A specific set of heterogeneous native interactions yields efficient knotting in protein folding

João NC Especial^{1,2} & Patrícia FN Faisca^{1,2*}

¹Departamento de Física and ²BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, Lisboa, Portugal

*Corresponding author:

PFNF, email: pffaisca@fc.ul.pt

Abstract

Native interactions are crucial for folding, and non-native interactions appear to be critical for efficiently knotting proteins. Therefore, it is important to understand both their roles in the folding of knotted proteins. It has been proposed that non-native interactions drive the correct order of contact formation, which is essential to avoid backtracking and efficiently self-tie. In this study we ask if non-native interactions are strictly necessary to tangle a protein, or if the correct order of contact formation can be assured by a specific set of native, but otherwise heterogeneous, interactions. In order to address this problem we conducted extensive Monte Carlo simulations of lattice models of proteinlike sequences designed to fold into a pre-selected knotted conformation embedding a trefoil knot. We were able to identify a specific set of heterogeneous native interactions that drives efficient knotting, and is able to fold the protein when combined with the remaining native interactions modeled as homogeneous. This specific set of heterogeneous native interactions is strictly enough to efficiently self-tie. A distinctive feature of these native interactions is that they do not backtrack, because their energies ensure the correct order of contact formation. Furthermore, they stabilize a knotted intermediate state, which is en-route to the native structure. Our results thus show that - at least in the context of the adopted model - non-native interactions are not necessary to knot a protein. However, when they are taken into account into protein energetics it is possible to find specific, non-local non-native interactions that operate as a scaffold that assists the knotting step.

Keywords: knotted proteins; protein folding; lattice model; Monte Carlo

Introduction

Knotted proteins embed a physical (i.e. open) knot in their native structure. The development of knot detection methods and loop closure procedures allowed determination of the fraction of knotted proteins in the protein data bank (PDB), as well as their knot type¹. Knotted proteins exhibit most frequently the trefoil (or 3_1) knot, but there are also proteins with 4_1 , 5_2 , and even one protein with the so-called Stevedore's knot that comprises 6 crossings². Depending on its location within the polypeptide chain the knot can be classified as shallow (when removing just a few residues from one of the termini is enough to untangle the fold) or deep.

These topologically complex proteins have attracted considerable attention and there is a large and growing body of experimental and theoretical work dedicated to understand their folding and knotting mechanisms³⁻¹³, their folding/unfolding properties¹⁴⁻¹⁸, their biological role¹⁹⁻²¹, and how they evolved²²⁻²⁴, just to mention a few examples.

Tracing and characterizing the folding pathway(s) is crucial to understanding the folding process of a globular protein. However, most standardly used experimental methods cannot yet adequately resolve the folding pathway with atomic detail and hence one must resort to molecular simulation in order to gain insights towards this goal. So far, several computational models based on protein representations with different levels of coarse graining, ranging from simple lattice to full atomistic representations, combined with different force fields and sampling methods, have been used to explore different aspects of the folding process of tangled proteins^{25,26}.

Shakhnovich and co-workers carried out the first computational study³ that provided the very first vistas on the knotting mechanism of protein YibK (PDB ID: 1j85), which embeds

a deep trefoil knot located 77 residues away from the N-terminus and 39 residues away from the C-terminus. They combined Langevin Dynamics simulations with a C-alpha representation to find that knotting occurs primarily in compact, near native conformations (with 80% of its native contacts formed), where part of the polypeptide chain is arranged into the so-called knotting loop, which is threaded by the C-terminus; the threading step occurs either directly (i.e. without bending) as observed for the vast majority of trajectories, or through an intermediate conformation in which the terminus arranges itself into an hairpin, and first gives rise to a pseudo-knotted conformation called a slipknot⁵⁸. Interestingly, the authors reported a non-negligible probability of observing the formation of the knot in less structurally consolidated conformations (with only 20% of their native contacts formed). The most striking conclusion of this study concerns the energetics of the knotting step. Indeed, it was reported that when only native interactions (i.e. interactions that are present in the native structure) are considered in protein energetics (as in the native centric Gō models²⁷), the conformations are too compact to allow for threading events, and none of the 100 attempted folding runs was successful. The authors reported that non-native interactions were essential to observe 100% folding efficiency (i.e. all attempted folding trajectories leading to native conformations). In particular, it was found that a specific set of ad-hoc non-native attractive interactions established between a stretch of residues within the knotting loop (86-93), and residues pertaining to the C-terminus assist the threading step.

A subsequent study by Faccioli and co-workers²⁸ investigated the folding and knotting processes of AOTCase (PDB ID: 2g68), a 332 residue long protein embedding a deep trefoil knot in its native structure. The authors used Monte Carlo simulation of a C-alpha

representation combined with an intramolecular energy potential that besides native interactions, also considers sequence-specific non-native interactions (i.e. non-native interactions that are encoded in the primary sequence) in the protein folding energetics. The outcome of this study provided further support for a functional role played by non-native interactions in the folding of knotted proteins by showing that (when compared with a purely native-centric potential) they enhance the probability to knot the polypeptide chain along the whole folding process (including conformations with a relatively low degree of structural consolidation), while assisting a knotting step based primarily on a direct threading movement of the C-terminus through loops formed by other parts of the chain. Similarly, a native centric potential complemented with (sequence-dependent) non-native interactions also leads to an enhancement of knotting frequency in association with a direct threading movement of the C-terminus for protein MJ0366²⁹ (PDB ID: 2efv). The latter is the smallest knotted protein identified so far and features a shallow knot. Interestingly, a direct movement of the C-terminus was also observed in full atomistic Molecular Dynamics simulations of MJ0366 based on a classic force field²⁹, reinforcing the idea that this is likely the mechanism that nature uses to tangle this protein. Overall, these results do suggest that non-native interactions play a functional role in the folding of knotted proteins by assisting (or even driving) a knotting step that is based on a direct movement of the protein's C-terminus. Furthermore, they indicate that knotted conformations can form rather early in the folding pathway, in conformations which show a small degree of structural consolidation. Based on the analysis of folding trajectories obtained for MJ0366, Faccioli and co-workers suggested that non-native interactions promote the correct order of native contact formation, which is crucial for a protein to efficiently self-tie²⁹. Indeed, it is widely accepted that the folding of knotted proteins stands out for being remarkably

ordered³⁰ (e.g. in the case of MJ0366 the two beta strands that exist in its native structure need to form the beta-sheet prior to any threading events²⁹). At a microscopic level, an ordered folding process means that the establishment of native contacts must follow a specific sequence of events^{8,20}. This ordered folding is the reason why knotted proteins are especially prone to backtracking, i.e., the formation, breaking and re-establishment of contacts that form untimely³¹. Molecular simulation studies based on native centric potentials appear to be less efficient in tying a knotted protein, presumably because they favor the premature (and likely untimely) establishment of interactions that are in close proximity in the native conformation. Possibly because of backtracking, the folding efficiency of these models is lower, and knotting occurs later in nearly native conformations, mainly through the mechanism of slipknotting^{32,33}.

Since native interactions are critical for folding, and non-native interactions appear to be important for efficient knotting, it is important to understand their relative role, and this has been pointed out as a major question in a recent review of this research theme³⁴. The results outlined above lead us to question if non-native interactions are strictly necessary to efficiently tangle a protein, or if it is possible that the correct order of contact formation can be assured by a specific set of native, but otherwise heterogeneous, interactions. This is the major question we propose to address in the present study, and in order to answer it we will use Monte Carlo simulations of lattice models. The reason is two-fold. Firstly, we are interested in determining the order of contact formation, and lattice models define unequivocally native and non-native contacts. Secondly, we are interested in exploring the folding and knotting behavior of proteinlike sequences with an intramolecular potential that treats identically the contribution of sequence-specific native and non-native interactions to

the protein energy. Such an intramolecular potential has a long history in the study of protein folding³⁵⁻³⁷. Our results show that specific (non-local) non-native interactions indeed play a role in assisting the threading movement of the terminus in a particular ensemble of conformations with roughly half of the native contacts, but, more importantly, they also clearly show that there is a specific set of native interactions which are able to ensure optimal knotting performance by driving a folding pathway with minimal backtracking, which is populated by a knotted intermediate state that forms relatively early during folding.

Models and methods

We use a simple lattice model in which amino acids are represented by beads of uniform size that occupy the vertices of a regular cubic lattice. To enforce the excluded volume constraint, only one bead is allowed per lattice site. Consecutive beads along the chain are connected by sticks (with length of one lattice spacing) that represent the peptide bond. Intramolecular contacts between two non-bonded beads occur if they are separated by one lattice spacing. Intramolecular contacts are classified either as native, if they are present in the native conformation, or non-native if they are established between any two beads that are not in contact in the native conformation.

To model protein energetics we consider the native centric Gō potential²⁷, and two sequence-specific potentials: the one developed by Shakhnovich³⁸, which we used in prior studies^{36,37,39,40}, and a new native-centric potential with heterogeneous interactions. In all three cases, energy is measured in reduced units.

The Gō potential

The total energy of a conformation defined by bead coordinates $\{\vec{r}_i\}$ is given by

$$E(\{\vec{r}_i\}) = \epsilon \sum_{i,j>i+2}^N \Delta_{nat}(\vec{r}_i - \vec{r}_j)$$

In the equation above, ϵ is a uniform intramolecular energy parameter (taken as -1 in this study), N is the chain length measured in number of beads, and the so-called contact function, Δ_{nat} , is unity, if and only if, beads i and j are establishing a native contact.

Sequence specific potentials

In the case of a sequence specific potential the total energy of a conformation is given by the Shakhnovich potential

$$E(\{\vec{r}_i\}, \{\sigma_i\}) = \sum_{i,j>i+2}^N \epsilon(\sigma_i, \sigma_j) \Delta(\vec{r}_i - \vec{r}_j)$$

where $\{\sigma_i\}$ represents a specific sequence of amino acids that was designed to have minimum energy in the native conformation⁴¹, σ_i represents the chemical identity of bead i , and the intramolecular energy interaction parameters $\epsilon(\sigma_i, \sigma_j)$ are taken from the 20×20 Miyazawa-Jernigan interaction matrix⁴². Unlike for the Gō potential, the contact function, Δ , is unity if beads i and j are establishing an intramolecular contact, which means that both native and non-native interactions contribute to the energy of a conformation.

The difference between the Shakhnovich potential and the other sequence-specific potentials used in this study is that, in the latter, the contact function Δ_{ij} is unity if and only if beads i and j are establishing a native contact or some specific non-native contact(s).

The folding process is simulated by sampling canonically distributed conformational states generated using Metropolis Monte Carlo⁴³ combined with replica exchange⁴⁴. The conformational space was explored using a move set that comprised end-, corner-flip and crankshaft moves³⁹. Following previous studies^{9,13,20,33,45} we used the fraction of native contacts, Q , as a reaction coordinate, i.e., as an indicator of folding progress, with $Q \sim 0$ representing the denatured ensemble and $Q = 1$ representing the native state.

Simulations start from an unfolded conformation and terminate after 10^{11} Monte Carlo steps. The first half of the simulation is used to allow all replicas to relax to the equilibrium ensemble at their particular temperature, and relevant properties (E , Q , topological state, etc.) are then sampled during the second half. The topological state of a conformation (i.e. knotted or unknotted) was determined using the Koniaris–Muthukumar–Taylor (KMT) algorithm^{25,46}. The Weighted Histogram Analysis Method (WHAM)⁴⁷ was used to analyze the sampled data and produce maximum likelihood estimates of the density of states, from which expected values for thermodynamic properties were calculated as functions of temperature. WHAM was also used to project the density of states along the chosen reaction coordinate to obtain free energy and probability profiles along this coordinate. In particular, the specific heat capacity, C_V , defined as $C_V = (\langle E^2 \rangle - \langle E \rangle^2) / T^2$, was evaluated as a function of temperature, and the melting temperature, T_m , was determined as the temperature at which C_V reached its maximum value. The folding transition temperature, T_f , was determined as the temperature at which the free energy profile had two

minima of equal value, one near the unfolded state and the other near the native state. T_m and T_f should coincide for a two-state folding transition.

Conditional probability contact maps

To obtain a microscopic description of the folding pathway we used conditional probability contact maps (CPCMs), i.e., $N \times N$ matrices where each entry represents the probability of a contact (native or non-native) being formed in an ensemble of conformations that share the same Q under thermal equilibrium at T_m .

Sequence design

To design protein sequences we performed design in sequence space⁴¹. Starting from a random sequence of N amino acids, an optimized (i.e. designed) proteinlike sequence was generated so that the energy of the native structure was minimal. This was achieved by means of Monte Carlo simulated annealing. Briefly, once an initial annealing temperature was selected, the chemical identity of two randomly selected beads was permuted at each MC step, this move being accepted or rejected according to the Metropolis criterion, and the temperature decreased according to a predefined annealing schedule. Further details on the design procedure can be found elsewhere³⁹.

Structural clustering

In order to determine the relevant conformational classes present in an ensemble of conformations we used the hierarchical clustering algorithm *jclust* available in the MMTSB tool set⁴⁸. Since we are using a lattice model, clustering was based on contact map similarity. From each identified cluster we extracted the conformation that lies closest to the cluster's centroid.

Results

Model systems and proteinlike sequences

In this work we consider a lattice protein with chain length $N = 41$ which was designed by hand to embed a trefoil knot (a 3_1 knot) in its native structure⁴⁹, and which, henceforth, will be named $K3_1$ (Fig. 1A). The knotted core (KC) of $K3_1$, i.e., the minimal segment of the chain that contains the knot, extends from beads 2 to 21. Since it suffices to remove two beads to untangle the protein, $K3_1$ classifies as shallow.

Figure 1. Model systems. (A) Native conformation of $K3_1$ with knotted core (KC) highlighted in dark grey. Proteinlike sequences S1 (B), S2 (C), and S3 (D) and corresponding native energy.

Three proteinlike sequences, named S1, S2 and S3, were designed to have minimal (and similar) energy in the native structure of $K3_1$ (Fig. 1B-D). It is important to mention that all sequences S1-S3 have the same number of different amino acids (i.e. all of them have 0 Cys (C), 1 Met (M), 3 Phe (F), 3 Ile (I), 4 Leu (L), 2 Val (V), 0 Trp (W), 0 Tyr (Y), 2 Ala (A), 3 Gly (G), 3 Thr (T), 0 Ser (S), 2 Gln (Q), 0 Asn (N), 2 Glu (E), 4 Asp (D), 3 His (H),

0 Arg (R), 5 Lys (K) and 4 Pro (P)), and only differ in the order by which the amino acids are assigned to the beads.

Knotting transition

The major goal of this study is to determine if and how native and non-native interactions assist knotting. A possible strategy to address this problem is to perform a comparative analysis of the folding and knotting transitions of proteinlike sequences with different knotting performances, focusing on the behavior of native and non-native interactions. We define the knotting probability, p_{knot} , as being the fraction of knotted conformations in a given ensemble where all the conformations share the same Q . The dependence of p_{knot} on Q (evaluated at T_m) indicates that for all the three sequences S1-S3 the knotting transition is a sigmoidal curve. However, S1 displays the earliest knotting transition (i.e. knotting occurs in less structurally consolidated conformations), S3 the latest (i.e. knotting occurs in more structurally consolidated conformations), and S2 exhibits an intermediate profile, becoming knotted in the vicinity of the transition driven by the $G\ddot{o}$ potential (Fig. 2A).

Figure 2. Knotting and folding transitions. (A) Knotting probability (p_{knot}) as a function of the reaction coordinate (Q) for sequences S1-S3, and G0, and (B) free energy profiles (to allow for a direct comparison, the native energy for the G0 potential was normalized to be the same as in S1). The intermediate state I populated by S1 at $Q = 0.53$. (C) The probability contact map for the intermediate state I (above the diagonal), and the native contact map, comprising 40 native contacts (below the diagonal). The 16 most probable native contacts in the intermediate are shown in shades of green, saturated proportionately to the contact probability, and the remaining native contacts are similarly shown in shades of red. The 4 most probably non-native contacts are shown in shades of blue, and the remaining non-native contacts in shades of grey. (D) Representative conformations (centroids) of the two (knotted) clusters found in the conformational ensemble of the intermediate state I. In both conformations, the knotting loop is highlighted in red, and the threading terminus is highlighted in green.

To get a more quantitative measure of how each protein sequence determines (by enhancing or suppressing) the probability to knot at a certain Q , we adopted the convention that the transition starts when $p_{\text{knot}} = 0.2$ (and term $Q_{0.2}$ the corresponding Q value), and ends when $p_{\text{knot}} = 0.8$ (and term $Q_{0.8}$ the corresponding Q value). We focus on $Q_{0.8}$, which we

take as indicator that the transition is nearly complete. We found that S1 requires 20 native contacts to reach a p_{knot} higher than 0.8 ($Q_{0.8} = 0.49$), S2 requires 23 ($Q_{0.8} = 0.57$), the Gō sequence 24 ($Q_{0.8} = 0.60$), and S3 requires 28 native contacts ($Q_{0.8} = 0.68$). The different knotting performance of sequences S1-S3 makes them ideal model systems to explore our strategy.

Folding thermodynamics and intermediate states

Experimental results indicate that several knotted proteins with deep knots are prone to populate intermediate states^{4,50-52}. Because knotting is entropically costly^{53,54}, and intermediate states represent stabilized thermodynamic states, their formation along the folding pathway could serve the purpose of assisting the knotting step by decreasing the free energy barrier to knotting. In line with this hypothesis, and in order to determine what underlies the different knotting transitions of S1-S3, we investigated the existence intermediate states. The free energy profiles, i.e. the free energy projected along the reaction coordinate Q , show that the Gō sequence together with sequences S2 and S3 exhibit two-state transitions, but sequence S1, which is the one that shows the earliest knotting transition, populates an intermediate state (denoted by I in Fig. 2B) centered at $Q = 0.53$ (21 native contacts), i.e. slightly above the $Q_{0.8} = 0.49$ found for this sequence. Since the native conformation is conserved in the four model systems, the formation of the intermediate state (I) must result from the specific amino acid interactions encoded in S1.

To structurally characterize the intermediate state we analyzed an ensemble of 12544 conformations with 21 native contacts ($Q = 0.53$) sampled by the replica having T nearest to T_m (from above). In this conformational ensemble there are 11811 (94%) knotted conformations, which is consistent with S1 having $Q_{0.8} = 0.49$. The probability contact map

evaluated over the intermediate state ensemble (Fig. 2C) indicates that the intermediate state is stabilized by 20 contacts that form with highest probability, 16 of which are native (namely, 2-11, 2-15, 3-18, 3-20, 3-24, 3-30, 4-23, 5-20, 5-22, 10-21, 11-24, 12-15, 12-25, 15-30, 21-24 and 25-30), and 4 non-native (namely, 1-10, 1-20, 4-29 and 7-20).

By performing structural clustering over the ensemble of knotted conformations one is able to identify two clusters: a dominant cluster with 7118 (60%) elements, and a smaller cluster comprising 4693 (40%) elements. The dominant cluster is represented by a compact conformation, and the smaller cluster is represented by an extended C-tail conformation (Fig. 2D). Visual inspection shows that non-native contact (7-20) temporarily stabilizes the knotting loop in both conformations, while threading (the passage of the terminus through the knotting loop) is assisted by non-native contact (4-29) in the compact conformation, and by native contact (4-23) in the extended tail conformation.

Contact formation, ordered folding and backtracking

The results reported in the previous section suggest that the mechanism behind the particular knotting transition of S1 is the formation of a knotted intermediate state that is energetically stabilized by 16 native interactions, and 4 non-native interactions. The question, then, is why does S1 populate this intermediate state while the other two sequences - which fold to the same native structure, have nearly the same native energy and same amount of each different amino acid in their chemical composition – do not? To address this problem we decided to explore the individual behavior of each native and non-native contact by evaluating their probability profiles, i.e., their conditional probability of formation along the reaction coordinate Q . The rationale for doing so is the following. It is known that the folding of knotted proteins stands out for being remarkably ordered^{29,30,55}.

At a microscopic level this means that the establishment of native contacts must follow a specific sequence of events. The need for an ordered contact formation is the reason why knotted proteins are especially prone to backtracking^{20,32,56}, i.e., the formation, breaking and re-establishment of contacts that form untimely⁵⁷. Our working hypothesis is that the level of backtracking is smaller in S1 than in the other 2 sequences, possibly because the intramolecular energies that were assigned to each native contact of S1 drive their formation in an order that is compatible with the constructive accumulation of structure giving rise to the observed intermediate state.

We find that contact probability profiles come out in different shapes but, for native contacts, it is possible to distinguish two major profile types: a sigmoidal shaped probability profile (Fig. 3A), and an irregularly shaped profile, exhibiting humps, bumps and plateaus, which are indication of backtracking (Fig. 3B). Non-native contacts have somewhat bell shaped probability profiles, which often also exhibit considerable irregularities (Fig. 3C). Interestingly, the vast majority (35 out of 40) of native contacts in sequence S3, display this type of irregular profile (Fig. 3B), while for S1 it is possible to find 10 native contacts with a clear sigmoidal profile out of the 16 contacts that precede the knotting transition (Fig. 3A). These results thus indicate that sequence S1 is the one that exhibits a smaller level of backtracking, which allows the formation of a knotted intermediate that is on-pathway to the native state. The formation of this intermediate state is concomitant with an expressive lowering of the major free energy barrier (Fig. 2B), and thermodynamic stabilization of the full folding transition.

Figure 3. Contact probability profiles. Probability profiles of native contacts start at $p = 0$ ($Q = 0$) and end at $p = 1$ ($Q = 1$), whereas the probability profiles of non-native contacts may start at any value for $Q = 0$ and must end at $p = 0$ for $Q = 1$. (A) Profiles of the 16 native contacts that precede knot formation in sequence S1. These are the same contacts found to be most probable in this sequence's intermediate state (SI Table I). (B) Profiles of the same 16 native contacts in the folding pathway of S3. (C) Profiles of the 4 most probable non-local non-native contacts identified in the intermediate state of S1 (SI Table I). (D) Profiles of the same 4 non-native contacts in the folding pathway of S3.

Similarly to the knotting transition, the value of the reaction coordinate at which a native contact's probability profile crosses the 0.8 probability threshold is named the native contact's $Q_{0.8}$ (i.e. for all values of $Q \geq Q_{0.8}$, $p_{\text{contact}} \geq 0.8$), and is an indicator of the stage at which the native contact becomes locked (or permanently formed) during the folding process. The sequence in which the native contacts get locked is a possible way to characterize the folding pathway, and this sequence can be traced by sorting the native contacts by increasing order of their $Q_{0.8}$. We find that in the case of S1, 16 native contacts

have $Q_{0.8}$ lower than the $Q_{0.8} = 0.49$ of its knotting transition, and that these are exactly the same native contacts identified as most probable in the intermediate state I populated by this sequence.

Perhaps not surprisingly, when, for the native contacts of S1, we plot the (contact energy, $Q_{0.8}$) pairs we find a statistically significant trend [for S1 (S3): linear term has $t = 4.171$ (5.469), null hypothesis $p = 0.00017$ (3.05×10^{-6})] between $Q_{0.8}$ and the contact's intramolecular energy (SI: Fig. 1A, D), with the more stable native contacts tending to form before the less stable ones in less consolidated conformations. This means that, in principle, when designing a knotted protein, the sequence of native contact formation can be tuned by changing the relative contact energies of the native contacts.

For non-native contacts, the contact probability profile enables calculation of two indicators: $p_{\max}$ which is the maximum value reached by the conditional contact probability, and $Q_{\max}$ which is the value of the reaction coordinate at which $p_{\max}$ occurs.

Unlike $Q_{0.8}$ for native contacts, for non-native contacts, $Q_{\max}$, which indicates the stage in the folding pathway at which the contact is most relevant, is not significantly influenced by the contact's intramolecular energy [for S1 (S3): linear term has $t = -0.461$ (-0.758), null hypothesis $p = 0.645$ (0.449)] (SI: Fig. 1B, E). The stage at which a non-native contact becomes relevant during the folding process can thus only be a consequence of geometric and excluded volume constraints that result from already formed native contacts. However, when these constraints allow the formation of a non-native contact, its probability is strongly influenced by the contact's intramolecular energy [for S1 (S3): linear term has $t = -19.399$ (-17.611), null hypothesis $p < 2 \times 10^{-16}$ (2×10^{-16}), quadratic term has $t = -2.222$ (-

1.351), null hypothesis $p = 0.0269$ (0.178)] (SI: Fig. 1C, F) and the level of assistance that a non-native contact may give to the stabilization of temporary intermediate conformations is, consequently, strongly dependent on its contact energy.

Role of native and non-native interactions

The results reported in the previous sections indicate that the 16 native interactions that are the first to lock in the folding of S1 may, either alone, or together with specific non-native interactions, be responsible for the knotting performance of S1. However, it is possible that not all of the native interactions are strictly necessary for knotting. Since the ensemble representative of the intermediate state (I) populated by S1 contains knotted and unknotted conformations, we evaluated the frequency of native and non-native contacts, as well as the contact probability curves in each sub-ensemble. We find 5 non-local, non-native contacts (namely, 4-29, 1-10, 14-31, 1-20, 7-20) that are potentially important for knotting because they are clearly more frequent in the sub-ensemble of knotted conformations (SI: Table 1 and SI: Fig. 4A-E). On the other hand, we also identified 4 native contacts (namely, 25-30, 12-15, 4-23, 5-22, and 25-30) that may not be crucial for knotting, exhibiting similar probabilities in both sub-ensembles (SI Table I and SI Fig. 4F-I). In order to test these hypotheses we conducted several folding simulations of K3₁ under different interaction potentials that consider only native interactions, or selected combinations of native and non-native interactions (Table 1).

We start by assessing the role of native interactions in knotting by considering two intramolecular potentials. The first, which we term sequence-specific 40 (ss40), takes into account the native interactions of sequence S1, and ascribes null energies to all non-native contacts. The second, which we term ss16, conserves the interaction parameters of S1 for

the 16 native contacts found in the intermediate state, and combines them with homogeneous energies for the remaining 24 native contacts (so that the native state's energy of sequence S1 is preserved), and ascribes null energy to all non-native contacts.

Potential	Native inter. (from S1)	Native inter. (homogeneous)	Non-native inter.
ss40	40	0	0
ss40 ⁺	40	0	4-29 or 7-20
ss40 ⁺²	40	0	4-29 + 7-20
ss16	16	24	0
ss16 ⁺	16	24	4-29 or 7-20
ss16 ⁺²	16	24	4-29 + 7-20
ss16 ⁰	16	0	0
ss12 ⁺⁵	12	27	4-29 + 1-10 + 14-31 + 1-20 + 7-20
ss12 ⁰	12	0	0

Table 1. Intermolecular potentials used in this study. The 12 native interactions used in ss12 exclude native interactions 25-30, 12-15, 4-23 and 5-22 out of the 16 that are the first to lock in the folding of sequence S1.

The knotting probability profiles of K3₁ obtained under these two potentials are qualitatively similar, and show that the set of 16 native interactions are able to efficiently knot the chain. These two p_{knot} profiles, however, lag behind the p_{knot} profile of S1 between $Q = 0.45$ and $Q = 0.53$ and, hence, the knotting enhancement observed in S1 in this interval of the reaction coordinate cannot be explained by native contacts alone (Fig. 4A).

Figure 4: Role of native and non-native interactions in knotting. (A) Knotting probability and free energy profiles for potentials ss40 and ss16. (B) Profiles for potentials ss40⁺⁽⁴⁻²⁹⁾, ss40⁺⁽⁷⁻²⁰⁾ and ss40⁺². (C) Profiles for potentials ss16⁺⁽⁴⁻²⁹⁾, ss16⁺⁽⁷⁻²⁰⁾ and ss16⁺². (D) Profiles for potentials ss16⁰, ss12⁺⁵ and ss12⁰.

In fact, when, in addition to the native interactions, we consider non-native interactions (e.g. 4-29, or 7-20), we obtain the p_{knot} profiles labeled ss40⁺⁽⁴⁻²⁹⁾ and ss40⁺⁽⁷⁻²⁰⁾ (Fig. 4B), or ss16⁺⁽⁴⁻²⁹⁾ and ss16⁺⁽⁷⁻²⁰⁾ (Fig. 4C), which nearly coincide with the S1 profile in this interval. Adding both non-native interactions to ss40 (ss40⁺²) (Fig. 4B), or ss16 (ss16⁺²) (Fig. 4C), leads to even earlier knotting profiles.

The free energy profiles of ss40, ss16 show two major differences from that of sequence S1. In both cases, one observes the formation of a novel, more structurally consolidated ($Q \sim 0.7$), intermediate state (Fig. 4A). Structural clustering analysis over an ensemble of

16423 conformations with $Q = 0.7$ populated by ss16 reveals the existence of 5 clusters of knotted conformations (instead of only 2 detected for the intermediate state populated by S1) (SI Fig. 2). Secondly, and more importantly, the transition state (TS) that precedes the formation of the intermediate state (located at $Q = 0.43$ for S1 and $Q = 0.45$ for ss40) is substantially more stabilized in S1 than in ss40, but only 15% of its conformations are knotted. Under both potentials the thermodynamic stabilization of the TS is dominated by the energy component of the free energy ($U_{TS} = -14.95$ for S1, and $U_{TS} = -11.41$ for ss40), but in the case of S1 the latter represents 79% of the free energy, and only 59% in the case of ss40. The energy associated with non-native interactions represents 30% of the total energy of the TS preceding the formation of the intermediate of S1, which allows us to conclude that non-native interactions are responsible for its higher thermodynamic stability. Interestingly, a similar analysis carried out for the intermediate state of S1 reveals that its thermodynamic stabilization is also dominated by the energetic contribution, but the energy of non-native interactions represents only 12% of the total energy.

To test if the 16 native interactions were sufficient to drive the knotting transition, we considered a potential (ss16⁰) in which the 16 native contacts kept their S1 energies, and all other contacts (native and non-native) were given null energies. We find that the 16 native interactions are indeed sufficient to knot the chain through the formation of conformations with an even smaller degree of structural consolidation (Fig. 4D). It is interesting to note that the knotting mechanism changes, however. The contact that forms concomitantly with knotting is now (10-21) instead of (2-11) (SI Fig. 3A), which indicates that knotting is now completed through a mouse-trap motion^{10,29,59} rather than direct threading of the terminus.

The considered interactions, however, are not sufficient to fold the protein, which requires the contribution of all native interactions (Fig. 4D).

Are all of these 16 native interactions required for knotting? In order to answer this question, we considered yet another intermolecular potential that takes into account the S1 energies of the 12 native contacts and 5 non-native contacts (that are more relevant in the sub-ensemble of knotted conformations), ascribes null energies to all other non-native contacts, and uniform energies to the remain 27 native contacts so that the energy of the native structure is preserved. Remarkably, this potential (termed $ss12^{+5}$) leads to an even earlier knotting transition, showing that the corresponding interplay of native and non-native interactions is optimal to fold this particular lattice protein. It is worth mentioning that the addition of non-native interactions (in this case and the other cases analyzed before) does not change the free energy profiles, indicating that non-native interactions do not play a thermodynamic role in assisting the knotting step (Fig. 4D). Instead, the considered specific non-native interactions play a structural or kinetic role instead, by temporarily stabilizing the knotting loop or assisting the threading step (SI Fig. 3B).

A modification of this potential that switches-off the 27 native interactions and non-native interactions ($ss12^0$) leads to the earliest knotting transition observed in this study, but the protein is not able to fold (Fig. 4D), in line with what is observed for $ss16^0$.

Conclusions

The role of non-native interactions in the folding of knotted proteins has been pointed out as an important fundamental question in the field of protein folding. Results reported in the literature are consistent with the idea that, independently of the sampling scheme used to probe the conformational space (Monte Carlo or Molecular Dynamics), and of the

resolution of protein representation adopted in simulations (from single bead to full atomistic), non-native interactions are critical to knot the chain, at least when the latter involves a single threading event. While the mechanistic details according to which they operate remain elusive, it has been suggested that non-native interactions assist knotting by driving the correct order of contact formation thus avoiding backtracking.

Here we explored this hypothesis in the context of a lattice model. In lattice models, native and non-native contacts are defined unequivocally and, therefore, lattice models offer an appropriate framework to investigate the order of contact formation.

By using a sequence specific potential that naturally incorporates the participation of non-native interactions in protein energetics, we were able to find nativelike sequences with different knotting performances. In particular, when gauged against a native-centric Gō model, we found that some sequences are more efficient and knot in conformations with a small degree of structural consolidation, while others need a higher degree of structural consolidation to knot.

By making an in depth comparative study, which comprised a thermodynamics analysis, and a microscopic investigation at the level of each individual contact, we were able to find efficient knotting does not require non-native interactions. Notwithstanding, we were able to identify specific non-native interactions that play a functional role in the knotting process by acting like a scaffold that temporarily stabilizes the knotting loop, or assists the threading step. Efficient knotting is, on the other hand, driven by a critical set of specific native interactions, which do not backtrack, and contribute to stabilize an intermediate state that decreases the entropic cost to knotting. The reason behind the ‘ideal’ behavior of these

native interactions relies, at least to a significant extent, on their relative energies (with the more stable native interactions tending to consolidate first), and how these energies are distributed within the native fold to minimize energetic and topological bottlenecks, which are being particularly important in the case of knotted proteins. The fact that efficient knotting only requires 12 out of 40 native interactions (which represents a reduction of 30% of the original amino acid alphabet) is interesting *per se*, and suggests that while energetic heterogeneity is important to assure a well defined, and correct order of contact formation, a 20 amino acid alphabet may not be necessary to efficiently tie a polypeptide chain. In light of these results the principle of minimal frustration might thus be restated for the particular case of knotted proteins as: knotted proteins evolve towards primary structures that underlie sigmoidal shaped contact probability profiles for their native contacts. A relevant question is then, what is the minimal size of the amino acid alphabet that efficiently ties a protein chain?

In this study we conducted replica-exchange Monte Carlo simulations, because our major goal was to study equilibrium thermodynamics. Another interesting question is to investigate the performance of the utilized potentials in simulations at fixed temperature to investigate their kinetic performance. In particular, which potential knots the chain faster?

It is also pertinent to ask whether or not similar results might be obtained in the scope of an off-lattice model or if, on the other hand, non-native interactions are indeed critical to drive an efficient knotting mechanism in off-lattice models, and perhaps in real world natural proteins. One should bear in mind that conformational entropy is measured more accurately off-lattice; likewise, the energetic stabilization associated with non-native interactions at the level of the unfolded or partially folded states, may become only visible off-lattice.

Furthermore, it is also possible that non-native interactions may play a structural role (e.g. by enlarging the knotting loop to avoid steric clashes), which is smeared out by the discrete nature of the lattice representation.

The questions outlined above will be the focus of future work.

Author contributions

PFNF designed the research. JNCE performed the calculations. JNCE and PFNF analyzed the data. PFNF and JNCE wrote the paper.

Acknowledgments

Work supported by UID/MULTI/04046/2020 Research Unit grant from FCT, Portugal (through BioISI). JNCE acknowledges financial support from FCT, Portugal, through PhD grant SFRH/BD/144345/2019. PFNF would like to thank FCT, for financial support through grant number PTDC/FIS-OUT/28210/2017. PFNF and JNCE would like to acknowledge the contribution of the COST Action CA17139.

References

- 1 Dabrowski-Tumanski, P. *et al.* KnotProt 2.0: a database of proteins with knots and other entangled structures. *Nucleic Acids Research* **47**, D367-D375, doi:10.1093/nar/gky1140 (2018).
- 2 Bölinger, D. *et al.* A Stevedore's Protein Knot. *PLoS Computational Biology* **6**, e1000731, doi:10.1371/journal.pcbi.1000731 (2010).
- 3 Wallin, S., Zeldovich, K. B. & Shakhnovich, E. I. The Folding Mechanics of a Knotted Protein. *Journal of Molecular Biology* **368**, 884-893, doi:<https://doi.org/10.1016/j.jmb.2007.02.035> (2007).

- 4 Mallam, A. L. & Jackson, S. E. Folding Studies on a Knotted Protein. *Journal of Molecular Biology* **346**, 1409-1421, doi:<https://doi.org/10.1016/j.jmb.2004.12.055> (2005).
- 5 Andersson, F. I., Pina, D. G., Mallam, A. L., Blaser, G. & Jackson, S. E. Untangling the folding mechanism of the 52-knotted protein UCH-L3. *The FEBS Journal* **276**, 2625-2635, doi:doi:10.1111/j.1742-4658.2009.06990.x (2009).
- 6 Mallam, A. L., Morris, E. R. & Jackson, S. E. Exploring knotting mechanisms in protein folding. *Proceedings of the National Academy of Sciences* **105**, 18740-18745, doi:10.1073/pnas.0806697105 (2008).
- 7 Mallam, A. L. & Jackson, S. E. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol* **8**, 147-153, doi:<http://www.nature.com/nchembio/journal/v8/n2/abs/nchembio.742.html#supplementary-information> (2012).
- 8 Sułkowska, J. I., Sułkowski, P. & Onuchic, J. Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences* **106**, 3119 (2009).
- 9 Especial, J., Nunes, A., Rey, A. & Faísca, P. F. N. Hydrophobic confinement modulates thermal stability and assists knotting in the folding of tangled proteins. *Physical Chemistry Chemical Physics* **21**, 11764-11775, doi:10.1039/C9CP01701A (2019).
- 10 Najafi, S. & Potestio, R. Folding of small knotted proteins: Insights from a mean field coarse-grained model. *The Journal of Chemical Physics* **143**, 243121, doi:10.1063/1.4934541 (2015).
- 11 Chwastyk, M. & Cieplak, M. Cotranslational folding of deeply knotted proteins. *Journal of Physics: Condensed Matter* **27**, 354105, doi:10.1088/0953-8984/27/35/354105 (2015).
- 12 Noel, J. K., Onuchic, J. N. & Sułkowska, J. I. Knotting a Protein in Explicit Solvent. *The Journal of Physical Chemistry Letters* **4**, 3570-3573, doi:10.1021/jz401842f (2013).
- 13 Zhao, Y., Dabrowski-Tumanski, P., Niewieczeral, S. & Sułkowska, J. I. The exclusive effects of chaperonin on the behavior of proteins with 52 knot. *PLOS Computational Biology* **14**, e1005970, doi:10.1371/journal.pcbi.1005970 (2018).
- 14 Soler, M. A. & Faísca, P. F. N. Effects of Knots on Protein Folding Properties. *PLOS ONE* **8**, e74755, doi:10.1371/journal.pone.0074755 (2013).
- 15 King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L. & Yeates, T. O. Structure and folding of a designed knotted protein. *Proceedings of the National Academy of Sciences* **107**, 20732 (2010).
- 16 Sułkowska, J. I., Sułkowski, P., Szymczak, P. & Cieplak, M. Stabilizing effect of knots on proteins. *Proceedings of the National Academy of Sciences* **105**, 19714-19719, doi:10.1073/pnas.0805468105 (2008).
- 17 Mallam, A. L., Rogers, J. M. & Jackson, S. E. Experimental detection of knotted conformations in denatured proteins. *Proceedings of the National Academy of Sciences* **107**, 8189 (2010).
- 18 Liang-Wei, W., Yu-Nan, L., Ping-Chiang, L., Sophie, E. J. & Shang-Te Danny, H. Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of *Helicobacter pylori*. *Journal of Physics: Condensed Matter* **27**, 354106 (2015).
- 19 Sułkowska, J. I., Rawdon, E. J., Millett, K. C., Onuchic, J. N. & Stasiak, A. Conservation of complex knotting and slipknotting patterns in proteins. *Proceedings of the National Academy of Sciences* **109**, E1715-E1723, doi:10.1073/pnas.1205918109 (2012).
- 20 Soler, M. A., Nunes, A. & Faísca, P. F. N. Effects of knot type in the folding of topologically complex lattice proteins. *The Journal of Chemical Physics* **141**, 025101, doi:10.1063/1.4886401 (2014).

- 21 Dabrowski-Tumanski, P., Stasiak, A. & Sulkowska, J. I. In Search of Functional Advantages of Knots in Proteins. *PLOS ONE* **11**, e0165986, doi:10.1371/journal.pone.0165986 (2016).
- 22 Cardelli, C. *et al.* Heteropolymer Design and Folding of Arbitrary Topologies Reveals an Unexpected Role of Alphabet Size on the Knot Population. *Macromolecules* **51**, 8346-8356, doi:10.1021/acs.macromol.8b01359 (2018).
- 23 Norbiato, F., Seno, F., Trovato, A. & Baiesi, M. Folding Rate Optimization Promotes Frustrated Interactions in Entangled Protein Structures. **21**, 213 (2020).
- 24 Sułkowska, J. I., Noel, J. K. & Onuchic, J. N. Energy landscape of knotted protein folding. *Proceedings of the National Academy of Sciences* **109**, 17783-17788, doi:10.1073/pnas.1201804109 (2012).
- 25 Perego, C. & Potestio, R. Computational methods in the study of self-entangled proteins: a critical appraisal. *Journal of Physics: Condensed Matter* **31**, 443001, doi:10.1088/1361-648x/ab2f19 (2019).
- 26 Faísca, P. F. N. Knotted proteins: A tangled tale of Structural Biology. *Computational and Structural Biotechnology Journal* **13**, 459-468, doi:<https://doi.org/10.1016/j.csbj.2015.08.003> (2015).
- 27 Taketomi, H., Ueda, Y. & Gö, N. STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION. *International Journal of Peptide and Protein Research* **7**, 445-459, doi:10.1111/j.1399-3011.1975.tb02465.x (1975).
- 28 Škrbić, T., Micheletti, C. & Faccioli, P. The Role of Non-Native Interactions in the Folding of Knotted Proteins. *PLOS Computational Biology* **8**, e1002504, doi:10.1371/journal.pcbi.1002504 (2012).
- 29 a Beccara, S., Škrbić, T., Covino, R., Micheletti, C. & Faccioli, P. Folding Pathways of a Knotted Protein with a Realistic Atomistic Force Field. *PLOS Computational Biology* **9**, e1003002, doi:10.1371/journal.pcbi.1003002 (2013).
- 30 Lim, N. C. H. & Jackson, S. E. Mechanistic Insights into the Folding of Knotted Proteins In Vitro and In Vivo. *Journal of Molecular Biology* **427**, 248-258, doi:<https://doi.org/10.1016/j.jmb.2014.09.007> (2015).
- 31 Capraro, D. T., Roy, M., Onuchic, J. N. & Jennings, P. A. Backtracking on the folding landscape of the β -trefoil protein interleukin-1 β ? **105**, 14844-14848, doi:10.1073/pnas.0807812105 %J Proceedings of the National Academy of Sciences (2008).
- 32 Sułkowska, J. I., Sułkowski, P. & Onuchic, J. Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences* **106**, 3119-3124, doi:10.1073/pnas.0811147106 (2009).
- 33 Noel, J. K., Sułkowska, J. I. & Onuchic, J. N. Slipknotting upon native-like loop formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences* **107**, 15403-15408, doi:10.1073/pnas.1009522107 (2010).
- 34 Jackson, S. E., Suma, A. & Micheletti, C. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current Opinion in Structural Biology* **42**, 6-14, doi:<http://dx.doi.org/10.1016/j.sbi.2016.10.002> (2017).
- 35 Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. Specific Nucleus as the Transition State for Protein Folding: Evidence from the Lattice Model. *Biochemistry* **33**, 10026-10036, doi:10.1021/bi00199a029 (1994).
- 36 Krobath, H., Shakhnovich, E. I. & Faísca, P. F. N. Structural and energetic determinants of co-translational folding. *The Journal of Chemical Physics* **138**, 215101, doi:10.1063/1.4808044 (2013).

- 37 Faísca, P. F. N., Nunes, A., Travasso, R. D. M. & Shakhnovich, E. I. Non-native interactions play an effective role in protein folding dynamics. *Protein Science : A Publication of the Protein Society* **19**, 2196-2209, doi:10.1002/pro.498 (2010).
- 38 Shakhnovich, E. I. Proteins with selected sequences fold into unique native conformation. *Physical Review Letters* **72**, 3907-3910, doi:10.1103/PhysRevLett.72.3907 (1994).
- 39 Faísca, P. F. N. & Ball, R. C. Thermodynamic control and dynamical regimes in protein folding. *The Journal of Chemical Physics* **116**, 7231-7237, doi:10.1063/1.1466833 (2002).
- 40 Heinrich, K. & Patrícia, F. N. F. Interplay between native topology and non-native interactions in the folding of tethered proteins. *Physical Biology* **10**, 016002 (2013).
- 41 Shakhnovich, E. I. & Gutin, A. M. Engineering of stable and fast-folding sequences of model proteins. **90**, 7195-7199, doi:10.1073/pnas.90.15.7195 %J Proceedings of the National Academy of Sciences (1993).
- 42 Miyazawa, S. & Jernigan, R. L. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**, 534-552, doi:10.1021/ma00145a039 (1985).
- 43 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087-1092, doi:10.1063/1.1699114 (1953).
- 44 Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141-151, doi:[http://dx.doi.org/10.1016/S0009-2614\(99\)01123-9](http://dx.doi.org/10.1016/S0009-2614(99)01123-9) (1999).
- 45 Soler, M. A., Rey, A. & Faísca, P. F. N. Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding. *Physical Chemistry Chemical Physics* **18**, 26391-26403, doi:10.1039/C6CP05086G (2016).
- 46 Taylor, W. R. A deeply knotted protein structure and how it might fold. *Nature* **406**, 916-919 (2000).
- 47 Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C. & Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *Journal of Chemical Theory and Computation* **3**, 26-41, doi:10.1021/ct0502864 (2007).
- 48 Feig, M., Karanicolas, J. & Brooks, C. L. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling* **22**, 377-395, doi:<https://doi.org/10.1016/j.jmgn.2003.12.005> (2004).
- 49 Faísca, P. F. N., Travasso, R. M. D., Charters, T., Nunes, A. & M., C. The folding of knotted proteins: insights from lattice simulations. *Physical Biology* **7**, 016009 (2010).
- 50 I., A. F., G., P. D., L., M. A., Georg, B. & E., J. S. Untangling the folding mechanism of the 52-knotted protein UCH-L3. *The FEBS Journal* **276**, 2625-2635, doi:doi:10.1111/j.1742-4658.2009.06990.x (2009).
- 51 Ziegler, F. *et al.* Knotting and unknotting of a protein in single molecule experiments. *Proceedings of the National Academy of Sciences* **113**, 7533-7538, doi:10.1073/pnas.1600614113 (2016).
- 52 Wang, I., Chen, S.-Y. & Hsu, S.-T. D. Folding analysis of the most complex Stevedore's protein knot. *Scientific Reports* **6**, 31514, doi:10.1038/srep31514 (2016).
- 53 Bustamante, A. *et al.* The energy cost of polypeptide knot formation and its folding consequences. *Nature Communications* **8**, 1581, doi:10.1038/s41467-017-01691-1 (2017).

- 54 Rivera, M., Hao, Y., Maillard, R. A. & Baez, M. Mechanical unfolding of a knotted protein unveils the kinetic and thermodynamic consequences of threading a polypeptide chain. *Scientific Reports* **10**, 9562, doi:10.1038/s41598-020-66258-5 (2020).
- 55 Soler, M. A. & Faísca, P. F. N. How Difficult Is It to Fold a Knotted Protein? In Silico Insights from Surface-Tethered Folding Experiments. *PLOS ONE* **7**, e52343, doi:10.1371/journal.pone.0052343 (2012).
- 56 Noel, J. K., Sułkowska, J. I. & Onuchic, J. N. Slipknotting upon native-like loop formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences* **107**, 15403 (2010).
- 57 Gosavi, S., Chavez, L. L., Jennings, P. A. & Onuchic, J. N. Topological Frustration and the Folding of Interleukin-1 β . *Journal of Molecular Biology* **357**, 986-996, doi:<https://doi.org/10.1016/j.jmb.2005.11.074> (2006).
- 58 King, N. P., Yeates, E. O. & Yeates, T. O. Identification of Rare Slipknots in Proteins and Their Implications for Stability and Folding. *Journal of Molecular Biology* **373**, 153-166, doi:<https://doi.org/10.1016/j.jmb.2007.07.042> (2007).
- 59 Covino, R., Škrbić, T., a Beccara, S., Faccioli, P. & Micheletti, C. The Role of Non-Native Interactions in the Folding of Knotted Proteins: Insights from Molecular Dynamics Simulations. *Biomolecules* **4**, 1-19, doi:10.3390/biom4010001 (2014).