Topological descriptions of protein folding

Erica Flapan[a], Adam He[b], and Helen Wong[c,1]

[a]Department of Mathematics, Pomona College, Claremont, CA 91711; [b]Computational Biology Program, Cornell University, Ithaca, NY 14853; and [c]Department of Mathematical Sciences, Claremont McKenna College, Claremont, CA 91711

Edited by José N. Onuchic, Rice University, Houston, TX, and approved March 18, 2019 (received for review May 14, 2018)

How knotted proteins fold has remained controversial since the identification of deeply knotted proteins nearly two decades ago. Both computational and experimental approaches have been used to investigate protein knot formation. Motivated by the computer simulations of Bölinger et al. [Bölinger D, et al. (2010) *PLoS Comput Biol* 6:e1000731] for the folding of the 6_1-knotted α-haloacid dehalogenase (DehI) protein, we introduce a topological description of knot folding that could describe pathways for the formation of all currently known protein knot types and predicts knot types that might be identified in the future. We analyze fingerprint data from crystal structures of protein knots as evidence that particular protein knots may fold according to specific pathways from our theory. Our results confirm Taylor's twisted hairpin theory of knot folding for the 3_1-knotted proteins and the 4_1-knotted ketol-acid reductoisomerases and present alternative folding mechanisms for the 4_1-knotted phytochromes and the 5_2- and 6_1-knotted proteins.

protein topology | knot folding | protein knots

When protein knots were first identified, they were believed to be the result of randomly occurring misfolded conformations. However, the discovery of proteins containing deeply embedded knots forced a reevaluation of this belief (1–3). Systematic reviews of the ever-growing Protein Data Bank (PDB) and the development of specialized servers for detecting protein knots have led to the identification of hundreds of knotted proteins, and it is now generally accepted that a small but significant fraction of proteins contains knots (4–7). However, exactly how and why such knots form are still unknown.

As of now, only five distinct knot types have been found in proteins in the PDB. We illustrate these knots as closed curves in Fig. 1, although the proteins containing them are actually open chains. From a mathematical perspective, a knotted open chain is equivalent to an unknotted open chain, since an open chain can unknot via a continuous deformation. However, from a biophysical perspective, the energy necessary to undo a deeply embedded protein knot is prohibitively large, effectively trapping the knot in the open chain. Thus, it is common practice to close knotted proteins by bringing the ends of the knotted chain together to obtain a loop, to which a knot type can be assigned. Different methods of closing the chain can result in different knot types, and various approaches have been used to resolve this problem (1, 8–17).

The standard notation to represent knots (18) uses a large numeral to denote the minimum number of crossings among all projections of the knot and a subscript to identify the particular knot with that number of crossings as in Fig. 1. To distinguish the two enantiomorphs of a chiral knot, we use the algorithm described by Mislow and coworkers (19, 20) to assign $+$ to one form and $-$ to the other.

According to a recent survey, 23 families of knotted proteins have been identified (21). Of these families, 19 contain $\pm 3_1$ knots, only the ketol-acid reductoisomerases (KARIs) and the phytochromes contain 4_1 knots, only the ubiquitin C-terminal hydrolases (UCHs) contain -5_2 knots, and only the α-haloacid dehalogenase (DehI) contains $+6_1$ knots. It is unclear why the -5_2 and $+6_1$ knots have been found, but their mirror forms have not. It could be a matter of time before a protein is found to contain $+5_2$, -6_1, or any other knot (22).

The study of protein knotting has been approached using experimentation, simulation, and theoretical descriptions (refs. 21 and 22 have recent reviews). Inspired by the theory of knot folding put forth by Taylor (23) (*1. Taylor's Twisted Hairpin Theory*) together with the simulations of Bölinger et al. (24) for the folding of DehI (*2. Knot Folding via Loop Flipping*), this paper presents a theoretical description of protein knot folding, which could be applicable to any knotted protein. Because our theory builds on recent experimental and computational results about knot folding, it provides a step forward in current thinking about knot folding.

1. Taylor's Twisted Hairpin Theory

Taylor (23) introduced a theory of protein knot folding where the protein assumes the form of a "twisted hairpin." Then, one terminus threads through the "eye" of the hairpin to create a knot. We illustrate this in Fig. 2, with dotted segments added to make the chains into closed loops. Note that knotted protein conformations are more complex and contain more crossings than the projections drawn in Fig. 2. We use these simplified drawings to focus on the knotting mechanism.

We will refer to this mechanism of knot folding as a twisted hairpin pathway. Knots obtained in this way are in the family of twist knots. According to Taylor's theory, all protein knots identified in the future must also be members of this family (23). For example, the 5_1, 6_2, and 6_3 knots (Fig. 3) have not been found in any solved protein structures, although they have a projection with the same number of crossings as the 5_2 and 6_1 knots, which

Significance

Knotting in proteins was once considered exceedingly rare. However, systematic analyses of solved protein structures over the last two decades have demonstrated the existence of many deeply knotted proteins. Conservation of knotting across some protein families strongly suggests that knotting can be important for protein structure and function, and hence, significant interest has arisen around how protein knots form. We build on results of previous computer simulations and prior theories of protein knot formation to obtain theoretical pathways for protein knotting that could apply to any knotted protein. By comparing our theoretical pathways with structural data on solved proteins, we determine which of our pathways may be feasible for each of the known protein knot types.

Author contributions: E.F. and H.W. designed research; E.F., A.H., and H.W. performed research; E.F. and H.W. analyzed data; and E.F., A.H., and H.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: hwong@cmc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808312116/-/DCSupplemental.

Published online April 18, 2019.

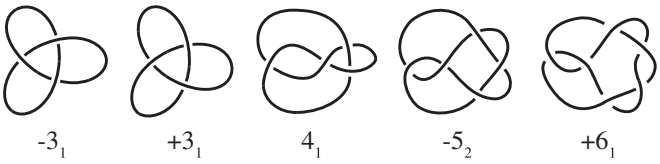

Fig. 1. As of now, these are the only knots that have been identified in proteins.

have been found. Since these three knots are not twist knots, Taylor's theory of knot folding would explain their absence among solved protein structures.

Taylor (23) argues that, in knot folding, loop penetration is the rate-limiting event and that knot formation depends primarily on the number of times that loop penetration occurs, then on the number of residues that must be threaded through the loop, and lastly, on the number of crossings in the resulting knot (23). This hierarchy suggests that proteins with deep knots should be less prevalent than those with shallow knots independent of the number of crossings. However, as can be seen on the database KnotProt (4, 5), proteins containing a deep 3_1 knot are vastly more common than those containing a shallow 5_2 knot. Thus, in contrast with Taylor's hierarchy, we would expect the number of crossings in a given knot to be higher in the hierarchy than the number of residues that must be threaded through the loop.

However, knot-folding rates are not only a function of the number of crossings and the depth of a knot. In particular, chaperones can speed up the kinetics of knot folding as has been observed for the 3_1-knotted proteins YibK, YbeA (25–27), VirC2, and DndE (28) and for the 5_2-knotted UCHs (29, 30). Furthermore, the work of Wallin et al. (31) shows that nonnative interactions increase the probability of correct knot folding for the deeply 3_1-knotted YibK protein, and simulations of Covino et al. (32) confirm this for the shallow trefoil knot in the MJ0366 protein.

In addition, the work of Chwastyk and Cieplak (33) shows that ribosomes may play a significant role in knot folding. In particular, they argue that the deep knot in the YibK protein is a result of on-ribosome folding. Recent computer simulations by Dabrowski-Tumanski et al. (34) also indicate that ribosomes play an active role in the folding of the protein with PDB ID code 5JIR, which contains the deepest 3_1 knot that has been identified in a protein. In particular, according to their simulations, one end of the nascent chain comes out of the ribosome and forms a twisted loop, which attaches to the ribosome around the exit tunnel. While this loop is held in place, the ribosome pushes a piece of the protein through the exit tunnel so that it is surrounded by the first loop, creating a slipknot. Finally, the rest of the chain is threaded through the exit tunnel to form a 3_1 knot.

While Taylor's twisted hairpin theory is useful to describe knot folding independent of any particular protein, there is computational and experimental evidence that this may not be the only pathway to protein knotting. In particular, it has been shown that encapsulation in a chaperonin can facilitate multiple folding pathways (28, 29). Even without chaperones (25, 35–39), knotted proteins can have complex energy landscapes that include knotted intermediates and parallel pathways. This is supported by simulations indicating that some trefoil-knotted proteins fold via multiple pathways (31, 40–42), including a newly described pathway where each terminus threads through a separate loop (42). Also, computational studies of the folding of the 5_2 knot in UCHs (29) and the folding of the 6_1 knot in DehI (24, 43) produced pathways that involved knotted intermediates. Such intermediates would not occur if the knots folded via a twisted hairpin pathway, because the chain remains unknotted until threading occurs at the last step.

For all of the above reasons, even if a twisted hairpin pathway is the primary folding mechanism for knotted proteins, it is worth considering alternative pathways that permit partially folded knotted intermediates. In the next section, we describe loop flipping as a knot-folding mechanism. Then, in *3. Our Proposed Theory of Knot Folding*, we introduce our theory of knot folding.

2. Knot Folding via Loop Flipping

While Taylor's theory of knot folding assumes that a terminus threads through the loop of a twisted hairpin, the same conformation would be produced if the loop of the hairpin was to flip over the terminus. In fact, the mobility of the loop may confer thermodynamical advantages, making it easier for knotting to occur by a loop-flipping motion rather than by threading. Furthermore, experimental results on the thermodynamic and kinetic properties of a -3_1-knotted protein similar to the MTase protein (44), a -5_2-knotted UCH protein (37), and the $+6_1$-knotted DehI protein (43) have all been consistent with loop flipping as a knotting mechanism. Loop flipping (also known as a "mousetrap-like" or "jump-rope-like" motion) is increasingly being observed in structure-based simulations of knot folding. For example, simulations show that some 3_1-knotted proteins (41, 42, 45) as well as the 5_2-knotted UCH proteins (29, 46) have at least one folding pathway involving a large loop flipping over a terminus.

Furthermore, using molecular dynamics simulations with a coarse-grained Gō model of the folding of DehI, Bölinger et al. (24) found two pathways to the $+6_1$ knot, which each involved a large loop flipping over a mostly folded smaller loop. They then used crystallographic B-factor data from the DehI protein to verify that the relevant pieces of the protein are flexible enough to permit the loop flipping required by this pathway. While their simulations did not take into account nonnative interactions, they assert based on the work of Wallin et al. (31) that nonnative interactions should, in fact, increase the rate of knot folding via their pathways.

In Fig. 4, we illustrate the steps of the simulation of Bölinger et al. (24). In Steps 1 and 2, the polypeptide forms a large green loop and a smaller red loop, which are aligned. At this point, the folding mechanism splits into two parallel pathways. In Step 3a, the red loop twists one more time, and the green loop flips over both the red loop and the blue end, causing the blue end to thread through the green loop. Then, the blue end threads through the red loop to obtain Step 4. In Step 3b, the red loop twists one more time, and the blue end threads through it. From here, the green loop flips over both the red loop and the blue end, causing the blue end to thread through the green loop to again obtain Step 4. In both pathways, loop flipping enables the efficient threading of the terminus through the two loops. Note that these pathways are distinct, because the intermediate in Step 3a is a 4_1 knot, while that of Step 3b is the unknot (*SI Appendix*, Fig. S40).

According to Bölinger et al. (24), the loop-flipping motion is facilitated by the existence of glycine and proline in the flexible regions of the protein. However, loop flipping is the

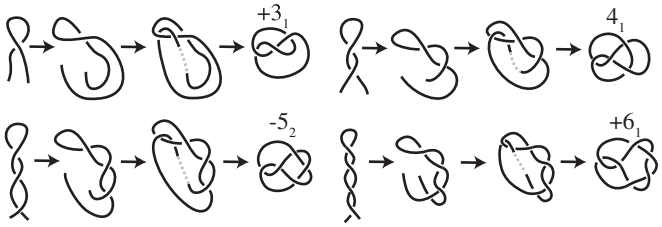

Fig. 2. The twisted hairpin folding mechanism proposed by Taylor (23).

Fig. 3. Neither these knots nor their enantiomorphs have been found in any protein.

rate-limiting step independent of how far the C terminus has threaded through the smaller loop. Thus, the depth of the knot does not slow the process down. This is in contrast to Taylor's twisted hairpin theory, which assumes that a deep knot will fold less efficiently than a shallow one (23).

While the simulation of Bölinger et al. (24) shows that loop flipping is implicated in the folding of a deep 6_1 knot in DehI, loop flipping has also been identified as a folding pathway for a shallow 3_1 knot in the protein MJ0366 (41). More recently, Chwastyk and Cieplak (42) have shown that MJ0366 has multiple folding pathways, including newly described two-loop mechanisms, and some of the pathways in the one-loop and two-loop mechanisms involve loop flipping. Given these examples of loop flipping as a folding mechanism for both deep and shallow knots, we assert that loop flipping should be considered as a possible folding mechanism for any protein knot whether deep or shallow.

3. Our Proposed Theory of Knot Folding

Motivated by the steps described in the simulation of Bölinger et al. (24) for the folding of DehI, we developed the following general theory of knot folding, which includes the pathways described by Bölinger et al. (24) and Taylor's twisted hairpin theory as special cases. Like the twisted hairpin theory, our theory is not obtained via a computer simulation and is not focused on any particular protein or family of proteins. In fact, we will show in *4. Knots That Can Be Obtained with Our Theory* that all known protein knots can be obtained by applying our steps. Thus, while we do not claim, as Taylor (23) did, that our theory is the only knot-folding mechanism, we believe that our theory is a possibility that should be considered for any knotted protein.

The Steps of Our Theory. An unknotted open chain is colored as in Fig. 4.

1) A small red loop and a large green loop each containing zero, one, or two twists form and come close together.
2) The blue end approaches the two loops, causing the black arc to pass either behind or in front of the red arc.
3) One of the following occurs.
   a) The green loop flips over the red loop and threads the blue end. Then, the loops align, and the blue end threads through the red loop.
   b) The blue end threads through the red loop. Then, the green loop flips over both the blue end and the red loop so that the loops are aligned and the green loop is threaded.

Fig. 5 illustrates how the -5_2 knot could be folded using these steps. In Step 1, red and green loops are formed and brought close together. In Step 2, the blue end approaches the two loops, causing the black arc to pass in front of the red arc. In Step 3a, the green loop flips over the red loop and threads the blue end, after which the red loop aligns with the green loop and the blue end threads through the red loop. Alternatively, in Step 3b, the blue end threads through the red loop, after which the green loop flips over both the blue end and the red loop so that the two loops are aligned and the green loop is now threaded.

The steps of our theory are closely related to the steps in the simulation of Bölinger et al. (24) for the knotting of DehI. The primary difference between our theory and the simulation of Bölinger et al. (24) is that we allow zero, one, or two twists in each of the loops, while they mandate two twists in the red loop and one twist in the green loop. The other differences are that we do not specify which direction the loops should twist in or whether the black arc should go behind or in front of the red arc.

Because of the parallels between our theory and the simulation of Bölinger et al. (24), we adopt the same assumptions. In particular, following Bölinger et al. (24), we assume that nonnative interactions and chaperones are not required for our steps to occur, although such interactions are likely to speed up the folding rate. Also, for our steps and those of Bölinger et al. (24), the blue terminus is the only one that is required to move during the knot folding. Thus, in a cotranslational model, where the red terminus is attached to the ribosome and the blue terminus remains free, knot folding could still occur with our steps.

In fact, Sorokina and Mushegian (47, 48) have argued that, for many proteins, knot formation is significantly facilitated when the protein is formed on the ribosome. This is consistent with the simulation of Chwastyk and Cieplak (33), which shows that the probability of knot formation for the protein YibK is increased substantially when the protein is formed on the ribosome. More recently, Dabrowski-Tumanski et al. (34) obtained similar results for the deep 3_1-knotted protein with PDB ID code 5JIR. Because of the role of the ribosome in promoting knotting in all of these studies, we expect a cotranslational model to promote knotting for our theory as well. In particular, for knots that are deeply embedded on the blue end or on both ends, the loop-flipping mechanism described by our steps might be difficult to achieve. In this case, the ribosome could facilitate the mechanism by acting as a scaffolding during the steps. For example, in our Step 1, the red loop and the green loop could exit from the ribosome and then be held in place while the blue end exits the ribosome in Step 2 and threads through the red loop in Step 3b. Afterward, the ribosome would release the green loop and hold the red loop and the blue end close together while the green loop flips over both to obtain the conformation in Step 4.

Our requirement in Step 1 that there are no more than two twists in each loop is related to an observation of Taylor (23) that the more twists in a hairpin, the farther the termini may be from the loop, making threading less likely. By an analogous argument, the more twists there are in either or both of the loops in our theory, the farther the blue terminus may be from the loops, making loop flipping over the terminus less likely (although this

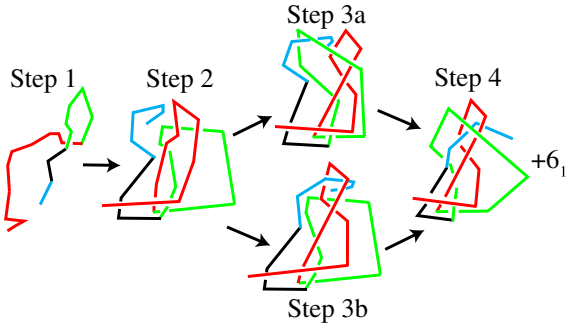

Fig. 4. The loop-flipping mechanism identified in structure-based simulations by Bölinger et al. (24). In Steps 1 and 2, green and red loops are formed. In Steps 3 and 4, the red loop adds a second twist, after which the larger green loop flips over the red loop and the blue end threads through the red and green loops in either of the orders illustrated in Steps 3a and 3b.

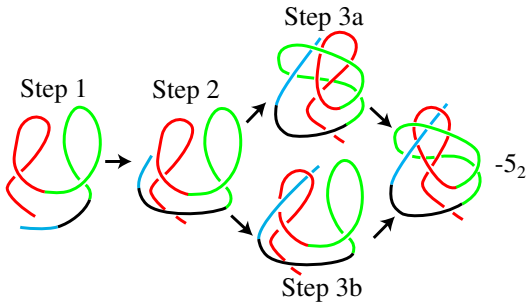

Fig. 5. An example of how a -5_2 knot could be folded with the above steps.

distance might be diminished if the protein is encapsulated in a chaperonin). In addition, according to Banavar and Maritan (49), proteins should be considered as tubes of nonzero thickness, and according to Taylor (23), this means that knots with more twisting require longer chains. Since a given protein has a fixed length, its thickness will favor a knotting mechanism requiring as few twists as possible in each loop. For all of these reasons, whenever multiple pathways produce the same knot, we assume that those with fewer twists in each loop will be more likely to describe successful knotting. We will refer to this principle henceforth as the Minimal Twisting Principle.

With this principle in mind, we cannot allow arbitrarily many twists in the loops described by our theory. Since the red loop in Fig. 4 has two twists and we want our theory to encompass the results of the simulation of Bölinger et al. (24), our upper bound must be at least two. However, as we will see in *4. Knots That Can Be Obtained with Our Theory*, all known protein knots can be obtained by applying our steps with at most two twists in each loop. Hence, we use two twists as an upper bound. If new protein knots are identified that require more twists in one or both of the loops, this upper bound can be increased accordingly.

While our steps describe a general knotting mechanism, the particular parameters involved can vary as follows.

The Parameters of Our Theory.

- The number of twists in the green and red loops and the direction in which they twist
- Whether the black arc crosses over or under the red arc
- Whether the left or right side of each loop goes in front or behind the blue arc after it threads

For example, in the final conformation of Fig. 5, the red loop and the green loop each have one twist but in opposite directions, the black arc crosses over the red arc, and the right sides of both loops are in front of the blue arc.

To symbolically represent different types of crossings, we introduce the following sign convention. If the slope of an overcrossing is positive, we designate the crossing by a $+$ sign, and if the slope of an overcrossing is negative, we designate the crossing by a $-$ sign. For example, in the final conformation of Fig. 5, the crossing of the red loop is negative, the crossing of the green loop is positive, and the red–black crossing is positive. In some illustrations, it is hard to tell the sign of the red–black crossing. Thus, we remark that the red–black crossing is always positive if the black arc goes over the red arc and always negative if the red arc goes over the black arc. Note that our sign convention does not agree with standard practice in knot theory, which requires a uniform orientation on an entire knot before determining the sign of any crossing.

Fig. 6 shows projections of all of the knots that can be obtained with our steps together with the notation that we will use to represent them. We refer to these projections as the configurations of our theory. Observe that configurations encode the steps of the pathways used to obtain them and are useful for determining the knot types resulting from these pathways. However, these are simplified drawings of the final conformations obtained with our knotting mechanism. The actual conformations are much more complicated.

We use the following notation for configurations. The letters L and R indicate whether the left or right side, respectively, of a loop goes in front of the blue arc. We always list an L or R for the red loop before we list it for the green loop. The first parameter inside of the parentheses indicates whether the red–black crossing at the bottom of the projection is positive or negative. The parameters a and b, which can be 0, ±1, or ±2, describe the number and slope of the vertical twists inside the boxes. As with L and R, we list the crossings of the red loop before we list the crossings of the green loop. For example, the -5_2 knot in Fig. 5 has configuration $RR(+,-1,1)$, and the projection of the knot resulting from the simulation of Bölinger et al. (24) has configuration $RR(-,2,-1)$ (see Fig. 12).

Although the blue and red ends illustrated in Fig. 6 are very short, either or both ends could be much longer, yielding a deeper knot. A very deep knot, like the 3_1 found in the protein with PDB ID code 5JIR, would correspond to a configuration where both ends are significantly longer. In this case, the protein would have three separate domains, with only the middle one knotted, and the external domains would remain unfolded until after the knotting mechanism begins.

4. Knots That Can Be Obtained with Our Theory

Table 1 lists the positive forms of all nontrivial knots that can be obtained with our theory together with the parameters of the configurations that are used to obtain them. This includes the positive forms of the knots $+3_1$, 4_1, $+5_2$, and $+6_1$, although 5_2 has only been found in its negative form in a protein. *SI Appendix*, Table S2 lists the configurations for the unknot and all the nontrivial knots that can be obtained with our steps. *SI Appendix*, Table S1 displays the same information, but organized according to parameters rather than according to knot type. In particular, this includes the negative forms -3_1 and -5_2, which have been found in proteins. Table 1 lists the positive forms of 10 additional knot types 5_1, 6_2, 6_3, 7_2, 7_5, 7_6, 7_7, 8_8, 8_{14}, and 9_{23}, which have not yet been identified in proteins. An explanation of how the tables were produced is given in *8. Materials and Methods*, and detailed computations are provided in *SI Appendix*.

Every knot in Table 1, except for 9_{23}, occurs with multiple configurations. Since each configuration represents a pair of pathways to a knot's formation, this means that our steps produce many pathways to fold most of the knots. This makes biological sense, since different families of proteins with the

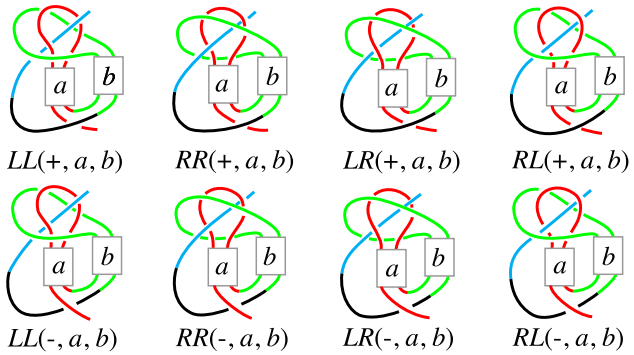

Fig. 6. We illustrate all configurations, where *a* and *b* are 0, ±1, or ±2 and a $+$ or $-$ sign indicates whether the slope of an overcrossing is positive or negative.

Table 1. Right-handed (+) and achiral knots produced by our model and the configurations used to obtain them

Knot	$RR(\pm, a, b)$	$RL(\pm, a, b)$	$LR(\pm, a, b)$	$LL(\pm, a, b)$	No. of configurations
$+3_1$	(+, 0, 0), (−, 1, −2), (−, −2, 1), (−, 2, 2)	(+, 0, −1), (−, −2, 0), (−, 2, 1)	(+, −1, 0), (−, 0, −2), (−, 1, 2)	(+, −1, −1), (−, 1, 1)	12
4_1	(+, 0, −1), (+, −1, 0), (−, 1, 2), (−, 2, 1)	(+, −1, −1), (−, 1, 1), (−, 2, 0), (+, 0, −2)	(+, −1, −1), (−, 1, 1), (+, −2, 0), (−, 0, 2)	(−, 0, 1), (−, 1, 0), (+, −1, −2), (+, −2, −1)	16
$+5_1$	(+, −2, −2), (−, 0, 0)	(−, 0, −1)	(−, −1, 0)	(−, −1, −1)	5
$+5_2$	(+, 0, −2), (+, −2, 0), (−, 0, 2), (−, 2, 0)	(−, 0, 1), (+, −2, −1), (−, 2, −1)	(−, 1, 0), (+, −1, −2), (−, −1, 2)	(−, −1, 1), (−, 1, −1)	12
$+6_1$	(−, −1, 2), (−, 2, −1)	(−, −1, 1), (−, 2, −2)	(−, 1, −1), (−, −2, 2)	(−, 1, −2), (−, −2, 1)	8
$+6_2$	(−, 0, −1), (−, −1, 0)	(−, −1, −1), (−, 0, −2)	(−, −1, −1), (−, −2, 0)	(−, −1, −2), (−, −2, −1)	8
6_3	(+, 1, −2), (+, −2, 1)	(+, −2, 0), (−, 0, 2)	(+, 0, −2), (−, 2, 0)	(−, −1, 2), (−, 2, −1)	8
$+7_2$	(−, 2, −2), (−, −2, 2)	(−, −2, 1)	(−, 1, −2)		4
$+7_5$	(−, 0, −2), (−, −2, 0)	(−, −2, −1)	(−, −1, −2)		4
$+7_6$	(+, 2, −2), (+, −2, 2)	(+, −2, 1)	(+, 1, −2)		4
$+7_7$	(+, 2, 2)	(+, 2, 1)	(+, 1, 2)	(+, 1, 1)	4
$+8_8$		(−, −2, 2)	(−, 2, −2)		2
$+8_{14}$	(−, −1, −2), (−, −2, −1)	(−, −2, −2)	(−, −2, −2)		4
$+9_{23}$	(−, −2, −2)				1

For a table of all knots produced by our model and their configurations, see *SI Appendix*, Tables S1 and S2.

same knot would not necessarily fold in the same way, and even one particular protein may have multiple knotting pathways.

To understand the relationship between a configuration and its mirror image, observe that all of the overcrossings and undercrossings are interchanged when a configuration is reflected in the plane of the paper. As a result, the mirror image of a configuration will interchange R and L and change the sign of each of the other parameters of the configuration. We summarize this in the following lemma.

Lemma 1. *Let a and b be integers. Then, the following relationships hold between configurations and their mirror forms (denoted by a minus sign in front of the configuration):*

$$RR(+, a, b) = -LL(-, -a, -b)$$
$$RR(-, a, b) = -LL(+, -a, -b)$$
$$RL(+, a, b) = -LR(-, -a, -b)$$
$$RL(-, a, b) = -LR(+, -a, -b).$$

For example, we see from Table 1 that the $+5_2$ knot is produced by 12 configurations:

$$RR(+, 0, -2),\ RR(+, -2, 0),\ RR(-, 0, 2),$$
$$RR(-, 2, 0), RL(-, 0, 1),\ RL(+, -2, -1),$$
$$RL(-, 2, -1),\ LR(-, 1, 0),\ LR(+, -1, -2),$$
$$LR(-, -1, 2),\ LL(-, -1, 1),\ LL(-, 1, -1).$$

It now follows from Lemma 1 that the -5_2 knot is produced by 12 configurations:

$$LL(-, 0, 2),\ LL(-, 2, 0),\ LL(+, 0, -2),$$
$$LL(+, -2, 0), LR(+, 0, -1),\ LR(-, 2, 1),$$
$$LR(+, -2, 1),\ RL(+, -1, 0), RL(-, 1, 2),$$
$$RL(+, 1, -2),\ RR(+, 1, -1),\ RR(+, -1, 1).$$

This means that, in total, there are 24 configurations for the $\pm 5_2$ knot. By contrast, because the 4_1 knot is achiral, the 16 configurations listed in Table 1 for the 4_1 knot are the only ones that can produce it.

One of the key tools that we used to construct the tables is the following result, the proof of which is given in *SI Appendix*.

Theorem 1. *Let a and b be integers, and let ε denote $+$ or $-$. If one of the following configurations has knot type K, then all of these configurations have knot type K:*

$$RR(\varepsilon, a, b),\ RR(\varepsilon, b, a),$$
$$RL(\varepsilon, a, b-1),\ LR(\varepsilon, b-1, a),$$
$$RL(\varepsilon, b, a-1),\ LR(\varepsilon, a-1, b),$$
$$LL(\varepsilon, a-1, b-1),\ LL(\varepsilon, b-1, a-1).$$

Furthermore, all of the following configurations have knot type $-K$ (the mirror image of K):

$$LL(-\varepsilon, -a, -b),\ LL(-\varepsilon, -b, -a),$$
$$LR(-\varepsilon, -a, -b+1),\ RL(-\varepsilon, -b+1, -a),$$
$$LR(-\varepsilon, -b, -a+1),\ RL(-\varepsilon, -a+1, -b),$$
$$RR(-\varepsilon, -b+1, -a+1),\ RR(-\varepsilon, -a+1, -b+1).$$

Thus, for an achiral knot K, if any configuration listed above has knot type K, then all 16 configurations have knot type K.

The following theorem, proved in *SI Appendix*, gives us information about the types of knots that can be produced by our theory (without restrictions on a and b).

Theorem 2. *All knots obtained by our steps can be deformed to a conformation with projection that has only two local maxima.*

This theorem does not imply that, if a protein becomes knotted via our steps, its final conformation has only two local maxima. In fact, due to physical and chemical properties, such as hydrophobic collapse, the final conformation of a knotted protein is quite complicated, containing many crossings and many local maxima. Saying that a knot can be deformed to have only two local maxima is saying something about its knot type rather than about its particular conformation.

To rephrase Theorem 2 in more mathematical language, all knots that can be produced with our steps are in the family of two-bridge knots. Such knots are a proper subset of the prime knots (i.e., those that cannot be split into two knotted arcs) (50). However, of the 84 prime knots with nine or fewer crossings, just 50 are two bridge (18), and only 14 of these can be obtained with our steps, where a and b are restricted to 0, ± 1, and ± 2.

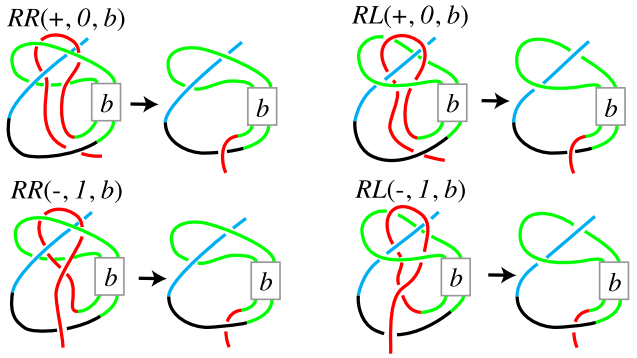

Fig. 7. The red loop does not play an essential role in these pathways, since it can be eliminated by pulling down on the red end.

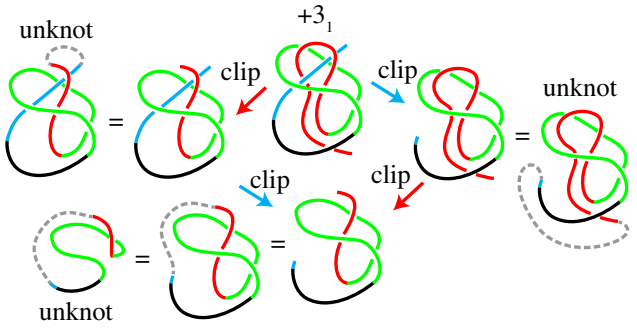

Fig. 8. The fingerprint of a $+3_1$ knot with configuration $RL(+, 0, -1)$ agrees with the fingerprints on KnotProt containing only one $+3_1$.

5. Configurations That Are Consistent with Twisted Hairpin Pathways

While our theory of knot folding was motivated by the simulation of Bölinger et al. (24), it is consistent with Taylor's twisted hairpin theory in the special case where the red loop does not play an essential role in the folding mechanism. In particular, we see in Fig. 7 that, if we start with the configuration $RR(+, 0, b)$, $RL(+, 0, b)$, $RR(-, 1, b)$, or $RL(-, 1, b)$ and tighten the knot by pulling downward on the red end while fixing the blue end, then the red arc will slide down along the blue and black arcs so that the red loop disappears, leaving a single red–black crossing at the bottom of the picture. This means that, for these configurations, the knotting was entirely due to the threading of the green loop. In this sense, the configurations $RR(+, 0, b)$, $RL(+, 0, b)$, $RR(-, 1, b)$, and $RL(-, 1, b)$ represent knotting mechanisms that are similar to a twisted hairpin pathway.

More generally, we say that a configuration is consistent with a twisted hairpin pathway if pulling down on the red end while fixing the blue end causes the red loop and all of the red twisting to disappear, leaving only a single red–black crossing at the bottom. Theorem 3 (proven in *SI Appendix*) says that the only configurations with this property are those illustrated in Fig. 7 together with their mirror images.

Theorem 3. *The only configurations that are consistent with a twisted hairpin pathway are* $RR(+, 0, b)$, $RR(-, 1, b)$, $RL(+, 0, b)$, $RL(-, 1, b)$, $LR(-, 0, b)$, $LR(+, -1, b)$, $LL(-, 0, b)$, *and* $LL(+, -1, b)$.

6. Knot Fingerprints of Configurations

King et al. (51) and Taylor (52) defined the fingerprint of a knotted protein to be the knot types of the protein and all of the partial structures obtained by clipping residues from each of the termini. Knot fingerprinting is useful, because it distinguishes different conformations of the same knot. For example, the 4_1 knot has been identified in KARIs and phytochromes (4). However, according to KnotProt, if both termini of the KARIs are clipped, we obtain a $+3_1$ knot, while no matter how much one or both termini of the phytochromes are clipped, we will not obtain a $\pm 3_1$ knot.

In this section, we compare the knot fingerprints of configurations from our theory with those of proteins on KnotProt to see if the pathways described by these configurations could correspond to folding pathways for the proteins. In particular, for each knot, we use Theorem 3 to determine which configurations are consistent with a twisted hairpin pathway for that knot. Then, we compare the knotted subchains of these configurations with those on KnotProt. For any protein where these do not agree, we propose an alternative configuration.

We use the following rules for creating fingerprints. At each step, we clip an end roughly at the first place where the knot type of the subchain is distinct from what it was before the cut, indicating which end has been cut with a red or blue arrow. Since the green and red loops are closely aligned, if we clip the blue end so that it goes through the red loop but not the green loop, then an extension of the blue end is likely to again pass through both loops. Thus, our first cut of the blue end will always remove enough of the blue arc so that it no longer passes through either loop, and hence, it will occur near where the blue and black arcs meet.

Each time that we clip one or both ends, we join the ends together with a dotted arc to show the most likely knot in a subchain. These dotted arcs are not part of the structure, and therefore, we remove them before we do any additional clipping.

As explained in *8. Materials and Methods*, wiggles can be added to any configuration to obtain repeated occurrences of a given knot. Thus, here, we focus only on comparing distinct knot types of subchains in configurations with those of proteins on KnotProt. More information on how fingerprints are determined by KnotProt and in the figures below is in *8. Materials and Methods* and *SI Appendix*, section 6.

Fingerprints of 3_1-Knotted Proteins. None of the $\pm 3_1$-knotted proteins on KnotProt have subchains containing any other knots. Thus, any configuration for $\pm 3_1$ with no other nontrivial knots in its fingerprint will agree with these data.

By the Minimal Twisting Principle, the configuration $RR(+, 0, 0)$ would describe the most likely pathways for folding the $+3_1$ knot, because it requires the least twisting of any configuration for the $+3_1$ knot (Table 1). We show in *SI Appendix*, Fig. S28 that the fingerprint of $RR(+, 0, 0)$ has two

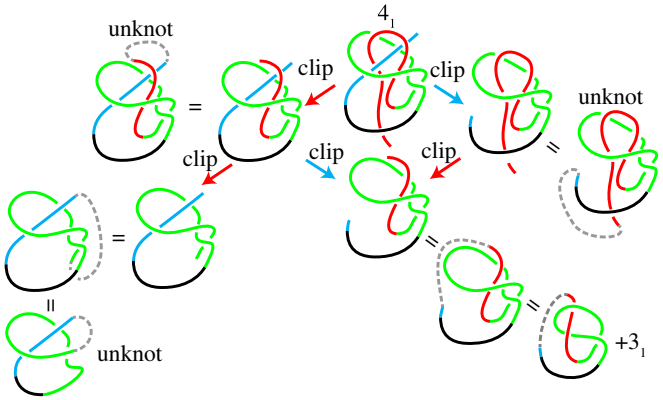

Fig. 9. The knot fingerprint of a 4_1 knot with configuration $RL(+, 0, -2)$ agrees with those of the 4_1-knotted KARIs.

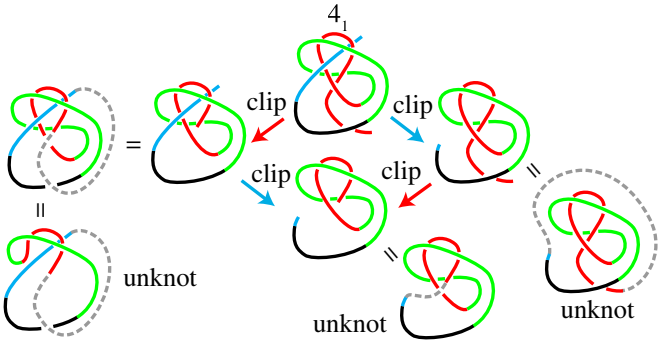

Fig. 10. The knot fingerprint of the 4_1 knot with configuration $RR(+,-1,0)$ is consistent with the 4_1-knotted phytochromes.

occurrences of the $+3_1$ separated by an unknot. This coincides with the fingerprints of some $+3_1$-knotted proteins (designated by $+3_1+3_1$ on KnotProt). Since most $+3_1$-knotted proteins have only one $+3_1$ knot in their fingerprint, we consider other configurations as well.

After $RR(+,0,0)$, the configurations for $+3_1$ with the least twisting are $RL(+,0,-1)$ and $LR(+,-1,0)$. We see in Fig. 8 that the fingerprint for $RL(+,0,-1)$ has only one occurrence of the $+3_1$ knot. In particular, at the top and center, we illustrate $RL(+,0,-1)$. Then, we clip the blue end on the right and the red end on the left (as indicated by the colored arrows). In both cases, we get the unknot. If we clip either or both ends any farther, we also get the unknot. Thus, the pathways described by $RL(+,0,-1)$ could correspond to folding pathways for $+3_1$-knotted proteins. We show in *SI Appendix*, Fig. S29 that the fingerprint of $LR(+,-1,0)$ contains a $+5_2$ knot, and hence, $LR(+,-1,0)$ is unlikely to describe folding pathways for the $+3_1$-knotted proteins.

The configurations $RR(+,0,0)$ and $RL(+,0,-1)$ are consistent with a twisted hairpin pathway (as shown by Theorem 3), and hence, the agreement of their fingerprints with KnotProt provides evidence that the $+3_1$-knotted proteins fold via a twisted hairpin pathway. By taking the mirror images of the configurations for $+3_1$, we obtain the configurations $LL(-,0,0)$ and $LR(-,0,1)$, which could describe pathways for the folding of -3_1-knotted proteins.

Fingerprints of 4_1-Knotted Proteins. We begin by considering the fingerprints of the 4_1-knotted KARIs. According to KnotProt, all subchains obtained by clipping either end alone are unknots, but removing a sufficient number of residues from both ends produces a $+3_1$ knot. We show in Fig. 9 that the fingerprint of the configuration $RL(+,0,-2)$ for the 4_1 knot agrees with this. Since this configuration is consistent with a twisted hairpin pathway by Theorem 3, our theory supports such a folding pathway for the KARIs.

However, $RL(+,0,-2)$ is not the only configuration for the 4_1 knot, which is consistent with a twisted hairpin pathway. In *SI Appendix*, we show that, among all of the configurations for 4_1 that are consistent with a twisted hairpin pathway, the only one other than $RL(+,0,-2)$ with a fingerprint that agrees with the KARIs is $LL(+,-1,-2)$. However, the configuration $LL(+,-1,-2)$ requires that both loops contain twists, while $RL(+,0,-2)$ requires only one loop to have twists. Because of the Minimal Twisting Principle, we propose that the KARIs are more likely to fold according to the twisted hairpin pathways described by $RL(+,0,-2)$.

Next, we consider the fingerprints for the 4_1-knotted phytochromes. According to KnotProt, no matter how much either or both ends of the phytochromes are clipped, the 4_1 knot is the only nontrivial knot that can be obtained. In *SI Appendix*, we show that all of the configurations for 4_1 that are consistent with a twisted hairpin pathway contain either a $\pm3_1$ knot or a $\pm6_1$ knot in their fingerprint. Thus, we suggest that the 4_1-knotted phytochromes fold via a configuration that is inconsistent with a twisted hairpin pathway.

Fig. 10 illustrates the fingerprint of the configuration $RR(+,-1,0)$ for the 4_1 knot. Clipping either or both ends enough to change the knot type results in the unknot. Thus, the fingerprint of $RR(+,-1,0)$ agrees with those of the phytochromes on KnotProt. Since the 4_1 knot is achiral, it follows that the fingerprint of its mirror form $LL(-,1,0)$ also agrees with those of the phytochromes on KnotProt.

In fact, we show in *SI Appendix* that the fingerprints of all of the configurations for 4_1 that are inconsistent with a twisted hairpin pathway contain no knots other than the 4_1. Thus, any of these configurations could describe the folding pathways of the phytochromes. However, the configurations $RR(+,-1,0)$ and $LL(-,1,0)$ are the only ones with just one twist. Thus, because of the Minimal Twisting Principle, we believe that one of these configurations is most likely to describe pathways for the folding of the phytochromes.

Observe that the knot fingerprints illustrated in Figs. 9 and 10 partition the 4_1-knotted proteins according to biological function. If these protein classes have different knotting pathways as indicated by their different configurations, it could suggest that the 4_1 knot plays a different functional role in the phytochromes than it does in the KARIs.

Fingerprints of 5_2-Knotted Proteins. The only proteins that are known to contain the -5_2 knot are the UCHs. These proteins are shallowly knotted; however, as shown on KnotProt, clipping the C terminus produces a -3_1 knot. There are no other knots in the fingerprints of the UCHs.

SI Appendix, Table S2 together with Theorem 3 show that the only configurations that are consistent with a twisted hairpin pathway for the -5_1 knot are $RL(-,1,2)$ and $LL(-,0,2)$. However, *SI Appendix*, Fig. S37 shows that, if we clip both ends of either of these configurations, we obtain a 4_1 knot. Since the fingerprints of the UCHs do not contain a 4_1 knot, the UCHs are unlikely to fold via a configuration that is consistent with a twisted hairpin pathway.

We see in Fig. 11 that the fingerprint of the configuration $RL(+,-1,0)$ agrees with that of the UCHs. In particular, by clipping the blue end, both ends, or the red end, we get the unknot, while clipping the red end more gives us a -3_1 knot. If we clip either end any farther, the unknot is produced. Additional support for this configuration comes from its intermediates shown in *SI Appendix*, Fig. S39, which agree with those obtained for the UCHs in computer simulations by Zhao et al. (29).

By the Minimal Twisting Principle, the folding pathways described by $RL(+,-1,0)$, which has only one twist, are more

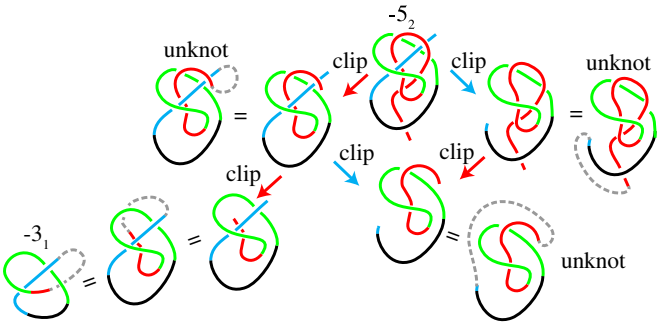

Fig. 11. The knot fingerprint of a -5_2 knot with configuration $RL(+,-1,0)$ agrees with those of the -5_2-knotted UCHs.

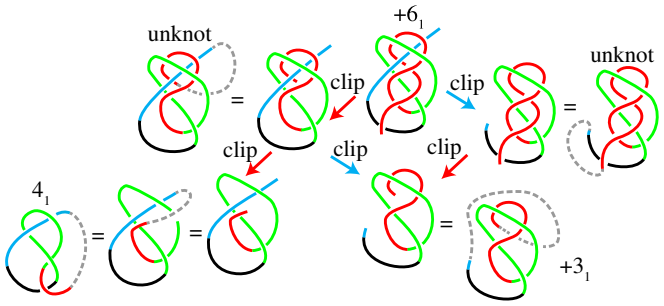

Fig. 12. The knot fingerprint of the knot $+6_1$ with configuration $RR(-, 2, -1)$ agrees with those of the $+6_1$-knotted DehIs.

likely than those described by a configuration with multiple twists in one loop or a single twist in each loop. The only other configuration for the -5_2 knot that has a single twist is $LR(+, 0, -1)$. However, we show in *SI Appendix*, Fig. S38 that the fingerprint of $LR(+, 0, -1)$ contains a 4_1 knot, which does not agree with the fingerprints for the UCHs on KnotProt. Thus, we assert that the configuration $RL(+, -1, 0)$ describes the most likely folding pathways for the UCHs.

Fingerprints of 6_1-Knotted Proteins. The only proteins known to contain a $+6_1$ knot are the DehIs. According to KnotProt, clipping either end of DehI a little yields an unknot, but clipping the N terminus a lot yields a 4_1 knot, and clipping both ends a moderate amount yields the $+3_1$ knot.

None of the configurations for $+6_1$ listed in Table 1 are of the form described in Theorem 3, and hence, no configuration for $+6_1$ is consistent with a twisted hairpin pathway. Thus, we begin with the $RR(-, 2, -1)$ configuration produced by the simulation of Bölinger et al. (24).

In Fig. 12, we see that clipping the blue end of $RR(-, 2, -1)$ produces the unknot. Clipping the red end a little also gives us an unknot, while clipping the red end a lot yields a 4_1 knot. Clipping both the red and blue ends gives us the $+3_1$ knot. Any additional clipping of either end enough to change the knot type yields the unknot. This fingerprint matches that of DehI on KnotProt, providing additional evidence that DehI may fold according to the pathways described by the simulations of Bölinger et al. (24).

To determine if another configuration could also describe the folding of the $+6_1$ knot in DehI, we considered the simplified illustration of the crystal structure obtained by Wang et al. (43). *SI Appendix*, Fig. S41 shows that, with only very minor changes, this simplified crystal structure corresponds to the configuration $LR(-, 1, -1)$.

In Fig. 13, we illustrate the fingerprint of $LR(-, 1, -1)$. As shown, clipping the red end a little yields the unknot, and clipping it more substantially yields the 4_1 knot. If we clip the blue end alone, we again get the unknot. However, if we clip both the red end and the blue end, we obtain the $+3_1$ knot. Additional clipping of either end enough to change the knot type yields the unknot.

Thus, both $RR(-, 2, -1)$ and $LR(-, 1, -1)$ have fingerprints that agree with DehI on KnotProt, and hence, either could describe the folding pathways of DehI. However, since $LR(-, 1, -1)$ requires less twisting and corresponds to the simplified crystal structure for DehI (43), the folding pathways described by $LR(-, 1, -1)$ are a reasonable alternative to the pathways described by $RR(-, 2, -1)$.

7. Discussion

Motivated by the simulations of Bölinger et al. (24) for the knotting of DehI, we introduced a theory that could describe folding pathways for any knotted protein. We expressed our theory in terms of steps that are encoded by the configurations in Fig. 6. Since multiple configurations produce the same knot (as listed in Table 1), our theory shows that different families of proteins containing the same knot could fold in distinct ways. This would apply, for example, to the KARIs and the phytochromes, which both fold into a 4_1 knot.

The differences between our theory and the twisted hairpin theory introduced by Taylor (23) are the number of folding pathways of a given knot, the number of loops, threading vs. loop flipping, and the possibility of knotted intermediates. According to Taylor's theory, all knots occur as the result of a terminus threading through the single loop of a twisted hairpin as illustrated in Fig. 2. This means that the complexity of a knot is entirely the result of the twists in the hairpin, and there can be no knotted intermediates. By contrast, according to our theory, a loop-flipping move causes a terminus to be threaded through two loops that are closely aligned but have no more than two twists each. Our theory produces two parallel folding pathways, which can each lead to knotted intermediates. This is consistent with recent experimental and computational results (24, 29, 35, 39).

In *6. Knot Fingerprints of Configurations*, we compared fingerprints of knotted proteins obtained by KnotProt with fingerprints of particular configurations. Our results show that the fingerprints of the configurations $RR(+, 0, 0)$ and $RL(+, 0, -1)$ for the $+3_1$-knotted proteins (Fig. 8), $LL(-, 00)$ and $LR(-, 0, 1)$ for the -3_1-knotted proteins, and $RL(+, 0, -2)$ for the 4_1-knotted KARIs (9) contain the same knots as the fingerprints for these proteins on KnotProt. Since these configurations are consistent with twisted hairpin pathways as shown in Theorem 3, our theory supports Taylor's twisted hairpin theory in these cases.

However, the knots in the fingerprints for the 4_1-knotted phytochromes, the -5_2-knotted UCHs, and the $+6_1$-knotted DehI do not correspond to those of configurations that are consistent with a twisted hairpin pathway. Rather, they agree with the configurations $RR(+, -1, 0)$ for the phytochromes (Fig. 10), $RL(+, -1, 0)$ for the UCHs (Fig. 11), and $RR(-, 2, -1)$ (Fig. 12) and $LR(-, 1, -1)$ (Fig. 13) for DehI. Thus, these configurations describe pathways that could correspond to the folding of these proteins. Furthermore, the configuration $LR(-, 1, -1)$ resembles the simplified crystal structure for DehI found by Wang et al. (43) and requires less twisting than $RR(-, 2, -1)$. Thus, the pathways described by the configuration $LR(-, 1, -1)$ could be a good alternative to both the pathways described by the simulation of Bölinger et al. (24) and the twisted hairpin pathway proposed by Taylor (23).

While Taylor's twisted hairpin theory of knot folding may be correct for most knotted proteins with three or four crossings, our results show that, for more complex protein knots, there may be other viable folding pathways. Furthermore, Taylor's theory predicts that all future protein knots will be members of the

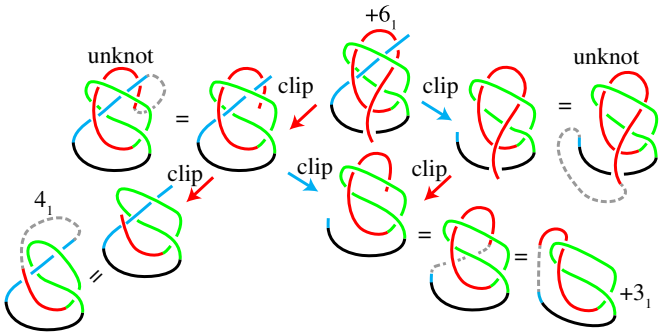

Fig. 13. The knot fingerprint of the knot $+6_1$ with configuration $LR(-, 1, -1)$ agrees with those of the $+6_1$-knotted DehIs.

twist knot family, whereas our theory predicts that some nontwist knots in the family of two-bridge knots might also eventually be found in proteins. Nonetheless, we predict that only the 14 knots listed in Table 1 (or their enantiomers) are likely to occur in proteins, although this list would be longer if we allowed three twists in each loop.

We see from Table 1 and *SI Appendix*, Table S2 that the $+3_1$, -3_1, 4_1, and -5_2 knots (which have been found in proteins) have 12, 12, 16, and 12 configurations, respectively, while enantiomers of the knots 5_1, 7_2, 7_5, 7_6, 7_7, 8_8, 8_{14}, and 9_{23} (which have not been identified in proteins) have only 5, 4, 4, 4, 4, 2, 4, and 1 configurations, respectively. Since configurations describe folding pathways, this means that, according to our theory, there are fewer pathways to obtain the knots in the latter group than to obtain those in the former group. This may explain why none of the latter knots have been found thus far in any protein. This is not surprising for complex knots with seven or more crossings, but it offers a different hypothesis for why the 5_1 knot has not been found.

The situation for the six-crossing knots is somewhat more subtle. The $+6_1$ knot has eight configurations as does the achiral 6_3 knot as well as each enantiomer of the 6_2 knot. However, all eight of the configurations for the 6_3 knot require at least one loop to contain two twists. Thus, by the Minimal Twisting Principle, we believe that, of the three six-crossing knots, the 6_3 knot is the least likely to occur in a protein. By contrast, for each enantiomer of the 6_2 knot, four of its eight configurations require no more than one twist in each loop. We compare this with the $+6_1$ knot, where only two of its eight configurations require no more than one twist in each loop. Because the $+6_1$ knot has been found in DehI, we predict that at least one of the enantiomers of the 6_2 knot will eventually be identified in a protein. If this turns out to be the case, it would be the first nontwist knot found in a protein.

8. Materials and Methods

Method for Obtaining Table 1. *SI Appendix*, Table S1 lists all possible configurations. To determine the knot types associated with these configurations, we started with a configuration and deformed it into a knot projection in the standard knot tables (18). Next, we applied Theorem 1 to that configuration to obtain other configurations that represent the same knot. We did this repeatedly until all 200 configurations listed in *SI Appendix*, Table S1 were identified. We then reorganized the information into Table 1 and *SI Appendix*, Table S2, which group the information by knot type rather than by the parameters of the configuration. Note that the configurations for the negative forms of the chiral knots in *SI Appendix*, Table S2 can be deduced from the configurations for their positive forms in Table 1 by applying Lemma 1; also, Table 1 does not include the unknot, which is among the configurations listed in *SI Appendix*, Table S1.

As an example, below we show how all configurations for the 4_1 knot were obtained. We begin by deforming the configuration $RR(+,-1,0)$ to the standard projection of 4_1 in Fig. 14.

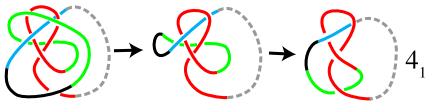

Fig. 14. $RR(+,-1,0)$ is the 4_1 knot.

Next, we apply Theorem 1 to the configuration $RR(+,-1,0)$ to conclude that the following configurations also produce the 4_1 knot: $RR(+,0,-1)$, $RL(+,-1,-1)$, $RL(+,0,-2)$, $LR(+,-1,-1)$, $LR(+,-2,0)$, $LL(+,-2,-1)$, and $LL(+,-1,-2)$. Since 4_1 is achiral, we can apply Lemma 1 to conclude that 4_1 is also produced by the configurations $LL(-,1,0)$, $LL(-,0,1)$, $LR(-,1,1)$, $LR(-,0,2)$, $RL(-,1,1)$, $RL(-,2,0)$, $RR(-,2,1)$, and $RR(-,1,2)$. This gives us all 16 configurations for 4_1 that are listed in Table 1.

Methods for Computing Knot Fingerprints. KnotProt (4, 5) determines fingerprints by starting with a crystal structure from the PDB, which is positioned in the center of a large ball. The termini are then extended to the boundary of the ball in several hundred randomly chosen directions. In each case, the endpoints are joined together by an arc in the boundary of the ball. To identify the knot type of the closed loop, the Alexander polynomial is computed. If the knot is chiral, the HOMFLYPT is computed to identify the exact enantiomer. Of the several hundred knot types obtained in this way, the most frequently occurring one is then assigned to the protein. To determine the subknots in a protein, the endpoints are clipped to specified residues, and the same method is used.

In contrast with KnotProt, we determine the fingerprints of our configurations qualitatively rather than quantitatively. In particular, we do not do a probabilistic analysis of hundreds of ways to extend the ends of a configuration to the boundary of a ball to get a closed knot. Rather, we assert that it is possible to draw the configurations and subchains as we have in *6. Knot Fingerprints of Configurations* and join the ends with dotted arcs in such a way that the knots that we obtain are indeed the most probable ones.

For simplicity, we do not include wiggles when we draw configurations, although wiggles occur in protein conformations and are important, because they can result in slipknots in subchains (45, 51, 53). Also, wiggles can cause the same knot to appear multiple times in a fingerprint, separated briefly by an unknot. For example, according to KnotProt, the fingerprint for the protein 5m4sA is $+3_1+3_1$, meaning that the entire chain contains a $+3_1$ knot, but there is also a $+3_1$ knot in a subchain. As we see in *SI Appendix*, Fig. S28, this fingerprint occurs for the configuration $RR(+,0,0)$. In *SI Appendix*, Fig. S29, we show that this same fingerprint can also occur with the configuration $RL(+,0,-1)$ by adding a wiggle. All of the fingerprints on KnotProt with multiple occurrences of a given knot can be similarly obtained from those in *6. Knot Fingerprints of Configurations* by adding wiggles at appropriate places.

ACKNOWLEDGMENTS. We thank Gregory Buck, Sophie Jackson, and Ken Millett for helpful conversations and the anonymous referees for their very constructive feedback. E.F. and H.W. thank the Institute for Advanced Study (IAS) and Carleton College for their hospitality while working on this project. E.F. and A.H. were supported in part by NSF Grant DMS-1607744, and H.W. was supported by NSF Grants DMS-1510453 and DMS-1841221 and a von Neumann Fellowship at the IAS.

1. Mansfield ML (1994) Are there knots in proteins? *Nat Struct Mol Biol* 1:213–214.
2. Taylor WR (2000) A deeply knotted protein structure and how it might fold. *Nature* 406:916–919.
3. Nureki O, et al. (2002) An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr D Biol Crystallogr* 58:1129–1137.
4. Jamroz M, et al. (2015) KnotProt: A database of proteins with knots and slipknots. *Nucleic Acids Res* 43:D306–D314.
5. Sułkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A (2012) Conservation of complex knotting and slipknotting patterns in proteins. *Proc Natl Acad Sci USA* 109:E1715–E1723.
6. Kolesov G, Virnau P, Kardar M, Mirny LA (2007) Protein knot server: Detection of knots in protein structures. *Nucleic Acids Res* 35:W425–W428.
7. Lai YL, Chen CC, Hwang JK (2012) pKnot v.2: The protein KNOT web server. *Nucleic Acids Res* 40:W228–W231.
8. Marcone B, Orlandini E, Stella A, Zonta F (2004) What is the length of a knot in a polymer? *J Phys A Math Gen* 38:L15–L21.
9. Millett KC, Sheldon BM (2005) Tying down open knots: A statistical method for identifying open knots with applications to proteins. *Physical and Numerical Models in Knot Theory*, eds Calvo JA, Millet KC, Rawdon EJ, Stasiak A (World Scientific, Singapore), pp 203–217.
10. Lua RC, Grosberg AY (2006) Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput Biol* 2:e45.
11. Virnau P, Mirny LA, Kardar M (2006) Intricate knots in proteins: Function and evolution. *PLoS Comput Biol* 2:e122.
12. Khatib F, Weirauch MT, Rohl CA (2006) Rapid knot detection and application to protein structure prediction. *Bioinformatics* 22:e252–e259.
13. Panagiotou E, et al. (2011) A study of the entanglement in systems with periodic boundary conditions. *Prog Theor Phys Suppl* 191:172–181.
14. Tubiana L, Orlandini E, Micheletti C (2011) Probing the entanglement and locating knots in ring polymers: A comparative study of different arc closure schemes. *Prog Theor Phys Suppl* 191:192–204.
15. Millett KC, Rawdon EJ, Stasiak A, Sułkowska JI (2013) Identifying knots in proteins. *Biochem Soc Trans* 41:533–537.
16. Alexander K, Taylor AJ, Dennis MR (2017) Proteins analysed as virtual knots. *Sci Rep* 7:42300.
17. Goundaroulis D, Dorier J, Benedetti F, Stasiak A (2017) Studies of global and local entanglements of individual protein chains using the concept of knotoids. *Sci Rep* 7:6309.
18. Cromwell PR (2004) *Knots and Links* (Cambridge Univ Press, Cambridge, UK).
19. Liang C, Cerf C, Mislow K (1996) Specification of chirality for links and knots. *J Math Chem* 19:241–263.

20. Liang C, Mislow K (1994) A left-right classification of topologically chiral knots. *J Math Chem* 15:35–62.
21. Jackson SE, Suma A, Micheletti C (2017) How to fold intricately: Using theory and experiments to unravel the properties of knotted proteins. *Curr Opin Struct Biol* 42: 6–14.
22. Dabrowski-Tumanski P, Sulkowska JI (2017) To tie or not to tie? That is the question. *Polymers* 9:454.
23. Taylor WR (2007) Protein knots and fold complexity: Some new twists. *Comput Biol Chem* 31:151–162.
24. Bölinger D, et al. (2010) A Stevedore's protein knot. *PLoS Comput Biol* 6: e1000731.
25. Lim NCH, Jackson SE (2015) Molecular knots in biology and chemistry. *J Phys Condens Matter* 27:354101.
26. Mallam AL, Jackson SE (2011) Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol* 8:147–153.
27. Sułkowska JI, et al. (2013) Knotting pathways in proteins. *Biochem Soc Trans* 41:523–527.
28. Niewieczerzal S, Sulkowska J (2017) Knotting and unknotting proteins in the chaperonin cage: Effects of the excluded volume. *PLoS One* 12:23.
29. Zhao Y, Dabrowski-Tumanski P, Niewieczerzal S, Sulkowska JI (2018) The exclusive effects of chaperonin on the behavior of proteins with 5_2 knot. *PLoS Comput Biol* 14:e1005970.
30. Sulkowska J, Zhao Y, Dabrowski-Tumanski P, Niewieczerzal S (2018) The exclusive effects of chaperonin on the free energy landscape of proteins with complex knots. *Biophys J* 114:552a–553a.
31. Wallin S, Zeldovich KB, Shakhnovich EI (2007) The folding mechanics of a knotted protein. *J Mol Biol* 368:884–893.
32. Covino R, Skrbic T, Beccara Sa, Faccioli P, Cristian M (2013) The role of non-native interactions in the folding of knotted proteins: Insights from molecular dynamics simulations. *Biomolecules* 4:1–19.
33. Chwastyk M, Cieplak M (2015) Cotranslational folding of deeply knotted proteins. *J Phys Condens Matter* 27:354105.
34. Dabrowski-Tumanski P, Piejko M, Niewieczerzal S, Stasiak A, Sulkowska JI (2018) Protein knotting by active threading of nascent polypeptide chain exiting from the ribosome exit channel. *J Phys Chem B* 122:11616–11625.
35. Andersson FI, Pina DG, Mallam AL, Blaser G, Jackson SE (2009) Untangling the folding mechanism of the 5_2-knotted protein uch-l3. *FEBS J* 276:2625–2635.
36. Andersson FI, et al. (2011) The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1. *J Mol Biol* 407:261–272.
37. Lee YTC, et al. (2017) Entropic stabilization of a deubiquitinase provides conformational plasticity and slow unfolding kinetics beneficial for functioning on the proteasome. *Sci Rep* 7:45174.
38. Lou SC, et al. (2016) The knotted protein UCH-L1 exhibits partially unfolded forms under native conditions that share common structural features with its kinetic folding intermediates. *J Mol Biol* 428:2507–2520.
39. Zhang H, Jackson SE (2016) Characterization of the folding of a 5_2-knotted protein using engineered single-tryptophan variants. *Biophys J* 111:2587–2599.
40. Tuszynska I, Bujnicki JM (2010) Predicting atomic details of the unfolding pathway for yibk, a knotted protein from the spout superfamily. *J Biomol Struct Dyn* 27:511–520.
41. Beccara Sa, Škrbić T, Covino R, Micheletti C, Faccioli P (2013) Folding pathways of a knotted protein with a realistic atomistic force field. *PLoS Comput Biol* 9:e1003002.
42. Chwastyk M, Cieplak M (2015) Multiple folding pathways of proteins with shallow knots and co-translational folding. *J Chem Phys* 143:045101.
43. Wang I, Chen SY, Hsu STD (2016) Folding analysis of the most complex Stevedore's protein knot. *Sci Rep* 6:31514.
44. Capraro DT, Jennings PA (2016) Untangling the influence of a protein knot on folding. *Biophys J* 110:1044–1051.
45. Sułkowska JI, Noel JK, Onuchic JN (2012) Energy landscape of knotted protein folding. *Proc Natl Acad Sci USA* 109:17783–17788.
46. Soler MA, Nunes A, Faísca PFN (2014) Effects of knot type in the folding of topologically complex lattice proteins. *J Chem Phys* 141:025101.
47. Sorokina I, Mushegian A (2016) The role of the backbone torsion in protein folding. *Biol Direct* 11:64.
48. Sorokina I, Mushegian A (2018) Modeling protein folding in vivo. *Biol Direct* 13:13.
49. Banavar JR, Maritan A (2003) Colloquium: Geometrical approach to protein folding: A tube picture. *Rev Mod Phys* 75:23–34.
50. Schubert H (1954) Über eine numerische knoteninvariante. *Math Z* 61:245–288.
51. King NP, Yeates EO, Yeates TO (2007) Identification of rare slipknots in proteins and their implications for stability and folding. *J Mol Biol* 373:153–166.
52. Taylor WR (2005) Protein folds, knots and tangles. *Physical and Numerical Models in Knot Theory: Including Applications to the Life Sciences* (World Scientific, Singapore), pp 171–202.
53. Noel JK, Sułkowska JI, Onuchic JN (2010) Slipknotting upon native-like loop formation in a trefoil knot protein. *Proc Natl Acad Sci USA* 107:15403–15408.