HIGHLIGHTS

** 42. Lim NC, Jackson SE: **Mechanistic insights into the folding of knotted proteins in vitro and in vivo**. *Journal of Molecular Biology* 2015, **427**:248–258.

Experimental study of the folding of nascent chains of trefoil-knotted methyl-transferases and engineered fusion proteins. The results clearly demonstrate that threading occurs through motions of the C-terminus and that chaperonins are most likely to accelerate folding through a mechanism involving backtracking or partial unfolding of a kinetically trapped intermediate.

** 59. Ziegler F, Lim NCH, Mandal SS, Pelz B, Ng W-P, Schlierf M, Jackson SE, Rief M: **Knotting and unknotting of a protein in single molecule experiments.** *Proc. Natl. Acad. Sci. U.S.A.* 2016, **113**:7533–7538.

The first single-molecule study that establishes the complexity of the energy landscape for folding of a knotted protein, as well as the impact of different knotted denatured states on folding. In addition, it characterizes a particularly large 5_2-knot in the denatured ensemble which has implications for cellular degradation pathways.

* 26. Pandey N, Kuypers BE, Nassif B, Thomas EE, Alnahhas RN, Segatori L, Silberg JJ: **Tolerance of a Knotted Near-Infrared Fluorescent Protein to Random Circular Permutation**. *Biochemistry* 2016, **55**:3763–3773.

This experimental work studies how the knotted topology of bacteriophytochrome photoreceptors is maintained in circular permutants.

* 29. Micheletti C, Di Stefano M, Orland H: **Absence of knots in known RNA structures.** *Proc. Natl. Acad. Sci. U.S.A.* 2015, **112**:2052–2057.

The study reports on the lack of knots in currently-available RNA structures and contrasts it to the occurrence of knots in proteins.

** 30. Liu L, Hyeon C: **Contact Statistics Highlight Distinct Organizing Principles of Proteins and RNA.** *Biophysical Journal* 2016, **110**:2320–2327.

This work presents a systematic characterization of local and non-local contact propensities in proteins and RNAs and relates it to their apparent different propensity to be knotted.

** 32. Wüst T, Reith D, Virnau P: **Sequence determines degree of knottedness in a coarse-grained protein model.** *Phys. Rev. Lett.* 2015, **114**:028102.

This computational study explores the principles through which the knot type and location can be encoded by the primary sequence of model knotted proteins.

* 53. Chwastyk M, Cieplak M: **Multiple folding pathways of proteins with shallow knots and co-translational folding**. *J. Chem. Phys.* 2015, **143**:045101.

A coarse-grained model is used to study the extent to which cotranslational folding can assist knot formation.

** 60. Wojciechowski M, Gómez-Sicilia À, Carrion-Vazquez M, Cieplak M: **Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins.** *Mol Biosyst* 2016, doi:10.1039/c6mb00214e.

A coarse-grained model is used to model the degradation of knotted proteins by pore translocation.

** 61. Suma A, Rosa A, Micheletti C: **Pore Translocation of Knotted Polymer Chains: How Friction Depends on Knot Complexity**. *ACS Macro Lett.* 2015, **4**:1420–1424.

Coarse-grained model simulations are used to study the compliance of different knot types to translocation through a narrow pore. Jamming is observed only at high driving forces and limitedly to twist knots.

How to fold intricately: using theory and experiments to unravel the properties of knotted proteins

S. E. Jackson[(1,*)], A. Suma[(2)] and C. Micheletti[(2,*)]

[(1)] Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

[(2)] SISSA, International School for Advanced Studies, via Bonomea 265, I-34136 Trieste, Italy

Abstract

Over the years, various experimental and theoretical advancements have helped understand the role of thermodynamic, kinetic and active (chaperone-aided) effects in coordinating folding steps required to achieving a knotted native state. Here, we review such developments by paying particular attention to the complementarity of experimental and computational studies. Key open issues that could be tackled with either or both approaches are finally pointed out.

Introduction

Despite the early evidence of a shallowly knotted carbonic anhydrase structure[1], the conviction that proteins had to be knot-free to avoid kinetic traps during folding held until the mid ’90s. At that time, a series of systematic surveys [2,3] of the growing protein databank (PDB)[4] proved unambiguously the occurrence of deeply knotted proteins.

We now know that knotted proteins are uncommon, but not exceptionally rare as they account for about 1% of all PDB entries[5-7]. Their *in vivo* abundance in specific contexts can be highly significant too. For instance, the human ubiquitin C-terminal hydrolase isoform 1 (UCH-L1), that is knotted, accounts for about 2-5% of soluble protein in neurons[8,9].

* Correspondence: sej13@cam.ac.uk, michelet@sissa.it

Both experimental and theoretical approaches have been used to understand the driving forces that coordinate the folding steps leading to knotted native states [5,10-14]. Experiments, unlike present-day simulations, can probe timescales that are sufficiently long to follow the spontaneous folding process. At the same time, folding simulations currently outcompete experiments for the level of detail they can provide of the folding routes. This review aims at conveying such complementarity by focusing on specific aspects that have been tackled with either or both strategies.

Overview of knotted proteins

A general overview of currently known knotted protein structures is given in the up-to-date non-redundant list of knotted representatives shown in Table I.

The range of functional families in Table I is noticeably broad. Indeed, no general functionally oriented rationale for the occurrence of knots in proteins has been found yet, although knots are known to occur in membrane proteins, and could be instrumental to avoid degradation and/or enhance the thermodynamic, kinetic or mechanical stability of proteins [10,15-19]. Indeed, it is intriguing that, in many cases, the knot is close to, or encompasses the active sites of several entangled enzymes, such as RNA methyltransferases [18,20,21], carbamoyltransferases [22], and bacterial phytochromes[23-26].

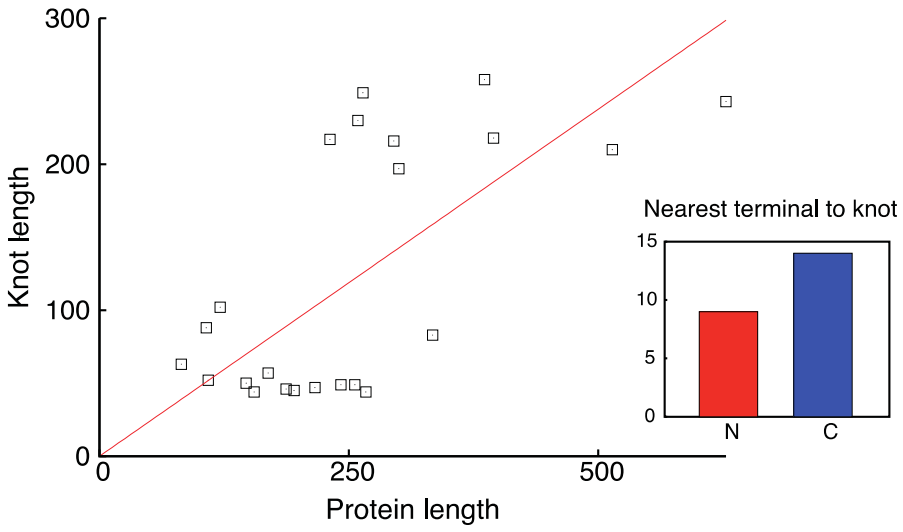


FIGURE 1. **Length and depth of knots in proteins.** The scatter plot presents the knot length versus protein length for the 23 minimally redundant representatives in Table I. The Kendall's correlation coefficient, tau = 0.31 and the one-sided *p*-value is 0.018. Of these 23 instances, 9 have the knotted region closest to the N terminus and 14 to the C one, see inset.

The physical knots listed in Table I cover four different topologies: the 3_1, 4_1, 5_2 and 6_1 knots. These are the simplest instances of twist knots that can be tied or untied with a single, suitably chosen, strand threading or passage. Non-twist knots with similar complexity, such as the 5_1 torus knot, are probably not observed because their folding would be more challenging, requiring at least two strand passages or threading events to be fully tied or untied[5].

Protein or protein function	**PDB code**	**Knot type**	**PDB length**	**Knot length**	**N-terminal depth**	**C-terminal depth**
Carbonic Anhydrase	3MDZ:A	3_1	259	230	26	3
RNA Methyltransferase	1X7O:A	3_1	267	44	190	33
	4H3Z:B	3_1	256	49	89	118
	4CND:A	3_1	169	57	79	33
	4E8B:A	3_1	242	49	164	29
	3O7B:A	3_1	216	47	143	26
	4JAK:A	3_1	155	44	77	34
	4JWF:A	3_1	187	46	98	43
	2QMM:A	3_1	195	45	124	26
S-Adenosylmethionine Synthetase	4ODJ:A	3_1	386	258	16	112
Carbamoyltransferase	3KZK:A	3_1	334	83	169	82
Hypothetical RNA Methyltransferase	1O6D:A	3_1	147	50	67	30
Hypothetical Protein MJ0366	2EFV:A	3_1	82	63	10	9
H+/Ca2+ Exchanger	4KPP:A	3_1	395	218	72	105
Na+/Ca2+ Exchanger	5HWY:A	3_1	300	197	33	70
N-Acetylglucosamine Deacetylase	5BU6:A	3_1	264	249	10	5
DNA Binding Protein	2RH3:A	3_1	121	102	7	12
	4LRV:A	3_1	107	88	8	11
Metal Binding (Zinc-Finger)	2K0A:A	3_1	109	52	21	36
Bacteriophytochrome	4GW9:A	4_1	628	243	20	365
Ketol-Acid Reductoisomerase	1QMG:A	4_1	514	210	236	68
Ubiquitin Carboxy-Terminal Hydrolase	2LEN:A	5_2	231	217	2	12
α-Haloacid Dehalogenase I DehI	3BJX:B	6_1	295	216	59	20

TABLE I. **Representative knotted proteins.** This up-to-date, non-redundant list of knotted representatives is based on a PDB survey specifically carried out for this review. The entries are presented in increasing complexity of the knot and, for each knot type, they are listed in order of decreasing representative molecular weight (number of represented PDB

entries at 10% sequence identity threshold). The complete list of knotted PDB entries, including a few where knots are likely artifacts due to limited structural resolution, are provided as supplementary information. Besides these physical knots other forms of protein entanglement have been reported. These include slipknots[13,27], which are observed when a the threading end is folded back onto itself, such that a knot is formed by part of, but not the full-length, chain, and pierced-lasso bundles that are observed when a part of the chain is threaded through a loop formed by a disulphide bond[28].

A key general question is whether the degree of entanglement observed in proteins differs from that of other compact, globular polymers [29-32]. In this regard, Lua and Grosberg[33] showed that naturally occurring proteins are knotted significantly less than equivalent models of globular homopolymers (which, unlike proteins, lack a defined native state). These results are compatible with the intuition that knots have been selected against in naturally occurring proteins, though not ruled out entirely. Similarly to general polymer models, however, proteins do exhibit a significant correlation between the length of the knotted region, and the overall protein length, see Fig. 1.

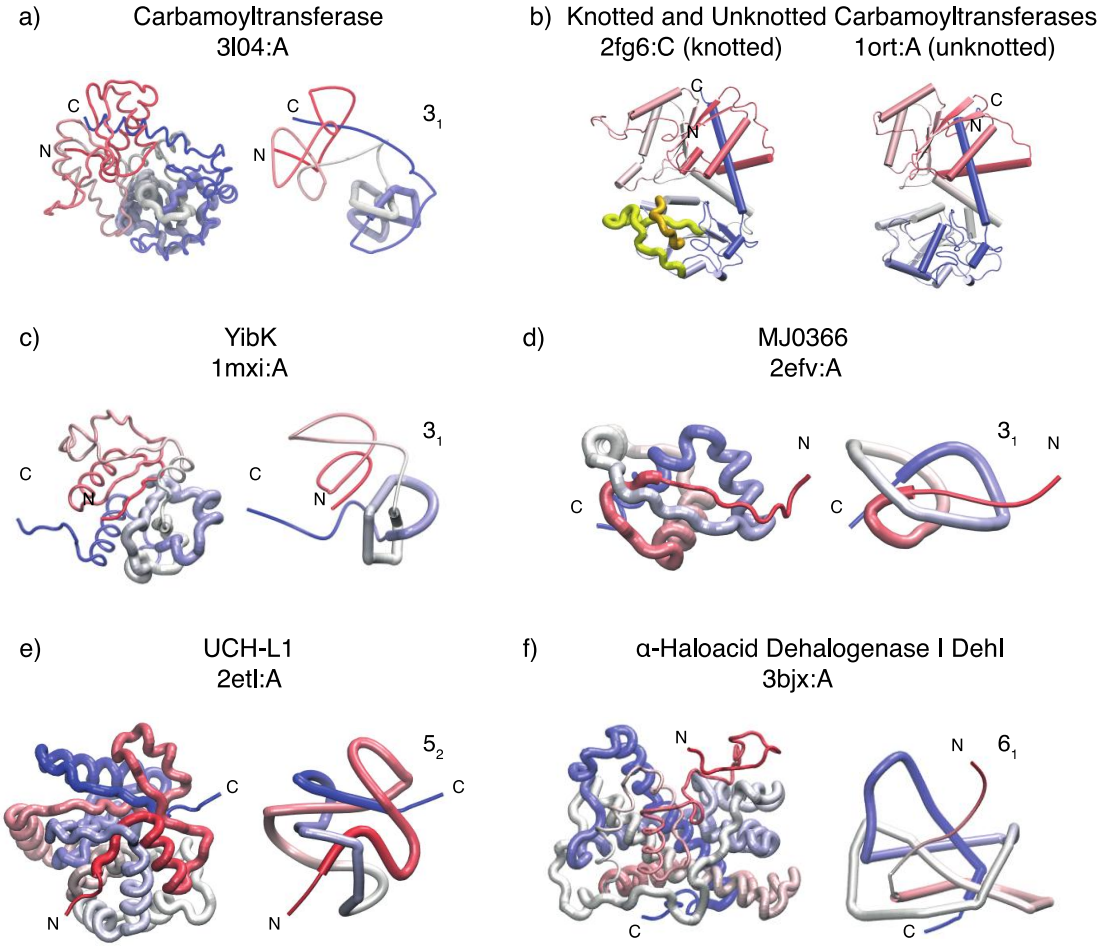


FIGURE 2. **Selected examples of knotted proteins**. In panels a, c-f, a smoothed structural representation is used to highlight the knot. Panel (b) presents a knotted/unknotted pair of carbamoyltransferases[6]. A virtual excision of either or both of the highlighted loops (colored in yellow and orange) unties the knotted variant.

Knotted and unknotted carbamoyltransferases

A systematic sequence-based comparison of knotted and unknotted proteins [6] showed that knotted carbamoyltransferases, see Fig. 2a, occupy a specific phylogenetic branch off the main trunk of unknotted precursors. Structurally, the key difference is the presence of additional short loops in the knotted variants [6,22], see Fig. 2b. Other types of knotted proteins, including UCH-L1, also have unknotted counterparts differing in the lack of short loops[6]. This suggests that mutation by loop addition may have been a recurrent step in the evolution of knotted proteins from unknotted precursors.

Skrbic et al.[34] took advantage of the sequence and structural similarities of knotted and unknotted carbamoyltransferases to compare their early folding stages and track differences associated with knotting. For simulations based on a pure Go model, that exclusively rewards native interactions, no propensity to knot formation was observed for either variant. However, when the Go-model was complemented with non-native quasi-chemical interactions, a small but systematic propensity to knot (about 0.5%) was found for the natively-knotted variant, while it remained negligible for the unknotted one. The knotting events typically involved the threading of the hydrophobic C-terminus through loosely structured loop regions.

The involvement of the C-terminus in topology-changing events was also seen in Go-model unfolding simulations of unknotted and knotted carbamoyltransferases[16]. The knotted variants were found to be more resilient to mechanical unfolding as well.

Yibk and YbeA

The bacterial homodimeric proteins YibK and YbeA are, so far, the smallest known members of the α/β-knot methyltransferases (MTases) family[18,35]. Their single-domain monomers are about 160 residues long and accommodate a trefoil knot at a depth of ca. 40 residues from the C-terminus, see Fig. 2c. They both fold relatively fast from their chemically denatured states [36,37], however, it is now well established that high concentrations of chemical denaturant do not remove their native 3_1-knotted topology. This remarkable and unexpected property was proven by trapping the denatured state topology

by ligating the protein termini and observing that the resulting cyclized proteins could refold to functional, knotted native states[38]. Later studies on another knotted MTase from *Thermotoga maritima* [39] confirmed that these knotted structures require several weeks under highly denaturing conditions to unfold and untie themselves[40].

Even more remarkable is the fact that both YibK and YbeA proteins are capable of attaining the correct knotted native topology even when molecular plugs are attached to either termini [41]. In the most recent demonstration of this, an *in vitro* transcription-translation system was used to study the *de novo* folding of YibK and YbeA as well as fusion variants obtained by attaching the rapidly folding ThiS domain at either or both termini [42]. The folding rate was slowed down by up to a factor of three when the ThiS plug was attached to the C terminus. No appreciable slowing down was observed when the plug was present solely at the N terminus. The results suggest that the rate-determining folding and knotting events take place at the C terminus, close to the native location of the knotted region.

One further aspect relevant for the *in vivo* folding of the knotted methyltransferases is the role of chaperonins. These have been shown to speed up the folding rates of YibK and YbeA by more than an order of magnitude[42,43]. Although the details of the chaperonins' action is still unclear, Jackson and coworkers proposed that they may help unfolding a highly native-like, but unknotted, misfolded state that would otherwise be kinetically trapped. Backtracking from similarly misfolded conformations had previously been observed in simulations[44].

Besides experiments, several folding simulations were carried out on MTases. The YibK study of Wallin et al. [45] was the first to address numerically the folding of a knotted protein. When using a pure Go-model, their folding simulations resulted in about 80% formation of native contacts, but not more. This happened because the near-native states were too compact to allow for threading events. However, after introducing non-native attractive interactions between the middle and C-terminal regions of the chain, the knotted native state was reached in all folding attempts[45]. These non-native interactions were crucial in establishing the correct topology to thread the chain through a loop. Interestingly, two different types of folding routes were observed, one where knotting occurred early and the other late (20% and 80% of native contacts formed, respectively).

In the studies of Sulkowska et al. [44] and Prentiss et al. [46] pure Go-models were used to generate folding trajectories from fully unfolded initial states. They concluded that pure native-centric potentials suffice to drive the folding process towards the lowest-energy, knotted, native state. However, in the

Sulkowska study, the yield of successful trajectories was low, around 1-2%, and in the Prentiss study, in order to estimate the fastest speed possible for folding a knotted protein, a minimal and shallowly knotted structure was simulated. In both cases, multiple pathways were observed involving the formation of the knot either at the N or C terminus [44,46] and knotting events occurred either at early [46] or late [44] stages of folding. The knotting modes involved either a direct threading or a slipknotting event. The relative weight of the events depending significantly on the level of structural detail in the model. [44]

A recent and interesting twist to the problem was addressed by Cieplak et al. [47] who used an optimized Go model to simulate the folding of a nascent chain of YibK. The study showed that aforementioned low yield of successful folding trajectories could be dramatically enhanced by including co-translational folding effects. In fact, the common knotting event consisted of the formation of a slipknot whilst the C-terminal region of the chain was still attached to the model ribosome, the knot was only able to form in full after the chain was released from it. The model of [47] adopted a tolerant criterion to define native interactions that included the key folding-promoting contacts of Wallin et al. [45], and therefore no *ad hoc* non-native contacts were needed to drive the correct folding either cotranslationally, or spontaneously. Interestingly, in the latter case, no slipknotting events were observed.

MJ0366

The homodimeric protein MJ0366 from *Methanocaldococcus jannaschii* is the smallest known knotted protein. Its monomers are 92 residues in length and feature a shallow trefoil knot at a depth of only 10 residues from the C-terminus, see Fig. 2d.

The folding mechanism of MJ0366 has recently been probed experimentally using a number of techniques [48], and evidence found for a highly structured, monomeric, on-pathway intermediate. It is highly likely that this intermediate is not yet knotted as it forms within a millisecond, towards the upper limit of folding rates observed for very small, unknotted proteins. This is followed by a slower second step that involves further folding, and likely knotting, as well as association to form the dimer, similar to results on the folding of 5_2-knotted UCHs.

The limited protein length and knot depth of MJ0366 has made it the focus of several computational studies. Computational models with different levels of structural detail, force fields and initial conditions have been employed [47,49-51]. In the study of Noel et al. [50], based on coarse-grained and atomistic native-centric models, knot formation was observed through both threading

and slipknotting mechanisms, the latter being dominant both below the folding temperature and upon extending the C-terminus. The study mostly focused on a single monomer of MJ0366 because, within the native-centric scheme used, knotting of the monomers precedes the formation of the dimer[50]. This conclusion, however, is in contrast to our interpretation of the experimental HDX data, which indicates that the region of the protein involved in knot formation is not highly structured in the intermediate state[48].

A later study by Beccara et al. [49] used, for the first time, a realistic atomistic force field, i.e. non-native centric, to study the folding of a monomer of MJ0366. In this case, computational demand was reduced by using a ratchet-and-pawl scheme to accelerate the evolution of the system to the native state. Only 1% of trajectories successfully reached the knotted native state, knotting occurred *via* direct threading of a loop formed by the earlier formation of the β-sheet, a mechanism later observed in the coarse-grained folding simulations of Najafi and Potestio too[52]. By comparison, slipknotting events were rare and, additionally, a novel mechanism, involving loop-flipping was reported. These studies were later followed by those of refs. [51,53] in which MJ0366 folding was studied from different specific initial conditions. In ref. [51] Noel et al. used an unbiased atomistic simulation to study the dynamic evolution from configurations that were unfolded, though slipknotted. The study of Chwastyk et al. [53] instead, used an optimised Go-model to study the co-translational folding of MJ0366. They found that the fraction of successful folding trajectories increased dramatically under nascent, co-translational conditions. Knotting events involved direct threading, slipknotting and loop-flipping, similarly to ref. [49] albeit in different proportions, and a further two-loop knotting event was added to the list of mechanisms observed computationally. Interesting, the study reported a reduction in knotting efficiency upon extension of the C terminus, unlike ref. [51] .

HP0242: a designed trefoil-knotted protein

In 2010, the Yeates group successfully designed a monomeric 3_1-knotted protein from the highly entwined homodimer HP0242 [54], thus demonstrating a potential pathway for how knotted proteins might have evolved from unknotted precursors. In experimental folding studies, the designed chain misfolded into a compact, probably unknotted, state before a slow transition, likely involving partial unfolding and knotting, to the knotted native state. Further kinetic studies by the Hsu group suggested that the designed protein folds through multiple intermediates states, only one of which can lead to productive folding [55].

Two computational studies on the knotted HP0242 variant exist. The first study [56], found a surprising lack of deep topological traps using a coarse-grained structure-based model. However, some aspects of the simulations did mimic the experimental

results such as the sensitive temperature dependence of successful folding trajectories[56]. The other study [57] used an improved atomic-interaction based coarse-grained model and reported a very high folding success rate, 96%. As had been shown before [56], the folding of the designed knotted protein was considerably slower that its unknotted counterpart. A number of intermediate states were observed including an off-pathway misfolded state that lacked the knot, consistent with the experiments.

Complex knotted proteins: UCHs and DehI

The most complex protein knots studied experimentally are from the ubiquitin C-terminal hydrolase (UCH) family, and have a 5_2 topology. The knot, located only a few residues from the N-terminus, has been hypothesized to be instrumental in the protein avoiding proteasomal degradation. The folding pathways of two isoforms UCH-L1 (associated with Parkinson's Disease and shown in Fig. 2e) and UCH-L3 have been probed with several techniques and very similar results obtained. Refolding after unfolding in chemical denaturants is fully reversible[8] and proceeds *via* two parallel pathways, each with a metastable intermediate [8,9,58]. The most recent studies of UCH-L1, based on NMR HDX [8], concluded that the two intermediates retain much of the β-sheet core, but differ in the degree to which flanking α-helices are packed against the β-sheet. Although somewhat circumstantial in nature, the results of this study also suggested that neither intermediate was knotted, consistent with the rate-limiting step being conversion of the intermediate to the native state, and associated with the threading event required to establish the final 5_2 native topology.

Recently, optical tweezers were used on UCH-L1 to take it to three different unfolded states: unknotted 0, 3_1 and 5_2 knotted and refolding from such states was measured [59]. This study showed that threading to form either a 3_1- or 5_2-knotted state slowed folding as expected. However, the inferred energy landscape was much more complex than previously envisioned, as many on- and off-pathway intermediate states were populated during unfolding and refolding. Furthermore, at low/moderate forces the 5_2-knotted region of the denatured state was much larger than expected, spanning about 40 residues. These results may have implications for the cellular degradation of this class of protein.

The latter problem has been recently tackled computationally in ref. [60]. In this study, the proteasome was simply represented by an effective potential along with constant and periodic pulling forces. Coarse-grained models of a 3_1-

knotted protein were used and the results showed that the knot can hinder or even jam the proteosomal machinery. It is likely that 5_2-knotted proteins such as UCH-L1, which have much larger twist knots in their denatured states. will have an even greater effect, as observed in the systematic pore translocation computational study of ref [61].

To our knowledge, no folding simulations have been carried out yet for UCHs. However, Faisca and coworkers have modeled the folding kinetics towards compact structures with a 5_2 topology on a lattice with pre-assigned geometry and topology [62,63]. They found that target structures with 3_1 topology were significantly more accessible and stable than 5_2 ones, that is interesting because both knot types can be tied/untied with a single strand passage.

Native-centric (Go-model) folding simulations have nevertheless been carried out for an alpha-haloacid dehalogenase DehI that accommodates a 6_1 knot at a depth of about 20 residues from the C terminus [7], see Fig. 2e. The analysis of the successful folding trajectories (6 out of 1000) showed that the complex native topology was achieved via slipknotting and depended on a large-twisted loop flipping over a smaller twisted loop.

Conclusions

In our view, the current understanding of knotted proteins could be significantly advanced by tackling the following questions:

What is the role of native and non-native contacts for the spontaneous folding of knotted proteins? In addition to the native structure, does the primary sequence encode for non-native interactions that are useful to avoid kinetic traps?

Do proteins tied in non-twist or composite knots exist? The spontaneous knotting dynamics of general polymer models indicates that the 5_1 torus knot and the "double" trefoil knot, 3_1-3_1 are kinetically accessible and have a lower, but still comparable incidence to 3_1 and 5_2 knots [64,65]. Can these knots be observed at all in naturally occurring proteins?

Is it possible to excise few short loops from a knotted protein and obtain unknotted variants? This fascinating possibility is suggested by structural comparative studies [6], but its viability has not yet been tested experimentally.

Most likely, future breakthroughs in these open issues will be fostered by a tighter integration of the complementary strengths of computational and

experimental approaches.

Acknowledgements

We acknowledge support from the Italian Ministry of Education, grant PRIN 2010HXAW77.

REFERENCES

1. Richardson JS: **Beta-sheet topology and the relatedness of proteins**. *Nature* 1977, **268**:495–500.

2. Mansfield ML: **Are there knots in proteins?** *Nat Struct Mol Biol* 1994, **1**:213–214.

3. Taylor WR: **A deeply knotted protein structure and how it might fold**. *Nature* 2000, **406**:916–919.

4. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J: **The Protein Data Bank and the challenge of structural genomics.** *Nature Struct Biol* 2000, **7 Suppl**:957–959.

5. Taylor WR: **Protein knots and fold complexity: Some new twists**. *Computational Biology and Chemistry* 2007, **31**:151–162.

6. Potestio R, Micheletti C, Orland H: **Knotted vs. unknotted proteins: evidence of knot-promoting loops.** *PLoS Comput Biol* 2010, **6**:e1000864.

7. Bolinger D, Sułkowska JI, Hsu H-P, Mirny LA, Kardar M, Onuchic JN, Virnau P: **A Stevedore's protein knot**. *PLoS Comput Biol* 2010, **6**:e1000731.

8. Lou S-C, Wetzel S, Zhang H, Crone EW, Lee Y-T, Jackson SE, Hsu S-TD: **The Knotted Protein UCH-L1 Exhibits Partially Unfolded Forms under Native Conditions that Share Common Structural Features with Its Kinetic Folding Intermediates**. *Journal of Molecular Biology* 2016, **428**:2507–2520.

9. Andersson FI, Werrell EF, McMorran L, Crone WJ, Das C, Hsu S-TD, Jackson SE: **The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1**. *Journal of Molecular Biology* 2011, **407**:261–272.

10. Virnau P, Mallam A, Jackson S: **Structures and folding pathways of topologically knotted proteins**. *J. Phys.: Condens. Matter* 2010, **23**:033101.

11. Lim NCH, Jackson SE: **Molecular knots in biology and chemistry.** *J. Phys.: Condens. Matter* 2015, **27**:354101.

12. Mallam AL: **How does a knotted protein fold?** *FEBS Journal* 2009, **276**:365–375.

13. Yeates TO, Norcross TS, King NP: **Knotted and topologically complex proteins as models for studying folding and stability**. *Current Opinion in Chemical Biology* 2007, **11**:595–603.

14. Faísca PF: **Knotted proteins: A tangled tale of Structural Biology**. *Computational and structural biotechnology journal* 2015, **13**:459–468.

15. Nureki O, Watanabe K, Fukai S, Ishii R, Endo Y, Hori H, Yokoyama S: **Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme.** *Structure/Folding and Design* 2004, **12**:593–602.

16. Sulkowska JI, Sulkowski P, Szymczak P, Cieplak M: **Stabilizing effect of knots on proteins**. *Proc Natl Acad Sci USA* 2008, **105**:19714–19719.

17. Sulkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A: **Conservation of complex knotting and slipknotting patterns in proteins**. *Proc Natl Acad Sci USA* 2012, **109**:E1715–E1723.

18. Nureki O, Shirouzu M, Hashimoto K, Ishitani R, Terada T, Tamakoshi M, Oshima T, Chijimatsu M, Takio K, Vassylyev DG, et al.: **An enzyme with a deep trefoil knot for the active-site architecture.** *Acta Crystallogr D Biol Crystallogr* 2002, **58**:1129–1137.

19. Sayre TC, Lee TM, King NP, Yeates TO: **Protein stabilization in a highly knotted protein polymer**. *Protein Engineering Design and Selection* 2011, **24**:627–630.

20. Ochi A, Makabe K, Yamagami R, Hirata A, Sakaguchi R, Hou Y-M, Watanabe K, Nureki O, Kuwajima K, Hori H: **The Catalytic Domain of Topological Knot tRNA Methyltransferase (TrmH) Discriminates between Substrate tRNA and Nonsubstrate tRNA via an Induced-fit Process**. *Journal of Biological Chemistry* 2013, **288**:25562–25574.

21. Swinehart WE, Jackman JE: **Diversity in mechanism and function of tRNA methyltransferases**. *RNA Biology* 2015, **12**:398–411.

22. Virnau P, Mirny LA, Kardar M: **Intricate Knots in Proteins: Function and Evolution**. *PLoS Comput Biol* 2006, **2**:e122.

23. Wagner JR, Brunzelle JS, Forest KT, Vierstra RD: **A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome**. *Nature* 2005, **438**:325–331.

24. Bornschlögl T, Anstrom DM, Mey E, Dzubiella J, Rief M, Forest KT: **Tightening the Knot in Phytochrome by Single-Molecule Atomic Force Microscopy**. *Biophysical Journal* 2009, **96**:1508–1514.

25. Stepanenko OV, Bublikov GS, Stepanenko OV, Stepanenko OV, Shcherbakova DM, Verkhusha VV, Turoverov KK, Kuznetsova IM: **A knot in the protein structure – probing the near-infrared fluorescent protein iRFP designed from a bacterial phytochrome**. *The FEBS Journal* 2014, **281**:2284–2298.

26. Pandey N, Kuypers BE, Nassif B, Thomas EE, Alnahhas RN, Segatori L, Silberg JJ: **Tolerance of a Knotted Near-Infrared Fluorescent Protein to Random Circular Permutation**. *Biochemistry* 2016, **55**:3763–3773.

27. King NP, Yeates EO, Yeates TO: **Identification of Rare Slipknots in Proteins and Their Implications for Stability and Folding**. *Journal of Molecular Biology* 2007, **373**:153–166.

28. Haglund E, Sulkowska JI, Noel JK, Lammert H, Onuchic JN, Jennings PA: **Pierced Lasso Bundles Are a New Class of Knot-like Motifs**. *PLoS Comput Biol* 2014, **10**:e1003613.

29. Micheletti C, Di Stefano M, Orland H: **Absence of knots in known RNA structures.** *Proc. Natl. Acad. Sci. U.S.A.* 2015, **112**:2052–2057.

30. Liu L, Hyeon C: **Contact Statistics Highlight Distinct Organizing Principles of Proteins and RNA.** *Biophysical Journal* 2016, **110**:2320–2327.

31. Grosberg AY: **Ensemble View of RNAs and Proteins: Loops, Knots, Territories, and Evolution.** *Biophysical Journal* 2016, **110**:2289–2290.

32. Wüst T, Reith D, Virnau P: **Sequence determines degree of knottedness in a coarse-grained protein model.** *Phys. Rev. Lett.* 2015, **114**:028102.

33. Lua RC, Grosberg AY: **Statistics of Knots, Geometry of Conformations, and Evolution of Proteins**. *PLoS Comput Biol* 2006, **2**:e45.

34. Škrbić T, Micheletti C, Faccioli P: **The role of non-native interactions in the folding of knotted proteins**. *PLoS Comput Biol* 2012, **8**:e1002504.

35. Michel G, Sauvé V, Larocque R, Li Y, Matte A, Cygler M: **The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot.** *Structure/Folding and Design* 2002, **10**:1303–1315.

36. Mallam AL, Jackson SE: **A Comparison of the Folding of Two Knotted Proteins: YbeA and YibK**. *Journal of Molecular Biology* 2007, **366**:650–665.

37. Mallam AL, Jackson SE: **Probing Nature's Knots: The Folding Pathway of a Knotted Homodimeric Protein**. *Journal of Molecular Biology* 2006, **359**:1420–1436.

38. Mallam AL, Rogers JM, Jackson SE: **Experimental detection of knotted conformations in denatured proteins.** *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:8189–8194.

39. Burban DJ, Haglund E, Capraro DT, Jennings PA: **Heterogeneous side chain conformation highlights a network of interactions implicated in hysteresis of the knotted protein, minimal tied trefoil**. *J. Phys.: Condens. Matter* 2015, **27**:354108.

40. Capraro DT, Jennings PA: **Untangling the Influence of a Protein Knot on Folding**. *Biophysical Journal* 2016, **110**:1044–1051.

41. Mallam AL, Onuoha SC, Grossmann JG, Jackson SE: **Knotted Fusion Proteins Reveal Unexpected Possibilities in Protein Folding**. *Molecular Cell* 2008, **30**:642–648.

42. Lim NC, Jackson SE: **Mechanistic insights into the folding of knotted proteins in vitro and in vivo**. *Journal of Molecular Biology* 2015, **427**:248–258.

43. Mallam A, Jackson SE: **Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins**. *Nat Chem Biol* 2012, **8**:147–153.

44. Sulkowska JI, Sulkowski P, Onuchic J: **Dodging the crisis of folding proteins with knots.** *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:3119–3124.

45. Wallin S, Zeldovich KB, Shakhnovich EI: **The Folding Mechanics of a Knotted Protein**. *Journal of Molecular Biology* 2007, **368**:884–893.

46. Prentiss MC, Wales DJ, Wolynes PG: **The energy landscape, folding pathways and the kinetics of a knotted protein**. *PLoS Comput Biol* 2010, **6**:e1000835.

47. Chwastyk M, Cieplak M: **Cotranslational folding of deeply knotted proteins**. *J. Phys.: Condens. Matter* 2015, **27**:354105.

48. Wang I, Chen S-Y, Hsu S-TD: **Unraveling the folding mechanism of the smallest knotted protein, MJ0366.** *J. Phys. Chem. B* 2015, **119**:4359–4370.

49. a Beccara S, Škrbić T, Covino R, Micheletti C, Faccioli P: **Folding pathways of a knotted protein with a realistic atomistic force field.** *PLoS Comput Biol* 2013, **9**:e1003002.

50. Noel JK, Sułkowska JI, Onuchic JN: **Slipknotting upon native-like loop formation in a trefoil knot protein**. *PNAS* 2010, **107**:15403–15408.

51. Noel JK, Onuchic JN, Sulkowska JI: **Knotting a Protein in Explicit Solvent**. *The Journal of Physical Chemistry Letters* 2013, **4**:3570–3573.

52. Najafi S, Potestio R: **Folding of small knotted proteins: Insights from a mean field coarse-grained model**. *J. Chem. Phys.* 2015, **143**:243121.

53. Chwastyk M, Cieplak M: **Multiple folding pathways of proteins with shallow knots and co-translational folding**. *J. Chem. Phys.* 2015, **143**:045101.

54. King NP, Jacobitz AW, Sawaya MR, Goldschmidt L, Yeates TO: **Structure and folding of a designed knotted protein.** *Proc Natl Acad Sci USA* 2010, **107**:20732–20737.

55. Wang L-W, Liu Y-N, Lyu P-C, Jackson SE, Hsu S-TD: **Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of Helicobacter pylori**. *J. Phys.: Condens. Matter* 2015, **27**:354106.

56. Sułkowska JI, Noel JK, Onuchic JN: **Energy landscape of knotted protein folding.** *Proc. Natl. Acad. Sci. U.S.A.* 2012, **109**:17783–17788.

57. Li W, Terakawa T, Wang W, Takada S: **Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot**. *PNAS* 2012, **109**:17789–17794.

58. Andersson FI, Pina DG, Mallam AL, Blaser G, Jackson SE: **Untangling the folding mechanism of the 5(2)-knotted protein UCH-L3.** *The FEBS Journal* 2009, **276**:2625–2635.

59. Ziegler F, Lim NCH, Mandal SS, Pelz B, Ng W-P, Schlierf M, Jackson SE, Rief M: **Knotting and unknotting of a protein in single molecule experiments.** *Proc. Natl. Acad. Sci. U.S.A.* 2016, **113**:7533–7538.

60. Wojciechowski M, Gómez-Sicilia À, Carrion-Vazquez M, Cieplak M: **Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins.** *Mol Biosyst* 2016, doi:10.1039/c6mb00214e.

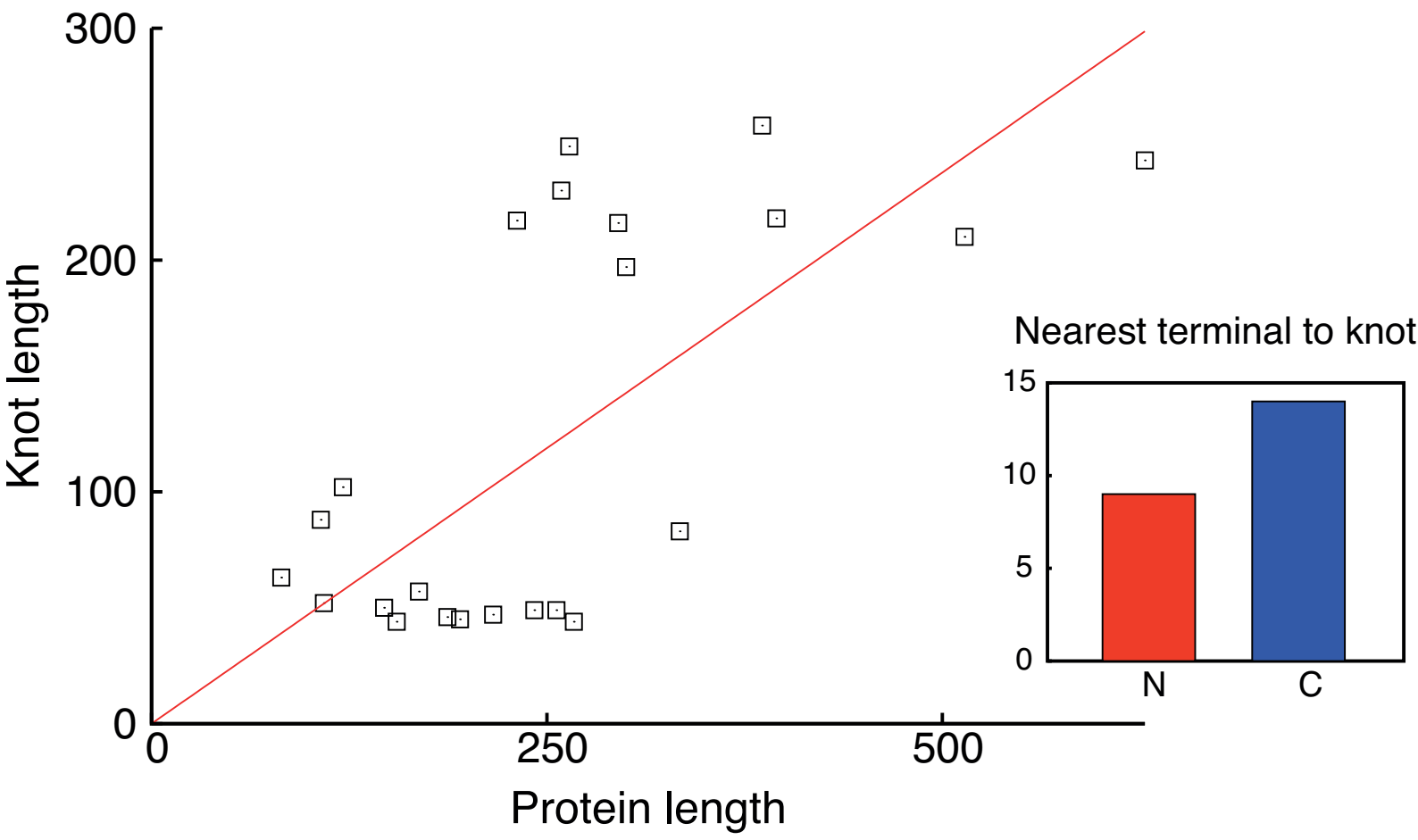
61. Suma A, Rosa A, Micheletti C: **Pore Translocation of Knotted Polymer Chains: How Friction Depends on Knot Complexity**. *ACS Macro Lett.* 2015, **4**:1420–1424.

62. Soler MA, Nunes A, Faísca PFN: **Effects of knot type in the folding of topologically complex lattice proteins**. *J. Chem. Phys.* 2014, **141**:025101.

63. Faísca PFN, Travasso RDM, Charters T, Nunes A, Cieplak M: **The folding of knotted proteins: insights from lattice simulations.** *Phys. Biol.* 2010, **7**:16009.

64. Tubiana L, Rosa A, Fragiacomo F, Micheletti C: **Spontaneous Knotting and Unknotting of Flexible Linear Polymers: Equilibrium and Kinetic Aspects**. *Macromolecules* 2013, **46**:3669–3678.

65. Micheletti C, Orlandini E: **Knotting and Unknotting Dynamics of DNA Strands in Nanochannels**. *ACS Macro Lett.* 2014, **3**:876–880.


300
200
100
0
Knot length
0
250
500
Protein length
Nearest terminal to knot
15
10
5
0
N
C

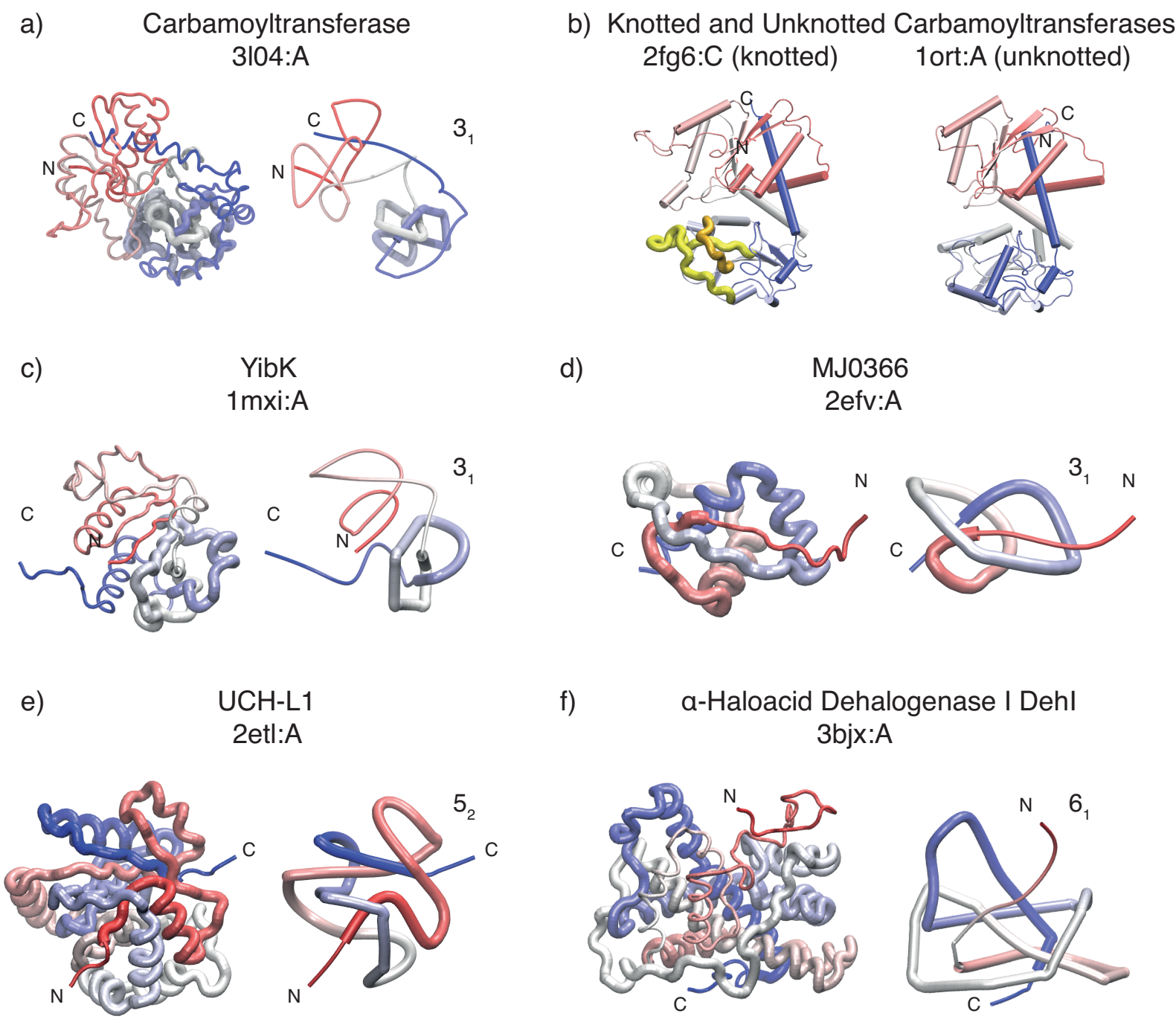

a)
Carbamoyltransferase
3l04:A
C
N
C
N
3 1
b)
Knotted and Unknotted Carbamoyltransferases
2fg6:C (knotted)
1ort:A (unknotted)
C
N
C
N
c)
YibK
1mxi:A
C
N
C
N
3 1
d)
MJ0366
2efv:A
N
C
3 1
N
C
e)
UCH-L1
2etl:A
C
N
5 2
C
N
f)
α-Haloacid Dehalogenase I DehI
3bjx:A
N
C
N
6 1
C

Table 1

Protein or protein function	PDB code	Knot type	PDB length	Knot length	N-terminal depth	C-terminal depth
Carbonic Anhydrase	3MDZ:A	3_1	259	230	26	3
RNA Methyltransferase	1X7O:A	3_1	267	44	190	33
	4H3Z:B	3_1	256	49	89	118
	4CND:A	3_1	169	57	79	33
	4E8B:A	3_1	242	49	164	29
	3O7B:A	3_1	216	47	143	26
	4JAK:A	3_1	155	44	77	34
	4JWF:A	3_1	187	46	98	43
	2QMM:A	3_1	195	45	124	26
S-Adenosylmethionine Synthetase	4ODJ:A	3_1	386	258	16	112
Carbamoyltransferase	3KZK:A	3_1	334	83	169	82
Hypothetical RNA Methyltransferase	1O6D:A	3_1	147	50	67	30
Hypothetical Protein MJ0366	2EFV:A	3_1	82	63	10	9
H+/Ca2+ Exchanger	4KPP:A	3_1	395	218	72	105
Na+/Ca2+ Exchanger	5HWY:A	3_1	300	197	33	70
N-Acetylglucosamine Deacetylase	5BU6:A	3_1	264	249	10	5
DNA Binding Protein	2RH3:A	3_1	121	102	7	12
	4LRV:A	3_1	107	88	8	11
Metal Binding (Zinc-Finger)	2K0A:A	3_1	109	52	21	36
Bacteriophytochrome	4GW9:A	4_1	628	243	20	365
Ketol-Acid Reductoisomerase	1QMG:A	4_1	514	210	236	68
Ubiquitin Carboxy-Terminal Hydrolase	2LEN:A	5_2	231	217	2	12
α-Haloacid Dehalogenase I DehI	3BJX:B	6_1	295	216	59	20

e-component- Supp. Info. Details for knotted protein survey

Click here to download e-component: Supporting_information_Knotted_Proteins.pdf

e-component- Supp. Info. complete list of knotted PDB entries

Click here to download e-component: Knotted_protein_survey_complete_list.dat