Supplementary material for "Energy landscape of knotted protein folding"

Contents

1 Topology and geometry of proteins

The knotted protein 2ouf was created by Todd Yeates group by mirroring the evolutional pathway by which several naturally knotted proteins can apparently become knotted, so called "domain duplication" [1]. This knotted protein is a fused, dimeric protein from Helicobacter pylori, HP0242 (PDB entires 2ouf/2bo3) by genetically linking the two subunits of the the dimer. The two monomers were fused using a flexible 9-residue linker (SGSGSGSGSSG) to construct the 2ouf knotted structure. Details of this work can be found in [1].

Table 1: Elements of secondary structures of the knotted protein 2ouf and unknotted protein 2ouf-ds.

Elements	Amino acids	Elements	Amino acids
1a	1-14	1a	1-14
2a	14-39	2a	14-39
3a	39-56	3a	39-56
4a	56-80	4a	56-80
linker	81-89	–	
1b	90-104	1b	90-104
2b	104-129	2b	104-129
3b	129-144	3b	129-144
4b	144-169	4b	144-169

2 Results and discussion

2.1 Folding mechanism of the designed protein 2ouf

Additional explanation about threading mechanism is shown in Fig. 2. In Fig. 2B, left vertical grain represents the ensemble of routes with very narrow range of Q, 4.9< Q <5.1 and over broad range of rmsd, 1.3-19 nm. These routes correspond to knotting by plugging mechanism [2], with linker breathing freely. Almost horizontal part of the grain represents routes with very similar rmsd, 1.1<rmsd<1.3 over a broader range of native contacts, 0.4< Q <4.9. The knotting event is the rate limiting step, and is located on the top of the barrier.

$K(Q, \text{rmsd})$ has bent ")" shape indicating optimal number of Q and rmsd to tie a knot, similar to the pass on the grain. Vertical grain represents the ensemble of routes with very narrow range of Q, 4.9< Q <5.1 over broad range of rmsd, 1.3-19nm. These routes correspond to knotting by

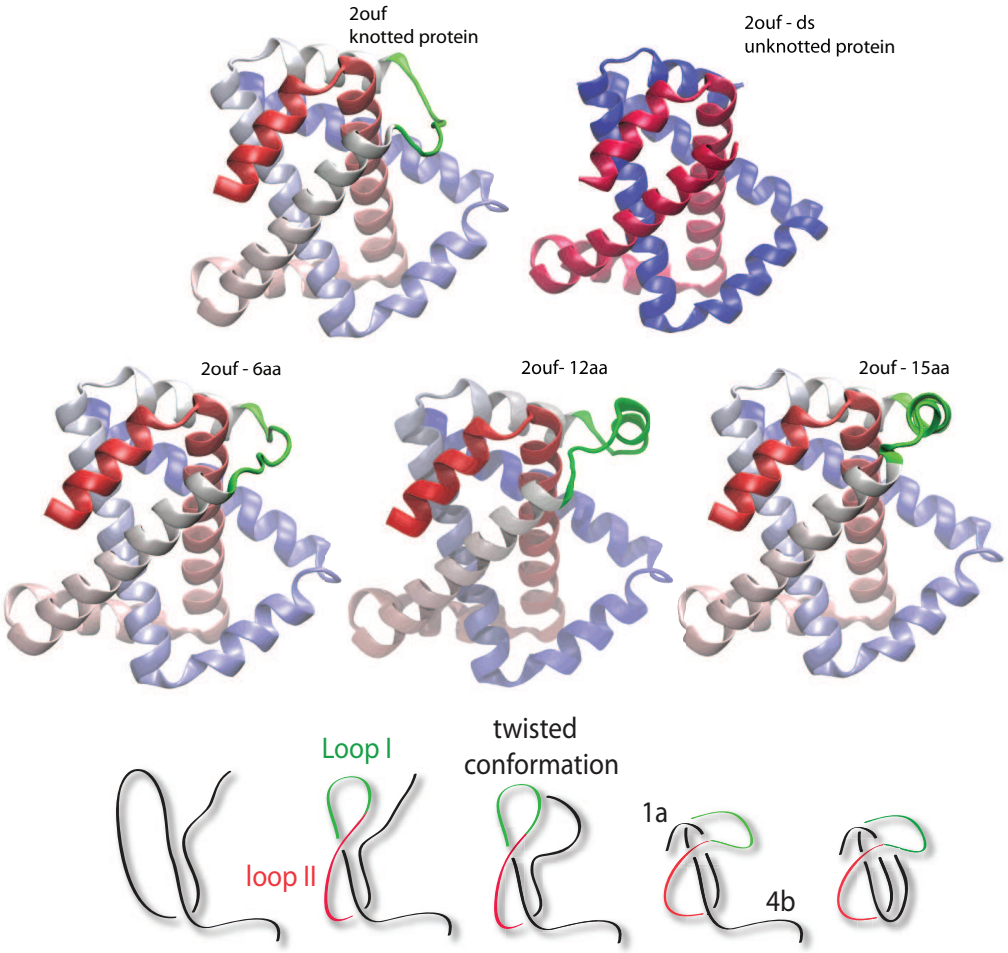

Figure 1: Cartoon representation of knotted (2ouf) and unknotted protein (2ouf-ds), first row. Second row: cartoon representation of knotted protein with 6aa, 12aa and 15aa. Bottom row: folding route of knotted protein. Red line shows loop II and green line shows loop I.

plugging mechanism [2], with linker breathing freely. Almost horizontal grain represents routes with very similar rmsd, 1.1<rmsd<1.3 over broader range of native contacts, 0.4< Q <4.9. The knotting mechanism is rate limiting step located on the top of the barrier.

2.2 Removing the native bias from the linker

The $K(Q, \text{rmsd}) = 0.5$ contour resembles a mirror image of "L". This shape of the contour hints at two different regimes of knot threading configurations separated by the elbow. One ensemble, broad in rmsd but at a well-defined and maximal Q, corresponds to knotting by the plugging mechanism [2], with the linker fluctuating freely. The constant Q value means that the native structural content needed to facilitate plugging is well determined. The other ensemble has a broad range below the maximal Q and well-defined rmsd values. These routes thread by slipknotting or loop flipping mechanisms. The polymer-like linker of 2ouf-free increases the general knotting propensity. $N_{transition}/N_k$ is smaller than for native 2ouf, a larger number of knotting events is observed. The knot topologies show the same characteristics as 2ouf, where complex knots like 5_2 where not found. This analysis suggests that free linker reduces frustration in the horizontal route but makes vertical route even more deterministic. This is one of the factors which is responsible for slower folding kinetics of 2ouf-free compared to 2ouf.

2.3 Varying the stiffness of the linker

Detailed analysis of the folding trajectories shows that 2ouf-stiff-N has a nearly identical folding mechanism as 2ouf (Tables S2 and S4). The only change is observed during threading the N-terminus across Loop I, 2a-1b forms ahead of 4a-2b, opposite to the situation in 2ouf. 2ouf-soft-N shows significant differences from 2ouf at late stage of folding. Both proteins first form twisted loop (2a2b, 3a2b, 2a3b), but then instead of formation of contacts between 2a-1b, contacts between 4a-2b, 2a-4a are formed. Then again contacts between 2a-4b are formed ahead of 1a-4a. When the packing of the N-terminal helix is not constrained there are more routes to cross the topological barrier, as implied by the lower route measure at the range of Q corresponding to forming contacts between 1a and rest of the structure (see Table S4).

2.4 Varying the helicity of the N-terminal helix

The conformation of the N-terminal helix is similar in both slipknotting and loop flipping. For the slipknot configuration, the N-terminal helix must bend back on itself in a hairpin-like configuration to thread Loop I. If it threads by loop flipping or plugging, the N-terminal helix can be native-like. Because of this conformational difference, the local bias towards a helical conformation for the N-terminal helix could change the folding of the knot. We studied two mutants, 2ouf-soft-N and 2ouf-stiff-N, which had their native dihedral bias (ϵ_{D} in [3]) reduced or strengthened by a factor of 2 relative to 2ouf. 2ouf-stiff-N has nearly identical folding behavior to 2ouf, but 2ouf-soft-N shows differences in late stages of folding. Both proteins first form twisted loop (2a2b, 3a2b, 2a3b), but then instead of formation of contacts between 2a-1b, contacts between 4a-2b, 2a-4a are formed. Then again contacts between 2a-4b are formed ahead of 1a-4a. When the packing of the N-terminal helix is not constrained there are more routes to cross the topological barrier, as implied by the lower route measure at the range of Q corresponding to forming contacts between 1a and rest of the structure (see Table S4).

As is typical in SBMs, the softer dihedrals shift the native basin towards lower Q and lower the barrier [5]. Free energy from the perspective of rmsd indicates different routes over the transition state, (Fig. S2). The route measure [4] indicates that the transition state ($0.42 < Q < 0.57$) is more routed/polarized for 2ouf-soft-N (Fig S3). But for late transition state events ($0.58 < Q < 0.68$) the situation is reversed, 2ouf-soft-N is less routed. This range corresponds to the appearance of the metastable state at high Q for 2ouf and 2ouf-stiff-N. Contour of knot formation shows a similar shape as 2ouf. It is interesting to notice that even though $F(Q, \mathrm{rmsd})$ is very similar to 2ouf (Fig. 2A and 2B), there is a pronounced decrease in total knot formation for both mutants. N_{trans}/N_{knot} for 2ouf-stiff-N is double that of 2ouf.

3 Folding routes via order of formation of tertiary structure

Table 2: Formation of secondary structures at transition state, based on thermodynamics data.

	Native knot	Linker with soft contacts	Free linker	Soft linker	Stiff linker	Stiffer linker	Soft N-terminal	Sitff N-terminal
Q_{tr}	0.46	0.5	0.53	0.47	0.46	0.44	0.44	0.467
	2a	2a	2a	2a	2a	1b¿2a	2a	2a
	2b	2b	2b	2b	1b !	2a	2b=3a	2b
	3a	3a	3a	3a	2b	1a	3a	3a
	1a	1a=4a	1a	1a=4a	1a=3a	4a	1b=4a	1a
	1b	1a=4a	4a	4a	3a	2b	4a	1b
	4a	3b=1b	3b	1b	4a	3a	3b	4a
	3b	1b	1b	3b	3b	3b	4a	3b
	4b	4b	4b	4b	4b	4b	1a !!!	4b

Table 3: Formation of tertiary structures during folding 2ou, 2ouf-soft-contacts-linker and 2ouf-free-linker.

description	2ouf	2ouf - weak contacts	2ouf-free linker
hydrophobic core	2a2b	2a2b	2a2b
(formation of	3a2b	3a2b	3a2b
twisted loop)	2a3b	2a3b	2a3b
knotting	2a1b=4a2b	4a2b	4a2b
	4a2b	2a4a	2a4a !
	2a4a	2a1b	2a1b
knotting	1a2b	1a2b	1a2b
N-terminal	1a4a	2a4b	2a4b !
final packing	2a4b	1a4a	1a4a !
	2b4b=1b4b	2b4b	2b4b
	1b4b	1b4b	1b4b

Table 4: Formation of tertiary structures during folding 2ouf, 2ouf-soft-N (protein with flexible N terminus,1a) and 2ouf-stiff-N (protein with stiff N terminus, 1a).

	2ouf	Soft N -terminal	Stiff N-terminal
hydrophobic core (formation of twisted loop)	2a2b 3a2b 2a3b	2a2b 3a2b 2a3b	2a2b 3a2b 2a3b
	2a1b=4a2b 4a2b 2a4a	4a2b 2a4a 2a1b !	2a1b ! 4a2b 2a4a
	1a2b 1a4a	1a2b 2a4b !	1a2b 1a4a
final packing	2a4b 2b4b=1b4b 1b4b	1a4a 2b4b 1b4b	2a4b 2b4b=1b4b 1b4b

Table 5: Formation of tertiary structures during folding of: 2ouf, 2ouf-soft linker, 2ouf-stiff-linker and 2ouf-stiffer linker.

	2ouf	Soft linker	Stiff linker	Stiffer linker
hydrophobic core (formation of twisted loop)	2a2b 3a2b 2a3b	2a2b 3a2b 2a3b	2a2b 2a1b ! 3a2b	2a1b ! 2a4a ! 2a2b
	2a1b=4a2b 4a2b 2a4a	4a2b 2a4a 2a1b	2a4a ! 2a3b 4a2b !	3a2b=1a2b 1a2b 4a2b = 2a1b
	1a2b 1a4a	1a2b 1a4a	1a2b 1a4a	2a1b 2a3b !
final packing	2a4b 2b4b=1b4b 1b4b	2a4b 2b4b 1b4b	2a4b 2b4b 1b4b	2a4b 1b4b 2b4b

4 Free energy landscape as $F(Q)$ and $F(Q, \mathrm{rmsd})$

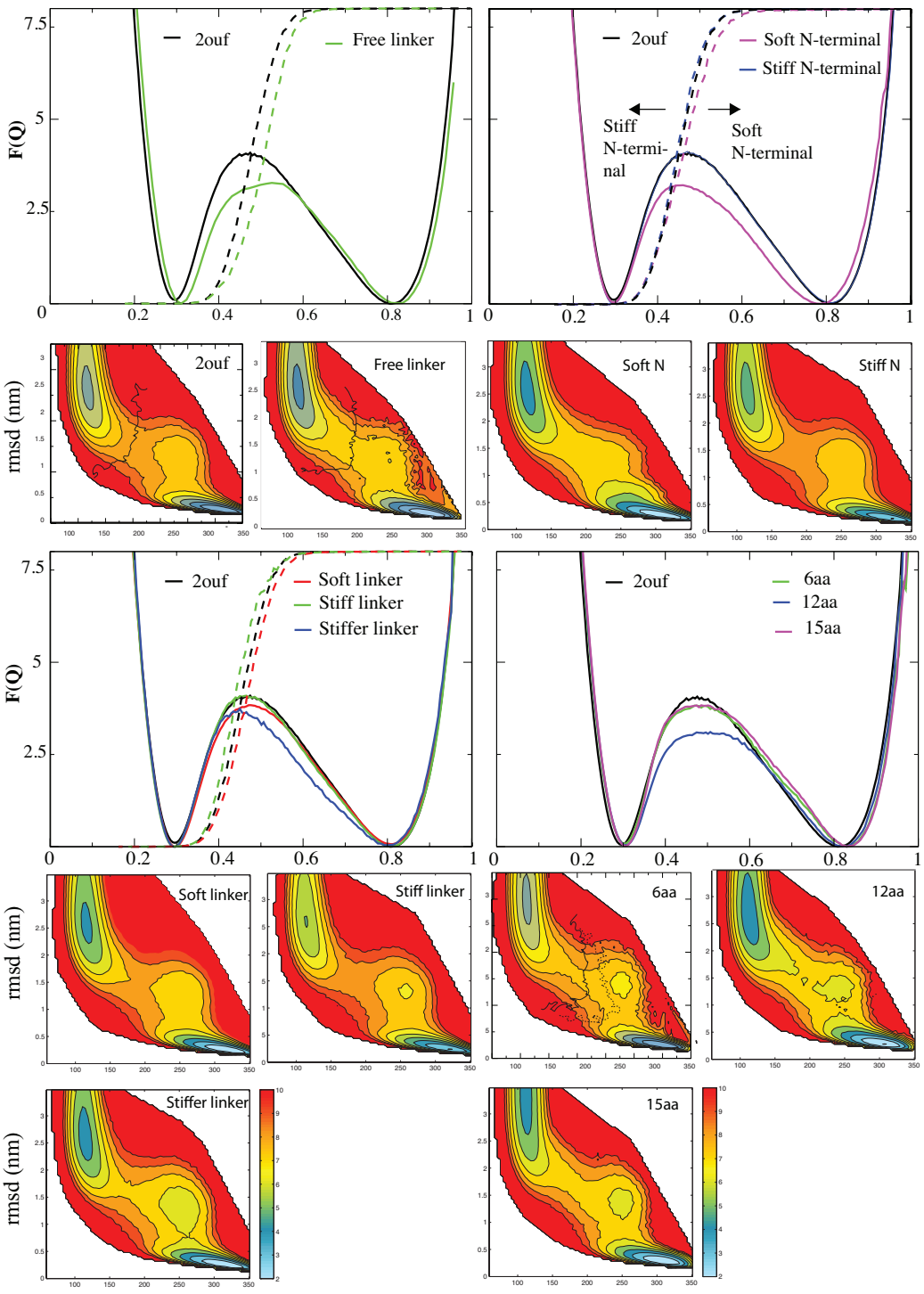

Figure 2: Free energy landscapes $F(Q)$, $F(Q, \text{rmsd})$ of all mutants of 2ouf.

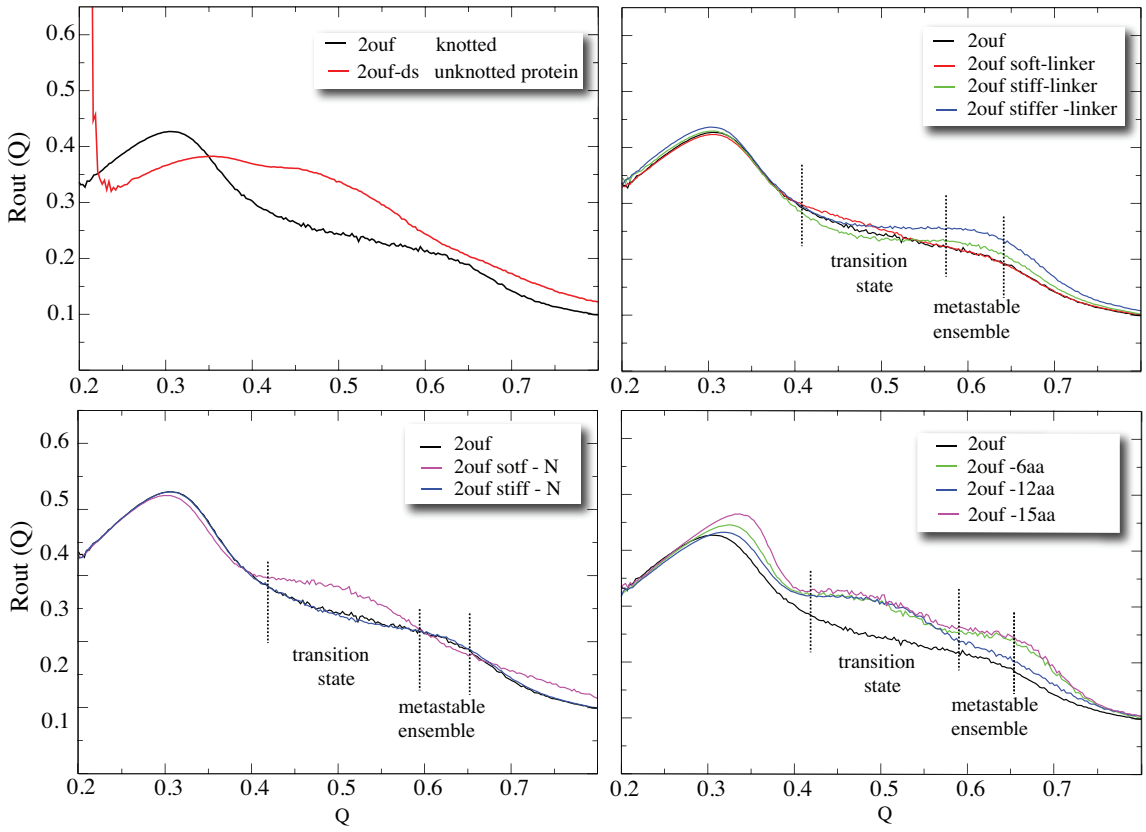

Figure 3: A comparison of the route measure $R(Q)$. The first peak at each $R(Q)$ plot corresponds to formation of twisted loop, where knot is formed at the transition state around $Q = 0.5$.

5 Route measure

$R(Q)$ is normalized between 0 and 1 and is defined by

$$R(Q) = \sum_{i=1}^{M} \frac{(\langle Q_i \rangle_Q - Q)^2}{MQ(1-Q)} \tag{1}$$

where M is the number of native contacts and $\langle Q_i \rangle_Q$ is the average formation of the i'th contact in all configurations with a particular global Q. $R(Q)$ quantifies the diversity of structures seen at each value of Q: $R(Q) = 0$ being maximum diversity and R(Q) = 1 being a single route. At R(Q) = 0, all $\langle Q_i \rangle_Q = Q$, meaning all possible configurations of native contacts are sampled equally. At R(Q) = 1 only a subset of QM contacts are formed with $\langle Q_i \rangle_Q = 1$, meaning only one configuration of native contacts is sampled.

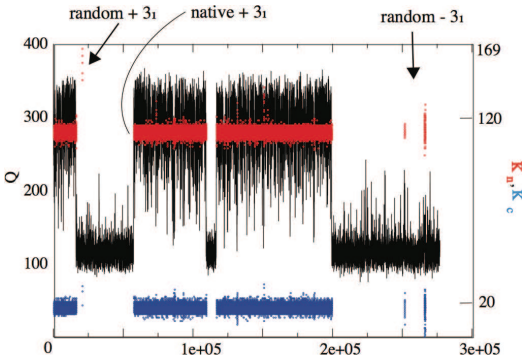

Figure 4: Knotting, folding and unknotting, unfolding events observed for 2ouf. Here we observed random knotting by the C-terminal (random $+3_1$ knot), folding via pre-order knotted loop, and random knotting with wrong chirality (random -3_1). Topological signature of the protein measured by the position of the knot along sequence, knot termini are shown by blue and red dots.

6 Folding/unfolding and knotting/unknotting events.

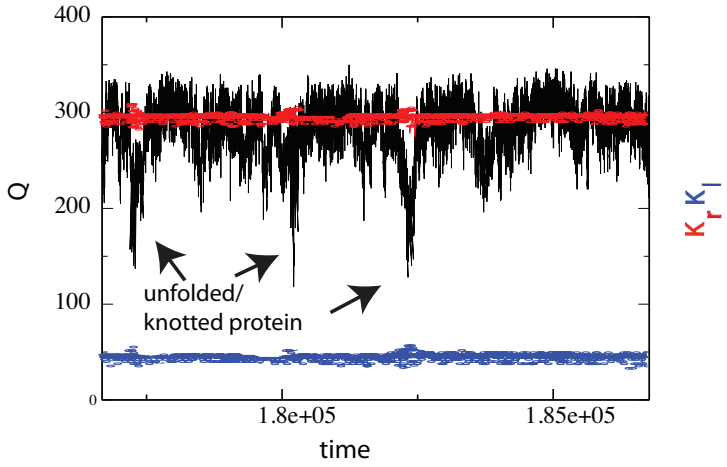

Figure 5: Unknotting of secondary structures while preserving the knot topology observed for 2ouf-6aa. Unfolded/ knotted protein shows very fast folding from this knotted state. Topological signature of the protein measured by the position of the knot along sequence, knot termini are shown by blue and red dots.

7 Comparison of contact maps for knotted (2ouf), and unknotted protein (2ouf-ds).

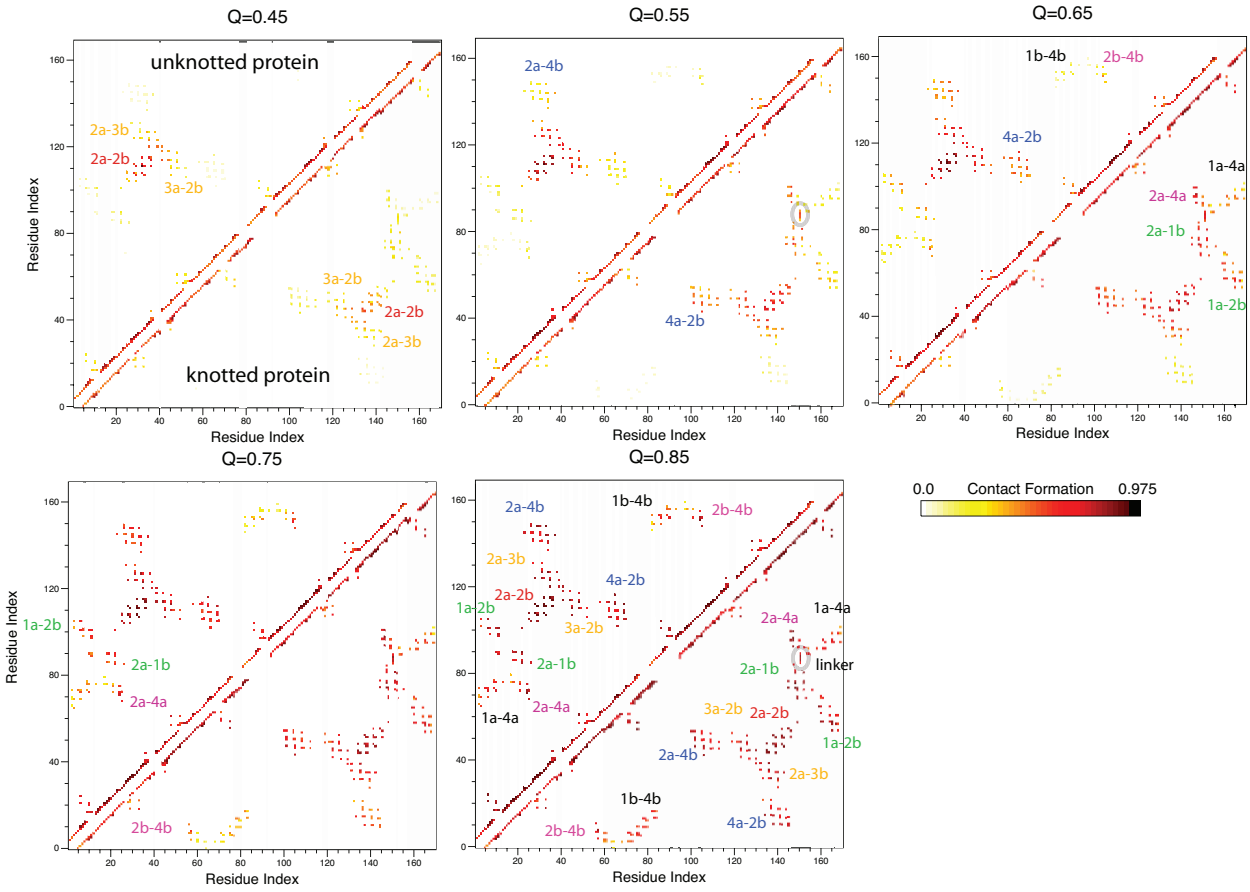

Figure 6: Comparison of contact maps for knotted, 2ouf and unknotted protein, 2ouf-ds at different Q based on thermodynamics data.

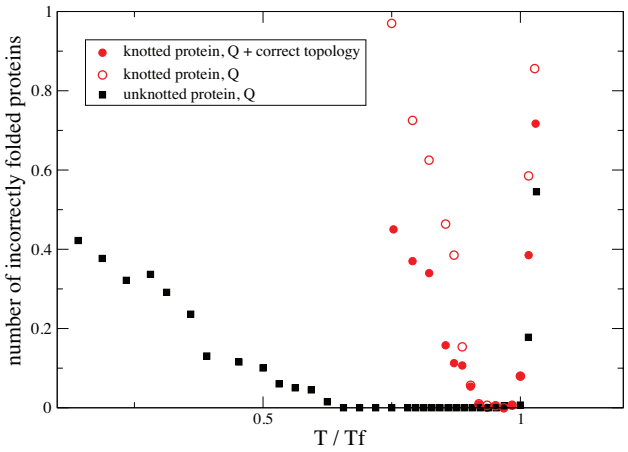

Figure 7: Number of incorrectly folded knotted (red circles) and unknotted (black squares) proteins versus temperature. Red solid circles correspond to number of folded proteins with correct topology and high Q. Red open circles correspond to number of folded proteins only for high Q value.

8 Kinetics and comparison to experimental results

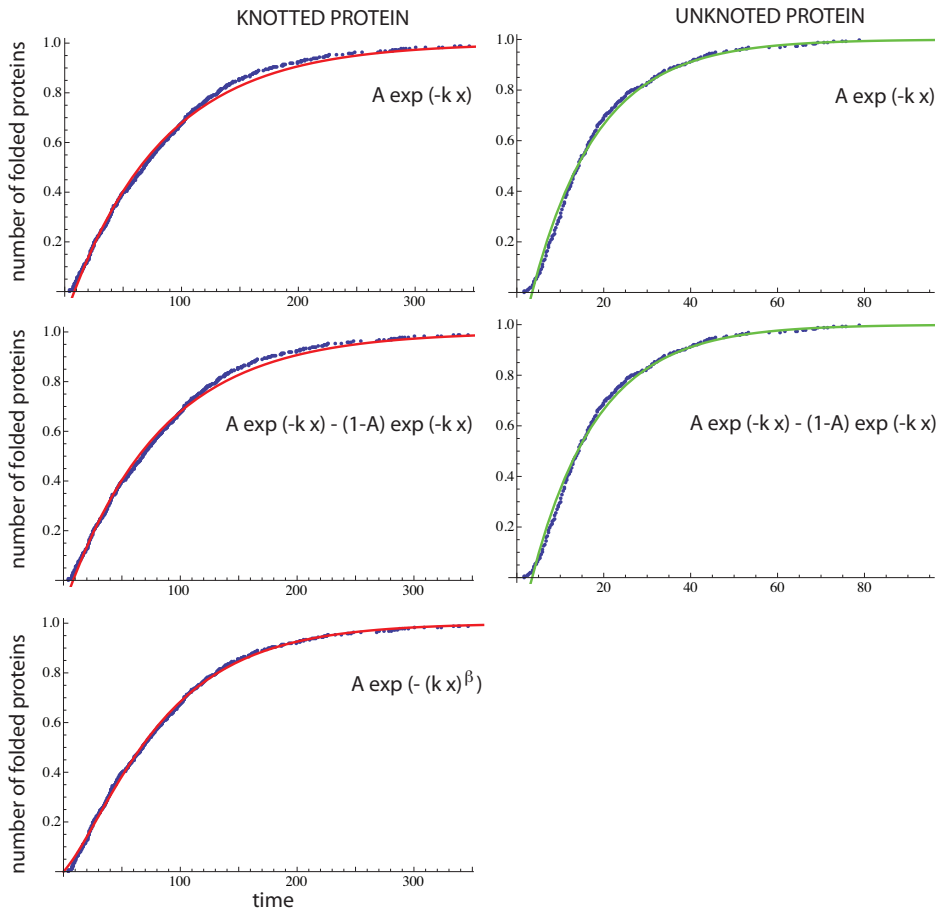

Figure 8: Distribution of number of folded proteins, knotted - left column and unknotted protein - right column. Distribution of number of folded proteins versus time are fitted to single, double and stretched exponent.

9 Knots detection

In order to define the knotted core (*i.e.* the minimal segment of amino acids that can be identified as a knot) we use the so-called Koniaris-Mutukhumar-Taylor algorithm [6, 7]. The structure is not knotted if this procedure can reduce the protein to just two termini, otherwise a knot must be present in the protein structure. Alternatively, cutting off amino acids from both sides of a knot until the knot ceases to be detected, allows one to determine the amino acids k_1 and k_2 spanning the knotted core.

$K(Q)$ (and similarly $K(Q, \mathrm{rmsd})$) is an ensemble average of the binary variable K over all structures at Q, where $K = 1$ if the KMT algorithm determines a structure has any knot and $K = 0$ otherwise.

References

[1] N. P. King, A.W. Jacobitz, M.R. Sawaya, L. Goldschmidt and T.O. Yeates (2010) Structure and folding of a designed knotted protein. *Proc Natl Acad Sci USA* 107: 20732–7.

[2] J.K. Noel, J.I Sulkowska, J.N. Onuchic (2010) Slipknotting upon native-like loop formation in a trefoil knot protein. *PNAS* 31: 15403-8.

[3] J.K. Noel, P.C. Whitford, K.Y. Sanbonmatsu, J.N. Onuchic (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38: W657-661

[4] S. Gosavi, L.L. Chavez, P.A. Jennings, J.N. Onuchic (2006) (Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.* 357: 986–996.

[5] P.C. Whitford, J.K. Noel, S. Gosavi, A. Schug, K.Y. Sanbonmatsu, J.N Onuchic (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430–441.

[6] K. Koniaris, M. Muthukumar (1991) Knottedness in ring polymers *Phys. Rev. Lett.* 66: 2211-2214.

[7] W.R. Taylor (2000) A deeply knotted protein structure and how it might fold. *Nature* 406: 916-919.