. In all of these cases, knots can be seen as the product of a cellular process or biological environment, and there is no evidence that the knot has any beneficial function on the system. In contrast, for RNA, although there are no naturallyoccurring knotted RNA species, pseudoknots, which have a number of topological crossings whilst not being knotted, are abundant. In this case, it is very clear that there is a close link between the topological form of the RNA and its function, as pseudoknots are found to play a role in frameshifting, transcriptional regulation including the initiation of protein synthesis and template recognition by viral replicases. There is also some evidence that the pseudoknot structure is required for optimum catalytic activity of a number of ribozymes or riboproteins.

For proteins, the role of the knot in the function of the biomolecule remains very unclear. Certainly, knots in polypeptide chains, may influence a number of the physical properties of the protein, such as stability (thermodynamic, kinetic, mechanical) or dynamics (rigidity) that could affect the halflife or the function of the protein *in vivo*. However, although some experimental and computational studies have demonstrated that there may be some effect of the knot on stability and/or dynamics, we are still far from understanding the role of knots in proteins, and definitive experiments, in particular, have yet to be performed. This is, of particular interest and importance, given the conservation of knotted and slipknotted protein structures.

# 8. Summary and future perspective

In summary, it is now clear that, just as in the macroscopic world, molecular knots are abundant in Nature. In addition, we are now able to design and synthesise different knotted species, be they based on nucleic acid building blocks (DNA and RNA), amino acid building blocks (proteins), or a very wide range of building blocks available to synthetic chemists.

For DNA, we already know a considerable amount on how knots can be introduced into DNA chains, how knots can be removed from chains, what biological processes result in knot formation, and some of the biological consequences of knotted structures. For RNA, we can design and make knotted RNA structures even though no naturally occurring species have yet to be found, which raises the fascinating possibility that RNA knots may still yet be found in Nature. In contrast to DNA, we know much less about how knots are formed in polypeptide chains, and the role of those knots in the structure and function of proteins. This is an area where both computational and experimental research programmes are beginning to reveal some of the facets of these systems, both in terms of how the knotted structures are formed, and to a lesser degree, how the knot might influence function. There remains considerable work to be done to fully address these questions and understand these systems. Recently, after decades of little progress, synthetic chemists have developed strategies for the design and synthesis of a number of molecular knots. This really now opens up the field and hopefully will allow many more knotted and other types of topologically complex species to be made and characterised in the near future. Characterisation of these synthetic molecular knots is crucial in order to understand how their properties differ from unknotted ones, and whether these types of molecule will be of use in nanotechnology, biotechnology, medicine, etc.

In the past few decades, interest in knotted systems has moved away from the realm of just the mathematicians and physicists, and chemists and biologists are now equally fascinated by these types of structure. Over the past ten years, it has been a great pleasure to see how these very disparate communities and disciplines have come together, to share knowledge and solve problems within the field. Going forward, we need to continue to do this and to combine a whole range of computational and experimental approaches on different knotted systems, in order to address some of the big questions in the field and to understand the formation and behaviour of knots both on a theoretical basis, but also, on a molecular basis, and in challenging heteropolymeric systems where different regions of the chain can interact with each other in complex ways, and where non-trivial behaviour can be expected.

# Acknowledgments

We would like to thank Dr Sarah Harris, Prof Henri Orland, Prof David Leigh, Augustinas Markevicius and Dr Guzman Gil-Ramirez for their helpful comments on the manuscript. We would also like to thank Dr Gokhan Barin and Prof J Fraser Stoddart for sending us the x-ray crystal structures in figures 13(b) and (d), and Dr G Dan Pantoş for sending us the structures shown in figure 1(g). We would also like to thank many people in the field for interesting discussions and useful advice. N C H Lim is supported by a UBD Chancellor scholarship from the Brunei Government. S E Jackson would like to dedicate this review to the memory of her mother, Ursula Margaret Jackson, who taught her a great deal even though she did not know any chemistry.

# References

- Airas U and Heinonen S 2002 Clinical significance of true umbilical knots: a population-based analysis *Am. J. Perinatol.* **19** 127–32
- [2] (http://shortwhitecoats.com/2011/how-to-suture-a-wound-avideo-tutorial
- [3] (www.womansday.com/style-beauty/fashion-style/how-to-tiea-reef-knot)-118095 - slide-1
- [4] (http://library.med.utah.edu/WebPath/PLACHTML/PLAC028. html)
- [5] Stray J E, Crisona N J, Belotserkovskii B P, Lindsley J E and Cozzarelli N R 2005 The saccharomyces cerevisiae Smc2/4 condensin compacts DNA into (+) chiral structures without net supercoiling *J. Biol. Chem.* 280 34723–34
- [6] Ponnuswamy N, Cougnon F B L, Clough J M, Pantoş G D and Sanders J K M 2012 Discovery of an organic trefoil knot Science 338 783–5
- [7] Arsuaga J, Vazquez M, McGuirk P, Trigueros S, Sumners D W L and Roca J 2005 DNA knots reveal a chiral organization of DNA in phage capsids *Proc. Natl Acad. Sci. USA* **102** 9165–9
- [8] Buck D 2009 DNA topology *Proc. Symp. in Applied Mathematics* vol 66, ed D Buck and E Flapan (Providence, RI: American Mathematical Society pp 1–33
- [9] Dean F B, Stasiak A, Koller T and Cozzarelli N R 1985 Duplex DNA knots produced by *E. coli* topoisomerase I. Structure and requirements for formation *J. Biol. Chem.* 260 4975–83
- [10] Giedroc D P and Cornish P V 2009 Frameshifting RNA pseudoknots: structure and mechanism Virus Res. 139 193–208
- [11] Gultyaev A P, Olsthoorn R C L, Pleij C W A and Westhof E 2012 RNA Structure: Pseudoknots. eLS (Chichester: Wiley)
- [12] Jackson S E 2012 Characterising the folding pathways of topologically knotted proteins *Curr. Phys. Chem.* 2 59–78
- [13] Mallam A L 2009 How does a knotted protein fold? *Febs J.* 276 365–75
- [14] Staple D W and Butcher S E 2005 Pseudoknots: RNA structures with diverse functions *PLoS Biol.* 3 e213
- [15] Virnau P, Mallam A and Jackson S 2011 Structures and folding pathways of topologically knotted proteins *J. Phys.*: *Condens. Matter* 23 033101

- [16] Ayme J-F, Beves J E, Campbell C J and Leigh D A 2013 Template synthesis of molecular knots *Chem. Soc. Rev.* 42 1700–12
- [17] Forgan R S, Sauvage J-P and Stoddart J F 2011 Chemical topology: complex molecular knots, links and entanglements *Chem. Rev.* 111 5434–64
- [18] Adams C C 2004 The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots (Providence, RI: American Mathematical Society)
- [19] Mishra R and Bhushan S 2011 Knot theory in understanding proteins J. Math. Biol. 65 1187–213
- [20] Freyd P, Yetter D, Hoste J, Lickorish W B R, Millett K C and Ocneanu A 1985 A new polynomial invariant of knots and links *Bull. Am. Math. Soc.* 12 239–46
- [21] Jones V F R 1985 A polynomial invariant for knots via von neumann algebras *Bull. Am. Math. Soc.* **12** 103–11
- [22] Ganzell S and Kapp A V 2008 Chirality versus HOMFLY and Kauffman polynomials J. Knot Theory Ramifications 17 1519
- [23] Ramadevi P, Govindarajan T R and Kaul R K 1994 Chirality of knots 9<sub>42</sub> and 10<sub>71</sub> and Chern–Simons theory *Mod. Phys. Lett.* A 9 3205–17
- [24] Taylor W R and Lin K 2003 Protein knots: a tangled problem Nature 421 25
- [25] Koniaris K and Muthukumar M 1991 Self-entanglement in ring polymers J. Chem. Phys. 95 2873–81
- [26] Taylor W R 2000 A deeply knotted protein structure and how it might fold *Nature* 406 916–9
- [27] Khatib F, Weirauch M T and Rohl C A 2006 Rapid knot detection and application to protein structure prediction *Bioinformatics* 22 e252–9
- [28] Millett K C, Dobay A and Stasiak A 2005 Linear random knots and their scaling behavior *Macromolecules* 38 601–6
- [29] Watson J D and Crick F H C 1953 A structure for deoxyribose nucleic acid *Nature* 171 737–8
- [30] Liu L F, Depew R E and Wang J C 1976 Knotted singlestranded DNA rings: a novel topological isomer of circular single-stranded DNA formed by treatment with *E.coli* omega protein *J. Mol. Biol.* **106** 439–52
- [31] Liu L F, Perkocha L, Calendar R and Wang J C 1981 Knotted DNA from bacteriophage capsids *Proc. Natl Acad. Sci.* USA 78 5498–502
- [32] Deweese J E, Osheroff M A and Osheroff N 2008 DNA topology and topoisomerases: teaching a 'knotty' subject *Biochem.Mol. Biol. Educ.* 37 2–10
- [33] Arsuaga J, Vazquez M, Trigueros S, Sumners D W L and Roca J 2002 Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids *Proc. Natl Acad. Sci. USA* **99** 5373–7
- [34] Shaw S Y and Wang J C 1993 Knotting of a DNA chain during ring closure Science 260 533–6
- [35] Sogo J M, Stasiak A, Martinez-Robles M L, Krimer D B, Hernandez P and Schvartzman J B 1999 Formation of knots in partially replicated DNA molecules *J. Mol. Biol.* 286 637–43
- [36] Bates A D and Maxwell A 1997 DNA topology: topoisomerases keep it simple *Curr. Biol.* 7 R778–81
- [37] Bates A D and Maxwell A 2005 DNA Topology (Oxford: Oxford University Press)
- [38] Buck D and Flapan E 2007 Predicting knot or catenane type of site-specific recombination products J. Mol. Biol. 374 1186–99
- [39] Wasserman S A, Dungan J M and Cozzarelli N R 1985 Discovery of a predicted DNA knot substantiates a model for site-specific recombination *Science* 229 171–4
- [40] Valencia K and Buck D 2011 Predicting knot and catenane type of products of site-specific recombination on twist knot substrates J. Mol. Biol. 411 350–67

- [41] Grindley N D, Whiteson K L and Rice P A 2006 Mechanisms of site-specific recombination *Annu. Rev. Biochem.* 75 567–605
- [42] Witz G and Stasiak A 2010 DNA supercoiling and its role in DNA decatenation and unknotting *Nucl. Acids Res.* 38 2119–33
- [43] Olavarrieta L, Hernandez P, Krimer D B and Schvartzman J B 2002 DNA knotting caused by head-on collision of transcription and replication J. Mol. Biol. 322 1–6
- [44] Lopez V, Martinez-Robles M L, Hernandez P, Krimer D B and Schvartzman J B 2012 Topo IV is the topoisomerase that knots and unknots sister duplexes during DNA replication *Nucl. Acids Res.* 40 3563–73
- [45] Liu L F, Davis J L and Calendar R 1981 Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases *Nucl. Acids Res.* 9 3979–89
- [46] Wolfson J S, McHugh G L, Hooper D C and Swartz M N 1985 Knotting of DNA molecules isolated from deletion mutants of intact bacteriophage P4 *Nucl. Acids Res.* 13 6695–702
- [47] Isaken M, Julien B, Calendar R and Lindgvist B H 1999
  Isolation of knotted DNA from coliphage P4 *Methods Mol. Biol.* 94 69–74
- [48] Micheletti C, Marenduzzo D, Orlandini E and Sumners D W L 2008 Simulations of knotting in confined circular DNA *Biophys. J.* 95 3591–9
- [49] Rollins G C, Petrov A S and Harvey S C 2008 The role of DNA twist in the packaging of viral genomes *Biophys. J.* 94 L38–40
- [50] Ashley C and Lee J S 2000 A triplex-mediated knot between separated polypurine-polyrimidine tracts in circular DNA blocks transcription by *E.coli* RNA polymerase *DNA Cell Biol.* 19 235–41
- [51] Portugal J and Rodriguez Campos A 1996 T7 RNA polymerase cannot transcribe through a highly knotted DNA template *Nucl. Acids Res.* 24 4890–94
- [52] Deibler R W, Mann J K, Sumners D W L and Zechiedrich E L 2007 Hin-mediated DNA knotting and recombining promote replicon dysfunction and mutation *BMC Mol. Biol.* 8 44
- [53] Buck G R and Zechiedrich E L 2004 DNA disentangling by type-2 topoisomerases J. Mol. Biol. 340 933–9
- [54] Deibler R W, Rahmati S and Zechiedrich E L 2001 Topoisomerase IV, alone, unkots DNA in *E. coli Genes Dev.* 15 748–61
- [55] Marenduzzo D, Micheletti C, Orlandini E and Sumners D W L 2013 Topological friction strongly affects viral DNA ejection *Proc. Natl Acad. Sci. USA* 110 20081–6
- [56] Mann J K 2007 DNA knotting: occurrences, consequences and resolution *Electronic Theses, Treatises and Dissertations Paper 2754*
- [57] Mirkin S M 2001 DNA topology: fundamentals *Encyclopedia* of Life Sciences (London: Nature Publishing Group) pp 1–11
- [58] Schvartzman J B, Martinez-Robles M L, Hernandez P and Krimer D B 2013 The benefit of DNA supercoiling during replication *Biochem. Soc. Trans.* 41 646–51
- [59] Doudna J A and Cate J H 1997 RNA structure: crystal clear? Curr. Opin. Struct. Biol. 7 310–6
- [60] Hermann T and Patel D J 1999 Stitching together RNA tertiary architectures J. Mol. Biol. 294 829–49
- [61] Rietveld K, van Poelgeest R, Pleij C W A, van Boom J H and Bosch L 1982 The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA *Nucl. Acids Res.* 10 1929–46
- [62] Pleij C W A, Rietveld K and Bosch L 1985 A new principle of RNA folding based on pseudoknotting *Nucl. Acids Res.* 13 1717–31
- [63] Westhof E and Jaeger L 1992 RNA pseudoknots Curr. Opin. Struct. Biol. 2 327–33

- [64] Kim N, Shiffeldrim N, Gan H H and Schlick T 2004 Candidates for novel RNA topologies *J. Mol. Biol.* 341 1129–44
- [65] Ke A, Zhou K, Ding F, Cate J H and Doudna J A 2004 A conformational switch controls hepatitis delta virus ribozyme catalysis *Nature* 429 201–5
- [66] Theimer C A, Blois C A and Feigon J 2005 Structure of the human telomerase RNA pseudokont reveals conserved tertiary interactions essential for function *Mol. Cell* 17 671–82
- [67] Baril M, Dulude D, Steinberg S V and Brakier-Gingras L 2003 The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot *J. Mol. Biol.* 331 571–83
- [68] Brierley I, Gilbert R J C and Pennell S 2008 RNA pseudoknots and the regulation of protein synthesis *Biochem. Soc. Trans.* 36 684–9
- [69] Brierley I, Pennell S and Gilbert R J C 2007 Viral RNA pseudoknots: versatile motifs in gene expression and replication *Nat. Rev. Microbiol.* 5 598–610
- [70] Mans R M W, Pleij C W A and Bosch L 1991 Transfer RNA-like structures: structure, function and evolutionary significance *Eur. J. Biochem.* 201 303–24
- [71] Michel F and Westhof E 1990 Modelling of the 3D architecture of group I catalytic introns based on comparative sequence analysis J. Mol. Biol. 216 585–610
- [72] Thill G, Vasseur M and Tanner N K 1993 Structural and sequence elements required for the self-cleaving activity of the hepatitis delta virus ribozyme *Biochemistry* 32 4254–62
- [73] Wadkins T S, Perrotta A T, Ferre-D'Amare A R, Doudna J A and Been M D 1999 A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes RNA 5 720–7
- [74] Tanaka Y, Hori T, Tagaya M, Sakamoto T, Kurihara Y, Katahira M and Uesugi S 2002 Imino proton NMR analysis of HDV ribozymes: nested double pseudoknot structure and Mg2<sup>+</sup> ion-binding site close to the catalytic core in solution *Nucl. Acids Res.* **30** 766–74
- [75] Lai M M 1995 The molecular biology of hepatitis delta virus Annu. Rev. Biochem. 64 259–86
- [76] Kuo M Y, Sharmeen L, Dinter-Gottlieb G and Taylor J 1988 Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus *J. Virol.* 62 4439–44
- [77] Ferre-D'Amare A R, Zhou K and Doudna J A 1998 Crystal structure of a hepatitis delata virus ribozyme *Nature* 395 567–74
- [78] Ferré-D'Amaré A R and Scott W G 2010 Small self-cleaving ribozymes Cold Spring Harb. Perspect. Biol. 2 a003574
- [79] Klein D J and Ferre-D'Amare A R 2006 Structural basis of glmS ribozyme activation by glucosamine-6-phosphate *Science* 313 1752–6
- [80] Salehi-Ashtiani K, Luptak A, Litovchick A and Szostak J W 2006 A genomewide search for ribozymes reveals an HDVlike sequence in the human CPEB3 gene Science 313 1788–92
- [81] Cech T R 2004 Beginning to understand the end of the chromosome *Cell* **116** 273–9
- [82] Blackburn E H 2001 Switching and signalling at the telomere Cell 106 661–73
- [83] Blackburn E H and Collins K 2011 Telomerase: an RNP enzyme synthesizes DNA Cold Spring Harb. Perspect. Biol. 3 a003558
- [84] Shay J W and Wright W E 2006 Telomerase therapeutics for cancer: challenges and new directions *Nat. Rev. Drug Discovery* 5 577–84
- [85] Wong J M and Collins K 2003 Telomere maintenance and disease *Lancet* 362 983–8

- [86] Marciniak R A, Johnson F B and Guarente L 2000 Dyskeratosis congenita, telomeres and human ageing *Trends Genetics* 16 193–5
- [87] Chen J L, Blasco M A and Greider C W 2000 Secondary structure of vertebrate telomerase RNA *Cell* 100 503–14
- [88] Collins K 2006 The biogenesis and regulation of telomerase holoenzymes Nat. Rev. Mol. Cell Biol. 7 484–94
- [89] Gilley D and Blackburn E H 1999 The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein *Proc. Natl Acad. Sci.* USA 96 6621–5
- [90] Tzfati Y, Knight Z, Roy J and Blackburn E H 2003 A novel pseudoknot element is essential for the action of a yeast telomerase *Genes Dev.* 17 1779–988
- [91] Comolli L R, Smirnov I, Xu L, Blackburn E H and James T L 2002 A molecular switch underlies a human telomerase disease *Proc. Natl Acad. Sci.USA* 99 16998–7003
- [92] Theimer C A, Finger L D, Trantirek L and Feigon J 2003 Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA *Proc. Natl Acad. Sci.USA* 100 449–54
- [93] Chen J L and Greider C W 2004 Telomerase RNA structure and function: implications for dyskeratosis congenita *Trends Biochem. Sci.* 29 183–92
- [94] Vulliamy T, Marrone A, Dokal I and Mason P J 2002 Association between aplastic anaemia and mutations in telomerase RNA *Lancet* 359 2168–70
- [95] Gesteland R F and Atkins J F 1996 Recoding: dynamic reprogramming of translation Annu. Rev. Biochem. 65 741–68
- [96] ten Dam E B, Pleij C W A and Bosch L 1990 RNA pseudoknots: translational frameshifting and readthrough on viral RNAs Virus Genes 4 121–36
- [97] Brierley I 1995 Ribosomal frameshifting viral RNAs J. Genetic Virol. 76 1885–92
- [98] Giedroc D P, Theimer C A and Nixon P L 2000 Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting *J. Mol. Biol.* 298 167–85
- [99] Somogyi P, Jenner A J, Brierley I and Inglis S C 1993 Ribosomal pausing during translation of an RNA pseudoknot *Mol. Cell Biol.* 13 6931–40
- [100] Takyar S, Hickerson R P and Noller H F 2005 mRNA helicase activity of the ribosome *Cell* **120** 49–58
- [101] Plant E P and Dinman J D 2005 Torsional restraint: a new twist on frameshifting pseudoknots *Nucl. Acids Res.* 33 1825–33
- [102] Plant E P, Muldoon Jacobs K L, Harger J W, Meskauskas A, Jacobs J L, Baxter J L, Petrov A N and Dinman J D 2003 The 9Å solution: how mRNA pseudoknots promote efficient programmed –1 ribosomal frameshifting RNA 9 168–74
- [103] Shen L X and Tinoco I 1995 The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus J. Mol. Biol. 247 963–78
- [104] Chen X, Kang H, Shen L X, Chamorro M, Varmus H E and Tinoco I 1996 A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA J. Mol. Biol. 260 479–83
- [105] Michiels P J, Versleijen A A, Verlaan P W, Pleij C W A, Hilbers C W and Heus H A 2001 Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting J. Mol. Biol. 310 1109–23
- [106] Berry K E, Waghray S, Mortimer S A, Bai Y and Doudna J A 2011 Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning *Structure* 19 1456–66
- [107] Lukavsky P J 2009 Structure and function of HCV IRES domains Virus Res. 139 166–71

- [108] Niepmann M 2013 Hepatitis C virus RNA translation Hepatitis C Virus L from Molecular Virology to Antiviral Therapy, Current Topics in Microbiology and Immunology ed R Bartenschlager vol 369 (Berlin: Springer) pp 143–66
- [109] Berry K E, Waghray S and Doudna J A 2010 The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit *RNA* 16 1559–69
- [110] Dreher T W and Miller W A 2006 Translational control in positive strand RNA plant viruses *Virology* 344 185–97
- [111] Kaur S, Gillet R, Li W, Gursky R and Frank J 2006 Cryo-EM visualization of transfer messenger RNA with two SmpBs in a stalled ribosome *Proc. Natl Acad. Sci. USA* 103 16484–9
- [112] Keiler K C, Waller P R and Sauer R T 1996 Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA Science 271 990–3
- [113] Akutsu T 2000 Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots *Discrete Appl. Math.* **104** 45–62
- [114] Rivas E and Eddy S R 1999 A dynamic programming algorithm for RNA structure prediction including pseudoknots J. Mol. Biol. 285 2053–68
- [115] Dirks R M and Pierce N A 2003 A partition function algorithm for nucleic acid secondary structure including pseudoknots J. Comput. Chem. 24 1664–77
- [116] Reeder J, Steffen P and Giegerich R 2007 pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows *Nucl. Acids Res.* 35 W320–4
- [117] Lyngsø R B and Pedersen C N 2000 RNA pseudoknot prediction in energy-based models J. Comput. Biol. 7 409–27
- [118] Chen X, He S M, Bu D, Zhang F, Wang Z, Chen R and Gao W 2008 FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space *Bioinformatics* 24 1994–2001
- [119] Andronescu M S, Pop C and Condon A E 2010 Improved free energy parameters for RNA pseudoknotted secondary structure prediction RNA 16 26–42
- [120] Sato K, Kato Y, Hamada M, Akutsu T and Asai K 2011 IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming *Bioinformatics* 27 i85–93
- [121] Bon M and Orland H 2011 TT2NE: a novel algorithm to predict RNA secondary structures with pseudoknots *Nucl. Acids Res.* **39** e93
- [122] Bon M, Micheletti C and Orland H 2013 McGenus: a Monte Carlo algorithm to predict RNA secondary structures with pseudoknots *Nucl. Acids Res.* 41 1895–900
- [123] Jabbari H and Condon A 2014 A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures *BMC Bioinf.* 15 1–17
- [124] Moss W N 2013 Computational prediction of RNA secondary structure *Methods Enzymol.* 530 3–65
- [125] Shapiro B A, Yingling Y G, Kasprzak W and Bindewald E 2007 Bridging the gap in RNA structure prediction *Curr. Opin. Struct. Biol.* 17 157–65
- [126] Akutsu T 2006 Recent advances in RNA secondary structure prediction with pseudoknots *Curr. Bioinf.* 1 115–29
- [127] van Batenburg F H D, Gultyaev A P, Pleij C W A, Ng J and Oliehoek J 2000 Pseudobase: a database with RNA pseudoknots *Nucl. Acids Res.* 28 201–4
- [128] Ren J, Rastegari B, Condon A and Hoos H H 2005 HotKnots: heuristic prediction of RNA secondary structures including pseudoknots RNA 11 1494–504
- [129] Byun Y and Han K 2006 Pseudoviewer: web application and web service for visualising RNA pseudoknots and secondary structures *Nucl. Acids Res.* 34 W416–22
- [130] Metzler D and Nebel M E 2008 Predicting RNA secondary structures with pseudoknots by MCMC sampling *J. Math. Biol.* 56 161–81

- [131] Bellaousov S and Matthews D H 2010 ProbKnot: fast prediction of RNA secondary structure including pseudoknots RNA 16 1870–80
- [132] Wang H, Di Gate R J and Seeman N C 1996 An RNA topoisomerase Proc. Natl Acad. Sci. USA 93 9477–82
- [133] Cao Z, Roszak A W, Gourlay L J, Lindsay J G and Isaacs N W 2005 Bovine mitochondrial peroxiredoxin III forms a two-ring catenane *Structure* 13 1661–4
- [134] Boutz D R, Cascio D, Whitelegge J, Perry L J and Yeates T O 2007 Discovery of a thermophilic protein complex stabilized by topologically interlinked chains *J. Mol. Biol.* 368 1332–44
- [135] Haglund E, Sulkowska J I, Noel J K, Lammert H, Onuchic J N and Jennings P A 2014 Pierced lasso bundles are a new class of knot-like motifs *Plos Comput. Biol.* 10 e1003613
- [136] McDonald N Q and Hendrickson W A 1993 A structural superfamily of growth factors containing a cystine knot motif *Cell* 73 421–4
- [137] Craik D J, Daly N L and Waine C 2001 The cystine knot motif in toxins and implications for drug design *Toxicon* 39 43–60
- [138] Craik D J, Cemazar M, Wang C K and Daly N L 2006 The cyclotide family of circular miniproteins: nature's combinatorial peptide template *Biopolymers* 84 250–66
- [139] Taylor W R 2007 Protein knots and fold complexity: some new twists *Comput. Biol. Chem.* 31 151–62
- [140] King N P, Yeates E O and Yeates T O 2007 Identification of rare slipknots in proteins and their implications for stability and folding J. Mol. Biol. 373 153–66
- [141] Noel J K, Sulkowska J I and Onuchic J N 2010 Slipknotting upon native-like loop formation in a trefoil knot protein *Proc. Natl Acad. Sci. USA* **107** 15403–8
- [142] Sulkowska J I, Rawdon E J, Millett K C, Onuchic J N and Stasiak A 2012 Conservation of complex knotting and slipknotting patterns in proteins *Proc. Natl Acad. Sci. USA* 109 E1715–23
- [143] Burman R, Gunasekera S, Stromstedt A A and Goransson U 2014 Chemistry and biology of cyclotides: circular plant peptides outside the box J. Nat. Prod. 77 724–36
- [144] Craik D J, Cemazar M and Daly N L 2007 The chemistry and biology of cyclotides *Curr. Opin. Drug Discovery Dev.* 10 176–84
- [145] Iyer S and Acharya K R 2011 Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines *FEBS J.* 278 4304–22
- [146] Jiang X, Dias J A and He X 2014 Structural biology of glycoprotein hormones and their receptors: insights to signaling *Mol. Cell Endocrinol.* 382 424–51
- [147] Mansfield M L 1994 Are there knots in proteins? Nat. Struct. Biol. 1 213–4
- [148] Lua R C and Grosberg A Y 2006 Statistics of knots, geometry of conformations, and evolution of proteins *PLoS Comput. Biol.* 2 e45
- [149] Virnau P, Mirny L A and Kardar M 2006 Intricate knots in proteins: function and evolution *Plos Comput. Biol.* 2 1074–9
- [150] Lai Y, Chen C and Hwang J K 2012 pKNOT v.2: the protein KNOT web server *Nucl. Acids Res.* 40 W228–31
- [151] Kolesov G, Virnau P, Kardar M and Mirny L A 2007 Protein knot server: detection of knots in protein structures *Nucl. Acids Res.* 35 W425–8
- [152] Jamroz M, Niemyska W, Rawdon E J, Stasiak A, Millett K C, Sulkowski P and Sulkowska J I 2014 KnotProt: a database of proteins with knots and slipknots *Nucl. Acids Res.* 43(DI) D306–14
- [153] Rawdon E J, Millett K C, Sulkowska J I and Stasiak A 2013 Knot localization in proteins *Biochem. Soc. Trans.* 41 538–41

- [154] Bölinger D, Sulkowska J I, Hsu H-P, Mirny L A, Kardar M, Onuchic J N and Virnau P 2010 A Stevedore's protein knot *Plos Comput. Biol.* 6 e1000731
- [155] Takusagawa F, Kamitori S and Markham G D 1996 Structure and function of S-adenosylmethionine synthetase: crystal structures of S-adenosylmethionine synthetase with ADP, BrADP, and PPi at 28 angstroms resolution *Biochemistry* 35 2586–96
- [156] Takusagawa F, Kamitori S, Misaki S and Markham G D 1996 Crystal structure of S-adenosylmethionine synthetase J. Biol. Chem. 271 136–47
- [157] Ahn H J, Kim H W, Yoon H J, Lee B I, Suh S W and Yang J K 2003 Crystal structure of tRNA(m1G37) methyltransferase: insights into tRNA recognition *EMBO J.* 22 2593–603
- [158] Finn R D et al 2014 Pfam: the protein families database Nucl. Acids Res. 42 D222–30
- [159] Chiang P K, Gordon R K, Tal J, Zeng G C, Doctor B P, Pardhasaradhi K and McCann P P 1996
   S-adenosylmethionine and methylation *FASEB J*. 10 471–80
- [160] Lim K, Zhang H, Tempczyk A, Krajewski W, Bonander N, Toedt J, Howard A, Eisenstein E and Herzberg O 2003 Structure of the YibK methyltransferase from haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot *Proteins-Struct. Funct. Genetics* **51** 56–67
- [161] Nureki O et al 2002 An enzyme with a deep trefoil knot for the active-site architecture Acta Crystallogr. D Biol. Crystallogr. 58 1129–37
- [162] Nureki O, Watanabe K, Fukai S, Ishii R, Endo Y, Hori H and Yokoyama S 2004 Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme *Structure* 12 593–602
- [163] Zarembinski T I, Kim Y, Peterson K, Christendat D, Dharamsi A, Arrowsmith C H, Edwards A M and Joachimiak A 2003 Deep trefoil knot implicated in RNA binding found in an archaebacterial protein *Proteins* 50 177–83
- [164] Shi D, Morizono H, Yu X, Roth L, Caldovic L, Allewell N M, Malamy M H and Tuchman M 2005 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 14366–9
- [165] Shi D, Gallegos R, DePonte J, 3rd Morizono H, Yu X, Allewell N M, Malamy M and Tuchman M 2002 Crystal structure of a transcarbamylase-like protein from the anaerobic bacterium bacteroides fragilis at 2.0 a resolution J. Mol. Biol. 320 899–908
- [166] van Roon A M, Loening N M, Obayashi E, Yang J C, Newman A J, Hernandez H, Nagai K and Neuhaus D 2008 Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif *Proc. Natl Acad. Sci. USA* 105 9621–6
- [167] Biou V, Dumas R, Cohen-Addad C, Douce R, Job D and Pebay-Peyroula E 1997 The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 A resolution *EMBO J.* 16 3405–15
- [168] Leung E W and Guddat L W 2009 Conformational changes in a plant ketol-acid reductoisomerase upon Mg(2<sup>+</sup>) and NADPH binding as revealed by two crystal structures *J. Mol. Biol.* 389 167–82
- [169] Das C, Hoang Q Q, Kreinbring C A, Luchansky S J, Meray R K, Ray S S, Lansbury P T, Ringe D and Petsko G A 2006 Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1 Proc. Natl Acad. Sci. USA 103 4675–80

- [170] Johnston S C, Larsen C N, Cook W J, Wilkinson K D and Hill C P 1997 Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 a resolution *EMBO J*. 16 3787–96
- [171] Johnston S C, Riddle S M, Cohen R E and Hill C P 1999 Structural basis for the specificity of ubiquitin C-terminal hydrolases *EMBO J.* 18 3877–87
- [172] 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* 438 325–31
- [173] Wagner J R, Zhang J, Brunzelle J S, Vierstra R D and Forest K T 2007 High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution *J. Biol. Chem.* 282 12298–309
- [174] Yeates T, 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
- [175] Bornschlogl T, Anstrom D M, Mey E, Dzubiella J, Rief M and Forest K T 2009 Tightening the knot in phytochrome by single-molecule atomic force microscopy *Biophys. J.* 96 1508–14
- [176] Soler M A, Nunes A and Faisca P F 2014 Effects of knot type in the folding of topologically complex lattice proteins *J. Chem. Phys.* 141 025101
- [177] Andersson F I, Werrell E F, McMorran L, Crone W J K, Das C, Hsu S-T D and Jackson S E 2011 The effect of parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1 J. Mol. Biol. 407 261–72
- [178] Sulkowska J I, Sulkowski P, Szymczak P and Cieplak M 2008 Stabilizing effect of knots on proteins *Proc. Natl Acad. Sci. USA* **105** 19714–9
- [179] 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* 24 627–30
- [180] Soler M A and Faisca P F N 2013 Effects of knots on protein folding properties *Plos One* 8 e74755
- [181] Wang T and Ikai A 1999 Protein stretching III: forceextension curves of tethered bovine carbonic anhydrase B to the silicon substrate under native, intermediate and denaturing conditions *Japan. J. Appl. Phys.* **38** 3912–7
- [182] Alam M T, Yamada T, Carlsson U and Ikai A 2002 The importance of being knotted: effects of the C-terminal knot structure on enzymatic and mechanical properties of bovine carbonic anhydrase II *FEBS Lett.* **519** 35–40
- [183] Dzubiella J 2013 Tightening and untying the knot in human carbonic anhydrase III J. Phys. Chem. Lett. 4 1829–33
- [184] Wang T, Arakawa H and Ikai A 2001 Force measurement and inhibitor binding assay of monomer and engineered dimer of bovine carbonic anhydrase B *Biochem. Biophys. Res. Commun.* 285 9–14
- [185] Szymczak P 2013 Tight knots in proteins: can they block the mitochondrial pores? *Biochem. Soc. Trans.* 41 620–4
- [186] Huang L and Makarov D E 2008 Translocation of a knotted polypeptide through a pore J. Chem. Phys. 129 121107
- [187] Daggett V and Fersht A 2003 The present view of the mechanism of protein folding *Nat. Rev. Mol. Cell Biol.* 4 497–502
- [188] Daggett V and Fersht A R 2003 Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28 18–25
- [189] Fersht A R and Daggett V 2002 Protein folding and unfolding at atomic resolution Cell 108 573–82
- [190] Jackson S E 1998 How do small single-domain proteins fold? Fold. Des. 3 R81–91
- [191] Jackson S E and Fersht A R 1991 Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition *Biochemistry* 30 10428–35

- [192] Dill K A and Chan H S 1997 From Levinthal to pathways to funnels *Nat. Struct. Biol.* **4** 10–9
- [193] Onuchic J N and Wolynes P G 2004 Theory of protein folding Curr. Opin. Struct. Biol. 14 70–5
- [194] Mallam A, Rogers J M and Jackson S 2010 Experimental detection of knotted conformations in denatured proteins *Proc. Natl Acad. Sci. USA* 107 8189–94
- [195] Mallam A L and Jackson S E 2005 Folding studies on a knotted protein J. Mol. Biol. 346 1409–21
- [196] Mallam A L and Jackson S E 2006 Probing nature's knots: the folding pathway of a knotted homodimeric protein *J. Mol. Biol.* 359 1420–36
- [197] Mallam A L and Jackson S E 2007 The dimerization of an alpha/beta-knotted protein is essential for structure and function *Structure* 15 111–22
- [198] Mallam A L and Jackson S E 2007 A comparison of the folding of two knotted proteins: YbeA and YibK J. Mol. Biol. 366 650–65
- [199] Mallam A L and Jackson S E 2012 Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins *Nat. Chem. Biol.* 8 147–53
- [200] Mallam A L, Morris E R and Jackson S E 2008 Exploring knotting mechanisms in protein folding *Proc. Natl Acad. Sci. USA* 105 18740–5
- [201] Mallam A L, Onuoha S C, Grossmann J G and Jackson S E 2008 Knotted fusion proteins reveal unexpected possibilities in protein folding *Mol. Cell* 30 642–8
- [202] Purta E, Kaminska K H, Kasprzak J M, Bujnicki J M and Douthwaite S 2008 YbeA is the m(3)Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA *Rna Soc.* 14 2234–44
- [203] Hsieh S J, Mallam A L, Jackson S E and Hsu S T 2013 Backbone NMR assignments of a topologically knotted protein in urea-denatured state *Biomol. NMR Assign.* 8 283–5
- [204] Andrews B T, Capraro D T, Sulkowska J I, Onuchic J N and Jennings P A 2013 Hysteresis as a marker for complex, overlapping landscapes in proteins J. Phys. Chem. Lett. 4 180–8
- [205] Wang P, Yang L, Liu P, Gao Y Q and Zhao X S 2013 Single-molecule detection reveals knot sliding in TrmD denaturation *Chemistry* 19 5909–16
- [206] Lim N C H and Jackson S E 2015 Mechanistic insights into the folding of knotted proteins *in vitro* and *in vivo J. Mol. Biol.* 427 248–58
- [207] King N P, Jacobitz A W, Sawaya M R, Goldschmidt L and Yeates T O 2010 Structure and folding of a designed knotted protein *Proc. Natl Acad. Sci. USA* 107 20732–7
- [208] Wang T, Arakawa H and Ikai A 2001 Force measurement and inhibitor binding assay of monomer and engineered dimer of bovine carbonic anhydrase B *Biochem. Biophys. Res. Commun.* 285 9–14
- [209] Andersson F I, Pina D G, Mallam A L, Blaser G and Jackson S E 2009 Untangling the folding mechanism of the 5(2)-knotted protein UCH-L3 *Febs J.* 276 2625–35
- [210] Wallin S, Zeldovich K B and Shakhnovich E I 2007 The folding mechanics of a knotted protein *J. Mol. Biol.* 368 884–93
- [211] Sulkowska J I, Sulkowski P and Onuchic J 2009 Dodging the crisis of folding proteins with knots *Proc. Natl Acad. Sci.* USA 106 3119–24
- [212] Sulkowska J I, Noel J K, Ramirez-Sarmiento C A, Rawdon E J, Millett K C and Onuchic J 2013 Knotting pathways in proteins *Biochem. Soc. Trans.* 41 523–7
- [213] Prentiss M C, Wales D J and Wolynes P G 2010 The energy landscape, folding pathways and the kinetics of a knotted protein *Plos Comput. Biol.* 6 e1000835
- [214] Tuszynska I and Bujnicki J M 2010 Predicting atomic details of the unfolding pathway for YibK, a knotted protein from the SPOUT superfamily J. Biomol. Struct. Dyn. 27 511–20

- [215] Sulkowska J I, Sulkowski P, Szymczak P and Cieplak M 2010 Untying knots in proteins J. Am. Chem. Soc. 132 13954–6
- [216] Faisca P F N, Travasso R D M, Charters T, Nunes A and Cieplak M 2010 The folding of knotted proteins: insights from lattice simulations *Phys. Biol.* 7 16009
- [217] Soler M A and Faisca P F N 2012 How difficult is it to fold a knotted protein? In silico insights from surface-tethered folding experiments *Plos One* **7** e52343
- [218] Skrbic T, Micheletti C and Faccioli P 2012 The role of nonnative interactions in the folding of knotted proteins *Plos Comput. Biol.* 8 e1002504
- [219] Covino R, Skrbić T, Beccara S A, Faccioli P and Micheletti C 2013 The role of non-native interactions in the folding of knotted proteins: insights from molecular dynamics simulations *Biomolecules* 4 1–19
- [220] Sulkowska J I, Sulkowski P and Onuchic J 2009 Jamming proteins with slipknots and their free energy landscape *Phys. Rev. Lett.* **103** 268103
- [221] He C, Genchev G Z, Lu H and Li H 2012 Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations J. Am. Chem. Soc. 134 10428–35
- [222] He C, Lamour G, Xiao A, Gsponer J and Li H 2014 Mechanically tightening a protein slipknot into a trefoil knot J. Am. Chem. Soc. 136 11946–55
- [223] Potestio R, Micheletti C and Orland H 2010 Knotted versus unknotted proteins: evidence of knot-promoting loops *PLoS Comput. Biol.* 6 e1000864
- [224] Krissinel E and Henrick K 2004 Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions Acta Crystallogr. D Biol. Crystallogr. 60 2256–68
- [225] Villeret V, Clantin B, Tricot C, Legrain C, Roovers M, Stalon V, Glansdorff N and Van Beeumen J 1998 The crystal structure of Pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures *Proc. Natl Acad. Sci. USA* **95** 2801–6
- [226] Sulkowska J I, Noel J K, Ramirez-Sarmiento C A, Rawdon E J, Millett K C and Onuchic J N 2013 Knotting pathways in proteins *Biochem. Soc. Trans.* 41 523–7
- [227] Covino R, Skrbić T, Beccara S A, Faccioli P and Micheletti C 2013 The role of non-native interactions in the folding of knotted proteins: insights from molecular dynamics simulations *Biomolecules* 4 1–19
- [228] Krobath H and Faisca P F 2013 Interplay between native topology and non-native interactions in the folding of tethered proteins *Phys. Biol.* **10** 016002
- [229] Skrbic T, Micheletti C and Faccioli P 2012 The role of non-native interactions in the folding of knotted proteins *PLoS Comput. Biol.* 8 e1002504
- [230] Onuchic J N and Wolynes P G 2004 Theory of protein folding *Curr. Opin. Struct. Biol.* **14** 70–5
- [231] Wasserman E 1960 The preparation of interlocking rings: a catenane J. Am. Chem. Soc. 82 4433–4
- [232] Beves J E, Campbell C J, Leigh D A and Pritchard R G 2013 Tetrameric cyclic double helicates as a scaffold for a molecular solomon link Angew. Chem. Int. Edn 52 6464–7
- [233] Schouwey C, Holstein J J, Scopelliti R, Zhurov O, Nagornov K O, Tsybin Y O, Smart O S, Bricogne G and Severin K 2014 Self-assembly of a giant molecular Solomon link from 30 subcomponents Angew. Chem. Int. Edn 53 11261–5
- [234] Chichak K S, Cantrill S J, Pease A R, Chiu S-H, Cave G W V, Atwood J L and Stoddart J F 2004 Molecular borromean rings *Science* **304** 1308–12
- [235] Forgan R S, Spruell J M, Olsen J-C, Stern C L and Stoddart J F 2009 Towards the stepwise assembly of molecular

borromean rings. A donor-acceptor ring-in-ring complex *J. Mex. Chem. Soc.* **53** 134–8

- [236] Leigh D A, Pritchard R G and Stephens A J 2014 A star of david catenane *Nat. Chem.* 6 978–82
- [237] Safarowsky O, Vogel E and Vögtle F 2000 Amide-based [3] catenanes and [2]catenanes with aliphatic chains *Eur. J. Org. Chem.* 2000 499–505
- [238] Crowley J D, Goldup S M, Lee A-L, Leigh D A and McBurney R T 2009 Active metal template synthesis of rotaxanes, catenanes and molecular shuttles *Chem. Soc. Rev.* 38 1530–41
- [239] Dietrich-Buchecker C O, Sauvage J-P and Kintzinger J P 1983 Une nouvelle famille de molecules: les metallocatenanes *Tetrahedron Lett.* 24 5095–8
- [240] Hamilton D G, Davies J E, Prodi L and Sanders J K M 1998 Synthesis, structure and photophysics of neutral pi-associated [2]catenanes *Chem.- A Eur. J.* 4 608–20
- [241] Beves J E 2012 Metal template synthesis of molecular knots and links *CHIMIA Int. J. Chem.* **66** 170–3
- [242] Lukin O and Vögtle F 2005 Knotting and threading of molecules: chemistry and chirality of molecular knots and their assemblies Angew. Chem. Int. Edn 44 1456–77
- [243] Frisch H L and Wasserman E 1961 Chemical topology J. Am. Chem. Soc. **83** 3789–95
- [244] Dietrich-Buchecker C O and Sauvage J-P 1989 A synthetic molecular trefoil knot Angew. Chem. Int. Edn Engl. 28 189–92
- [245] Seeman N C 1998 Nucleic acid nanostructures and topology Angew. Chem. Int. Edn 37 3220–38
- [246] Du S M, Stollar B D and Seeman N C 1995 A synthetic DNA molecule in three knotted topologies J. Am. Chem. Soc. 117 1194–200
- [247] Carina R F, Dietrich-Buchecker C and Sauvage J-P 1996 Molecular composite knots J. Am. Chem. Soc. 118 9110–6
- [248] Beves J E, Blight B A, Campbell C J, Leigh D A and McBurney R T 2011 Strategies and tactics for the metaldirected synthesis of rotaxanes, knots, catenanes and higher order links Angew. Chem. Int. Edn 50 9260–327
- [249] Cesario M, Dietrich-Buchecker C O, Gullhem J, Pascard C and Sauvage J-P 1985 Molecular-structure of a catenand and its copper(I) catenate—complete rearrangement of the interlocked macrocyclic ligands by complexation J. Chem. Soc. Chem. Commun. 1985 244–7
- [250] Sauvage J-P and Dietrich-Buchecker C 1999 Molecular knots- from early attempts to high-yield template synthesis Molecular Catenanes, Rotaxanes and Knots: a Journey Through the World of Molecular Topology ed J-P Sauvage and C Dietrich-Buchecker (Weinheim: Wiley-VCH)
- [251] Walba D M 1985 Topological stereochemistry Tetrahedron 41 3161–212
- [252] Dietrich-Buchecker C O, Sauvage J-P, De Cian A and Fischer J 1994 High-yield synthesis of a dicopper(I) trefoil knot containing 1, 3-phenylene groups as bridges between the chelate units J. Chem. Soc. Chem. Commun. 1994 2231–2
- [253] Dietrich-Buchecker C O, Rapenne G and Sauvage J-P 1997 Efficient synthesis of a molecular knot by copper(I)-induced formation of the precursor followed by ruthenium(II)-catalysed ring closing metathesis *Chem. Commun.* **1997** 2053–4
- [254] Rapenne G, Dietrich-Buchecker C O and Sauvage J-P 1999 Copper(I)- or Iron(II)-templated synthesis of molecular knots containing two tetrahedral or octahedral coordination sites J. Am. Chem. Soc. 121 994–1001
- [255] Perret-Aebi L-E, von Zelewsky A, Dietrich-Buchecker C and Sauvage J-P 2004 Stereoselective synthesis of a topologically chiral molecule: the trefoil knot Angew. Chem. Int. Edn 43 4482–5
- [256] Adams H, Ashworth E, Breault G A, Guo J, Hunter C A and Mayers P C 2001 Knot tied around an octahedral metal centre *Nature* 411 763

- [257] Sokolov V I 1973 Topological ideas in stereochemistry Russ. Chem. Rev. 42 452–63
- [258] Guo J, Mayers P C, Breault G A and Hunter C A 2010 Synthesis of a molecular trefoil knot by folding and closing on an octahedral coordination template *Nat. Chem.* 2 218–22
- [259] Barran P E, Cole H L, Goldup S M, Leigh D A, McGonigal P R, Symes M D, Wu J and Zengerle M 2011 Active-metal template synthesis of a molecular trefoil knot Angew. Chem. Int. Edn 50 12280–4
- [260] Ayme J-F, Gil-Ramirez G, Leigh D A, Lemonnier J-F, Markevicius A, Muryn C A and Zhang G 2014 Lanthanide template synthesis of a molecular trefoil knot J. Am. Chem. Soc. 136 13142–5
- [261] Carver F J, Hunter C A and Shannon R J 1994 Directed macrocyclisation reactions J. Chem. Soc. Chem. Commun. 1994 1277–80
- [262] Safarowsky O, Nieger M, Fröhlich R and Vögtle F 2000 A molecular knot with twelve amide groups- one-step synthesis, crystal structure, chirality *Angew. Chem. Int. Edn* 39 1616–8
- [263] Feigel M, Ladberg R, Engels S, Herbst-Irmer R and Fröhlich R 2006 A trefoil knot made of amino acids and steroids Angew. Chem. Int. Edn 45 5698–702
- [264] Ponnuswamy N, Cougnon F B L, Pantoş G D and Sanders J K M 2014 Homochiral and meso figure eight knots and a solomon link J. Am. Chem. Soc. 136 8243–51
- [265] Ashton P R, Matthews O A, Menzer S, Raymo F M, Spencer N, Stoddart J F and Williams D J 1997 Molecular Meccans, 27. A template-directed synthesis of a molecular trefoil knot *Liebigs Annalen der Chem.* 1997 2485–94
- [266] Woods C R, Benaglia M, Toyota S, Hardcastle K and Siegel J S 2001 Trinuclear copper(I)-bipyridine triskellion: template/cantilever control of coordination complex stereochemistry in a trefoil knot precursor Angew. Chem. Int. Edn 40 749–51
- [267] Arias K I, Zysman-Colman E, Loren J C, Linden A and Siegel J S 2011 Synthesis of a D3-symmetric 'trefoil' knotted cyclophane *Chem. Commun.* 47 9588–90
- [268] Prakasam T, Lusi M, Elhabiri M, Platas-Iglesias C, Olsen J-C, Asfari Z, Cianférani-Sanglier S, Debaene F, Charbonnière LJ and Trabolsi A 2013 Simultaneous selfassembly of a [2]catenane, a trefoil knot, and a solomon link from a simple pair of ligands *Angew. Chem. Int. Edn* 125 10140–4
- [269] Ayme J-F, Beves J E, Leigh D A, McBurney R T, Rissanen K and Schultz D 2011 A synthetic molecular pentafoil knot *Nat. Chem.* 4 15–20
- [270] Rapenne G, Dietrich-Buchecker C and Sauvage J-P 1996 Resolution of a molecular trefoil knot J. Am. Chem. Soc. 118 10932–3
- [271] Hasenknopf B and Lehn J M 1996 Trinucleaer double helicates of iron(II) and nickel(II)- self-assembly and resolution into helical enantiomers *Helv. Chim. Acta* 79 1643–50
- [272] Meyer M, Albrecht-Gary A-M, Dietrich-Buchecker C O and Sauvage J-P 1997 Dicopper(I) trefoil knots: topological and structural effects on the demetalation rates and mechanism J. Am. Chem. Soc. 119 4599–607

- [273] Meluzzi D, Smith D E and Arya G 2010 Biophysics of knotting Annual Review of Biophysics vol 39 ed D C Rees and J R Williamson (Palo Alto, CA: Annual Reviews) pp 349–66
- [274] Trigueros S and Roca J 2007 Production of highly knotted DNA by means of cosmid circularization inside phage capsids *BMC Biotechnol.* 7 94
- [275] Bao X R, Lee H J and Quake S R 2003 Behaviour of complex knots in single DNA molecules *Phys. Rev. Lett.* 91 265506
- [276] Hagerman P J 1988 Flexibility of DNA Annu. Rev. Biophys. Biophys. Chem. 17 265–86
- [277] Murphy M C, Rasnik I, Cheng W, Lohman T M and Ha T 2004 Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy *Biophys. J.* 86 2530–7
- [278] Kebbekus P, Draper D E and Hagerman P 1995 Persistence length of RNA *Biochemistry* 34 4354–7
- [279] Bizarro C V, Alemany A and Ritort F 2012 Non-specific binding of Na<sup>+</sup> and Mg2<sup>+</sup> to RNA determined by force spectroscopy methods *Nucl. Acids Res.* 40 6922–35
- [280] Bucka A and Stasiak A 2002 Construction and electrophoretic migration of single-stranded DNA knots and catenanes *Nucl. Acids Res.* 30 e24
- [281] Seeman N C 2007 An overview of structural DNA nanotechnology Mol. Biotechnol. 37 246–57
- [282] Seeman N C 1998 DNA nanotechnology: novel DNA constructions Annu. Rev. Biophys. Biomol. Struct. 27 225–48
- [283] Seeman N C 1996 The design and engineering of nucleic acid nanoscale assemblies *Curr. Opin. Struct. Biol.* 6 519–26
- [284] Seeman N C 2003 Biochemistry and structural DNA nanotechnology: an evolving symbiotic relationship *Biochemistry* 42 7259–68
- [285] Wang H, Du S M and Seeman N C 1993 Tight singlestranded DNA knots J. Biomol. Struct. Dyn. 10 853–63
- [286] Au-Yeung H Y, Pantos G D and Sanders J K 2011 Dynamic combinatorial donor-acceptor catenanes in water: access to unconventional and unexpected structures J. Org. Chem. 76 1257–68
- [287] Lam R T S, Belenguer A, Roberts S L, Naumann C, Jarrosson T, Otto S and Sanders J K M 2005 Amplification of acetylcholine-binding catenanes from dynamic combinatorial libraries *Science* 308 667–9
- [288] Au-Yeung H Y, Dan Pantos G and Sanders J K M 2009 Dynamic combinatorial synthesis of a catenane based on donor-acceptor interactions in water *Proc. Natl Acad. Sci.* USA 106 10466–70
- [289] Micheletti C, Di Stefano M and Orland H 2015 Absence of knots in known RNA structures *Proc. Natl Acad. Sci. USA* 112 2052–7
- [290] Wüst T, Reith D and Virnau P 2015 Sequence determines degree of knottedness in a coarse-grained protein model *Phys. Rev. Lett.* **114** 028102
- [291] Virnau P, Rieger F C and Reith D 2013 Influence of chain stiffness on knottedness in single polymers *Biochem. Soc. Trans.* 41 528–32
- [292] Cebrian J et al 2015 Electrophoretic mobility of supercoiled, catenated and knotted DNA molecules Nucl. Acids Res. 43 e24