Supplementary Information for

Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding

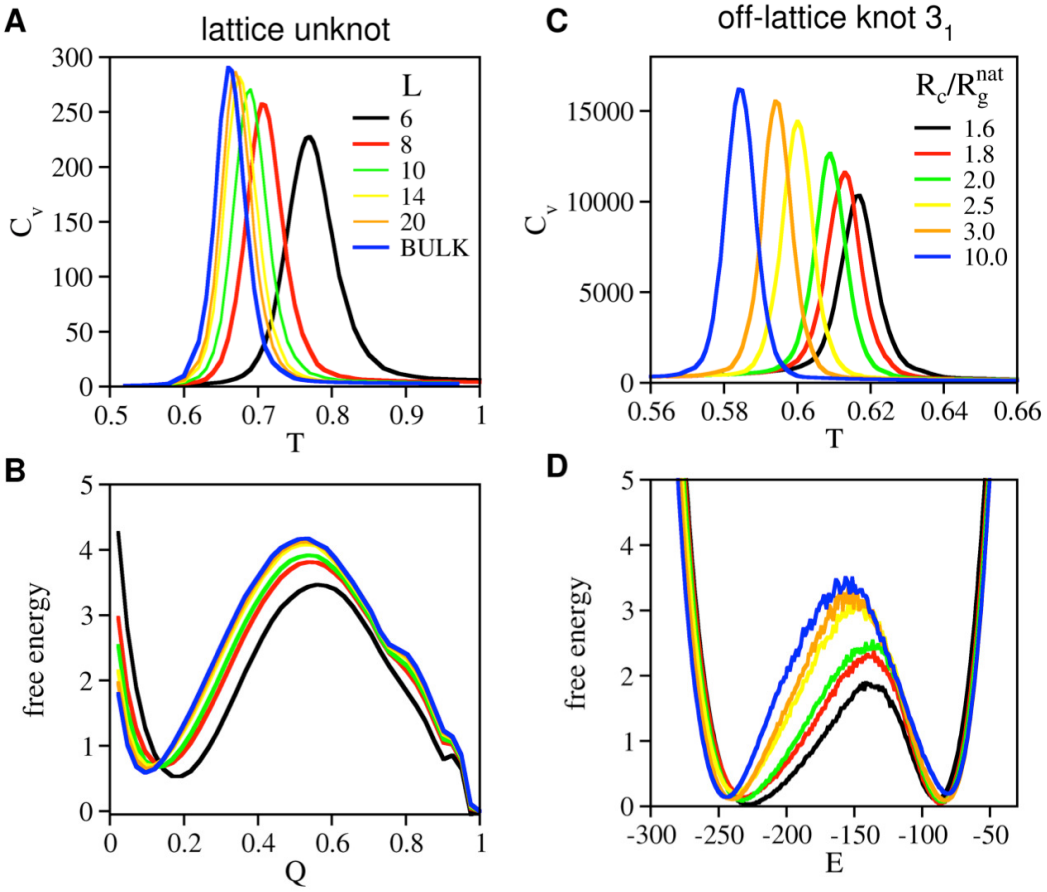
Miguel A Soler[1], Antonio Rey[2]* & Patrícia FN Faísca[3]*
[1]Dipartimento di Scienze Mediche e Biologiche, Universita' di Udine, Piazzale Kolbe 4, 33100 Udine, Italy
[2]Departamento de Química Física I, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain
[3]Departamento de Física and BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, 1749-016 Lisboa, Portugal

Corresponding authors:

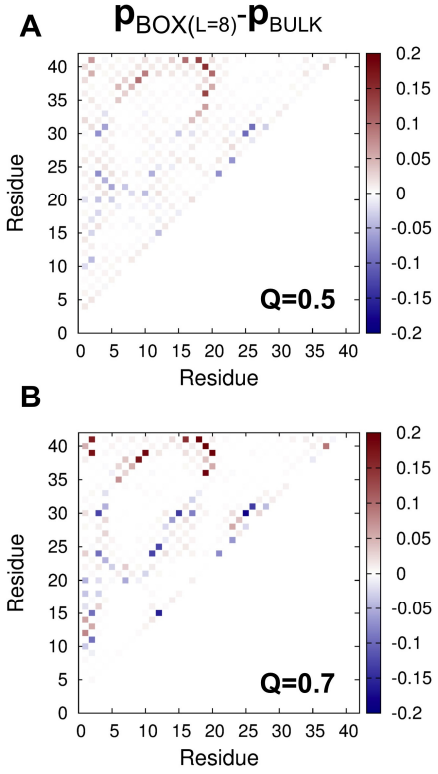
Patrícia F.N. Faísca, Email: pffaisca@fc.ul.pt
Antonio Rey, Email: areygayo@ucm.es

This file includes:

Supplementary Figures SI: Fig. 1-SI: Fig. 6
Supplementary Table 1



SI: Figure 1. Thermodynamics of the folding transition. The dependence of the heat capacity on temperature for the unknotted lattice system under confinement (A) and free energy profiles evaluated at T_m (B). Dependence of the heat capacity on temperature for protein MJ0366 (C) under confinement and free energy profiles evaluated at T_m (D). L is the linear size of the cubic box and R_c/R_g^{nat} is the ratio between the radius of the confinement sphere and the radius of gyration of the native structure of protein MJ0366.

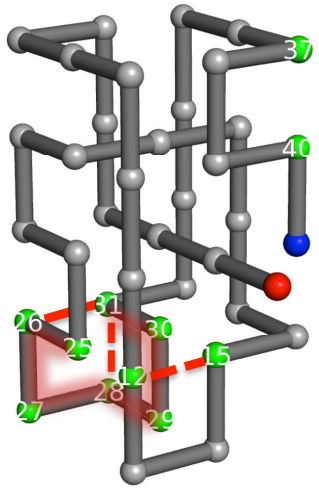


Contact	$p_{BOX(L=8)}$-p_{BULK}
21-24	-0.072
23-26	-0.069
25-30	-0.097
26-31	-0.096
28-31	-0.030
3-24	-0.073
3-30	-0.071
25-30	-0.181
12-15	-0.159
15-30	-0.135
26-31	-0.134
3-30	-0.127
11-24	-0.129
12-25	-0.127

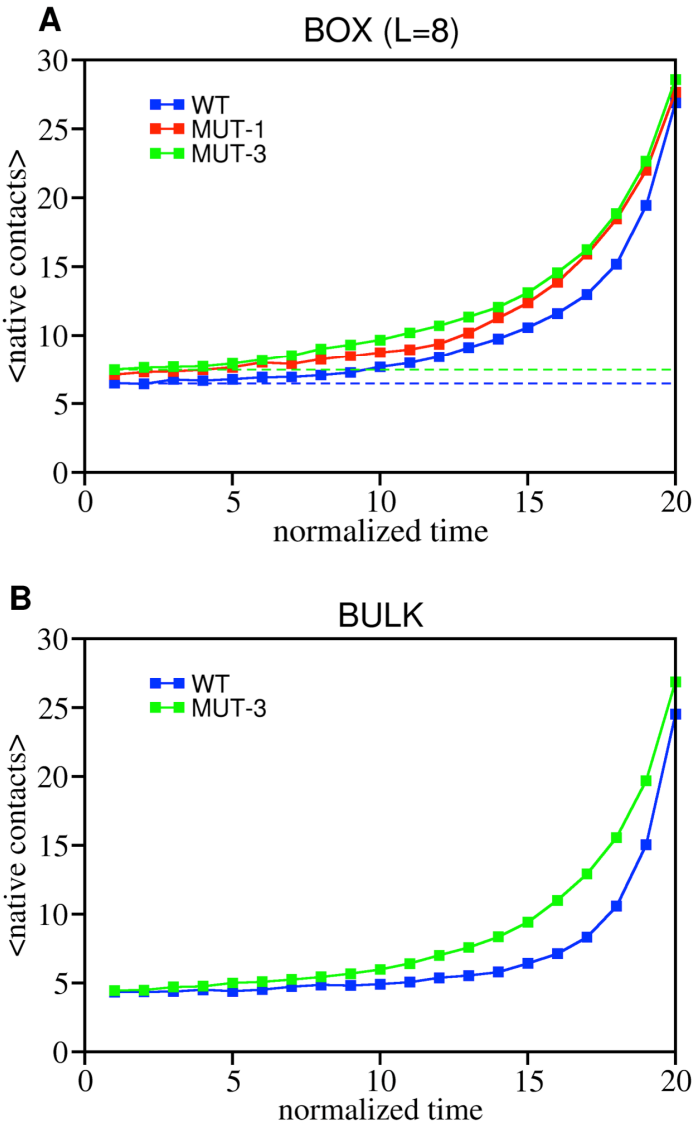
SI: Figure 2. **Contact statistics under confinement**. Difference probability maps evaluated for ensembles of conformations representative of (A) the transition state (Q = *0.5*) and (B) post-transition state (Q = *0.7*) for the lattice knotted system. Reported data show the difference between the probability of contact formation when the lattice protein folds under confinement and in the bulk. The contacts whose formation becomes less likely under confinement (at both folding stages) are identified on the right, with local contacts being highlighted in bold.

Interaction(s) switched-off	Folding rate (at T_m)
WT	**7.78e-08**
WT (BULK)	**4.82e-08**
21-24	6.47e-08
23-26	7.00e-08
25-30	9.14e-08
26-31	8.46e-08
28-31	8.41e-08
25-30	9.14e-08
12-15 (MUT-1)	**10.82e-08**
MUT-1 (BULK)	**6.70e-08**
37-40	8.21e-08
28-31•26-31	9.27e-08
25-30•24-29	11.65e-08
12-15•37-40	12.54e-08
12-25•15-30	13.74e-08
12-15• 37-40•26-31	14.78e-08
12-15• 28-31•26-31 (MUT-3)	**16.13e-08**
MUT-3 (BULK)	**7.50e-08**

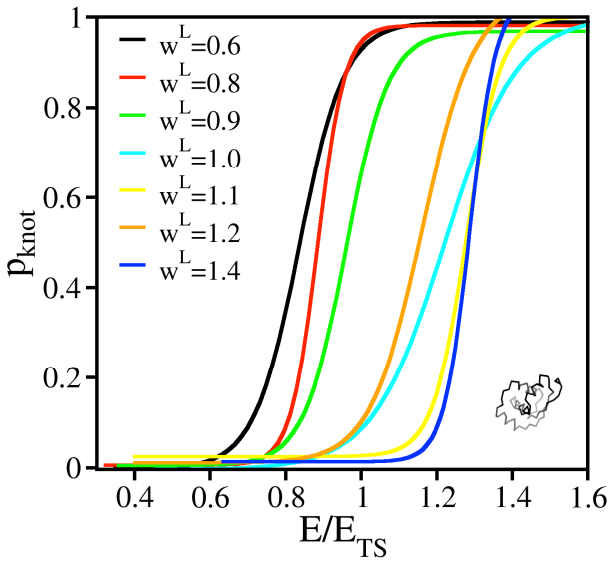
SI: Table 1. Local structure and folding speed. Effect of selected single, double and triple point mutations on the folding rate of the lattice system evaluated at T_m. The contact between residues 37 and 40 was used as a control because it is a local contact whose frequency is not affected by confinement.



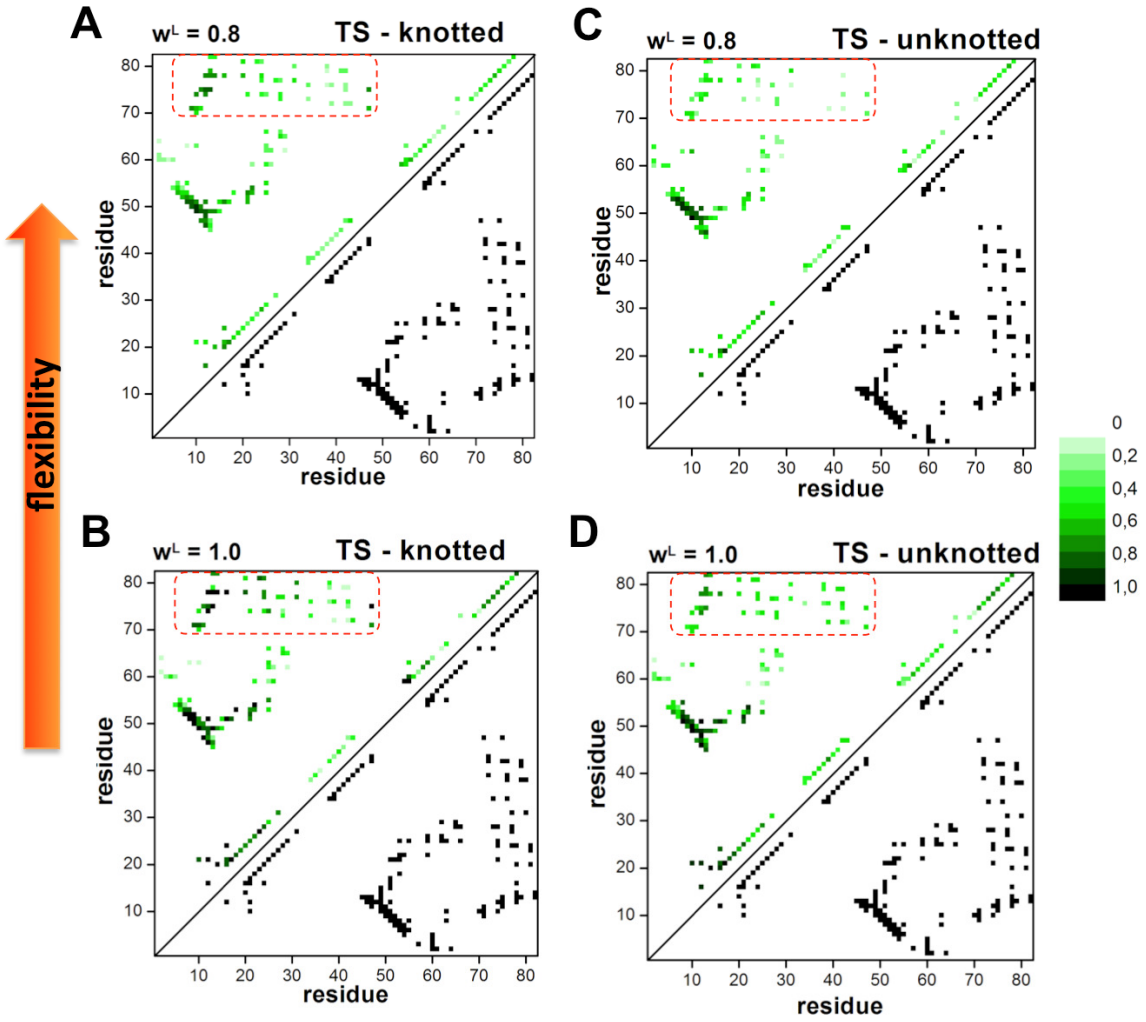
SI: Figure 3. Mapping of selected interactions on the native structure. The triple mutation MUT-3 perturbs three interactions that define a region of local structure (highlighted in red) whose formation delays folding.



SI: Figure 4. **Effect of local interactions on backtracking**. Dependence of the average number of formed native contacts on the simulation time under confinement (A) and in the bulk (B). To compute these plots we considered 2000 Monte Carlo folding runs and divided the last 1% of each MC run in 20 intervals. The number of MC steps defining each interval was normalized to the run's folding time. The average number of native contacts was evaluated for each interval by taking into account data from the 2000 MC runs. We observe that in the case of the wild-type (WT) system, the average number of native contacts keeps constant for a longer time, meaning that the protein is forming and breaking contacts until very late in folding. When the local interactions are disrupted, the protein is able to grow a larger number of native contacts early on during folding, indicating less backtracking for the mutants with higher flexibility.



SI: Figure 5. **Effect of flexibility on the knotting probability**. The knotting probability, p_{knot}, as a function of energy E (normalized to the energy of the transition state, E_{TS}) for protein MJ0366. As before, p_{knot} was evaluated for ensembles of conformations collected from replica-exchange simulations (replica at T_m). The curves represent sigmoidal fitting of the raw data.



SI: Figure 6. **Flexibility and knotting mechanism**. Probability maps of the knotted (A and B) and unknotted (C and D) conformations representing the transition state for the balanced model and for the model with enhanced flexibility. The dotted rectangle indicates the contacts involving the C-terminal helix and the knotting loop.