Folding Mechanisms of Entangled Proteins

Numerical studies within coarse-grained structure-based models

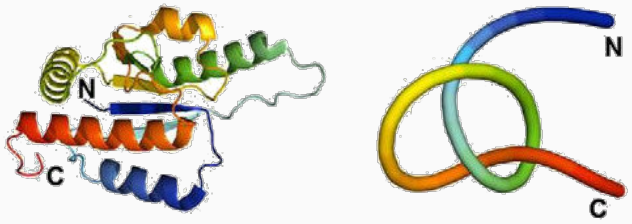
---

Leonardo Salicari | leonardo.salicari@phd.unipd.it

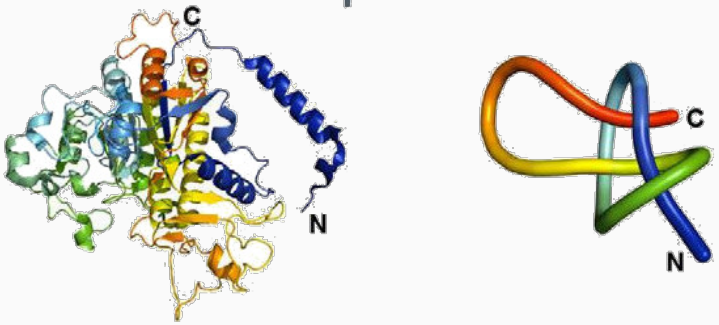
University of Padua, Department of Physics and Astronomy

Entanglement in proteins: the example of knotted proteins

3_1 Knot



Slipknot

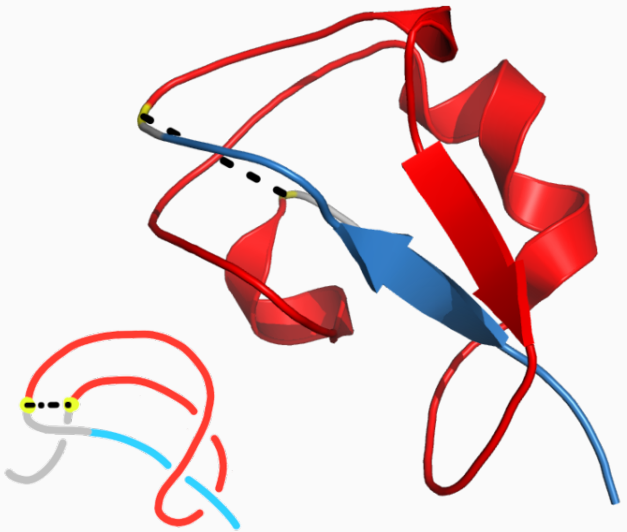


Lim, N. *et al.*; J. Phys.: Condens. Matter (2015)

A simple topological observable: Gaussian Entanglement

$$G'(i,j) = \frac{1}{4\pi} \sum_{i=i_1}^{i_2-1} \sum_{j=j_1}^{j_2-1} \frac{\boldsymbol{R}_i - \boldsymbol{R}_j}{\left|\boldsymbol{R}_i - \boldsymbol{R}_j\right|^3} \cdot \left(\Delta \boldsymbol{R}_i \times \Delta \boldsymbol{R}_j\right)$$

Type III Antifreeze protein RD1



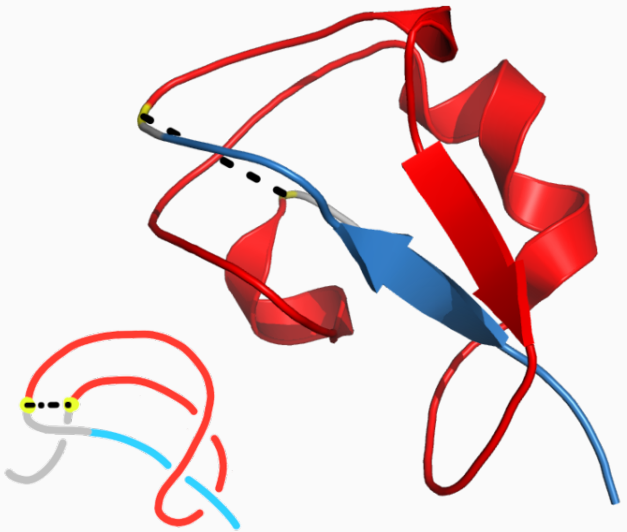
32% of proteins domains contains at least one *entangled loop* motif

Baiesi, M. *et al.*; A. Sci. Rep. (2017)

A simple topological observable: Gaussian Entanglement

$$G'(i,j) = \frac{1}{4\pi} \sum_{i=i_1}^{i_2-1} \sum_{j=j_1}^{j_2-1} \frac{\boldsymbol{R}_i - \boldsymbol{R}_j}{\left|\boldsymbol{R}_i - \boldsymbol{R}_j\right|^3} \cdot \left(\Delta \boldsymbol{R}_i \times \Delta \boldsymbol{R}_j\right)$$

Type III Antifreeze protein RD1



32% of proteins domains contains at least one *entangled loop* motif

Baiesi, M. *et al.*; A. Sci. Rep. (2017)

Statistics of entangled loops

- Entangled loops enriched in the C-terminal side of the entanglement they are in
- high $|G'|$ contacts tend to be weakly bounded

Hypothesis

Contacts with high $|G'|$ tend to form in the later stages of the folding

Baiesi, M. *et al.*; A. Sci. Rep. (2019)

Statistics of entangled loops

- Entangled loops enriched in the C-terminal side of the entanglement they are in
- high $|G'|$ contacts tend to be weakly bounded

Hypothesis

Contacts with high $|G'|$ tend to form in the later stages of the folding

Baiesi, M. *et al.*; A. Sci. Rep. (2019)

Statistics of entangled loops

- Entangled loops enriched in the C-terminal side of the entanglement they are in
- high $|G'|$ contacts tend to be weakly bounded

Hypothesis

Contacts with high $|G'|$ tend to form in the later stages of the folding

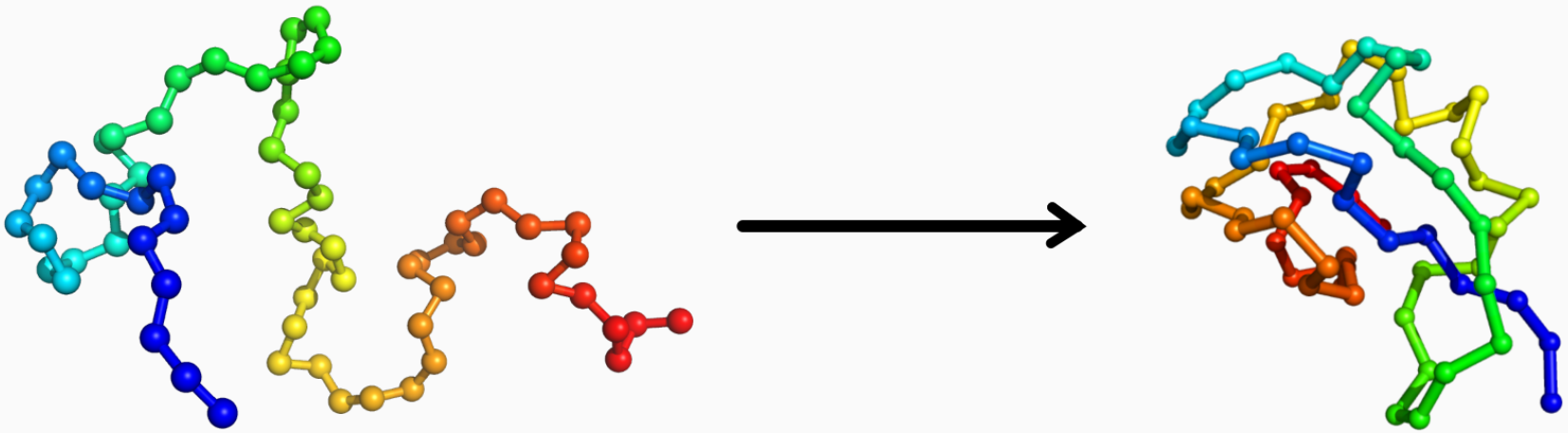
Baiesi, M. *et al.*; A. Sci. Rep. (2019)

Scientific question

Goal

Use MD to sample folding events of a protein having an entangled loop in its native state.
Is the formation of contacts with high $|G'|$ postponed?

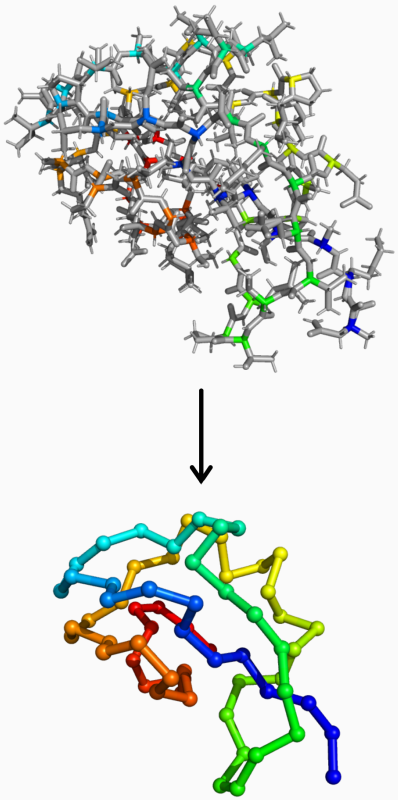
Type III Antifreeze Protein RD1 (PDB: 1UCS)



Ko, Tzu-Ping *et al.*; Biophys. J (2003) | García-Arribas, O. *et al.*; Protein Sci. (2007) | Kundu, S. *et al.*; J. Mol. Graph. Model. (2008) | Chen, H. L. *et al.*; Proteins (2010)

The model

Alpha Carbon Rep



Structure-based

$$
\begin{aligned}
V = & \sum_{i=1}^{N-1} \varepsilon_r^i \left(r^i - r_o^i\right)^2 \\
& + \sum_{i=1}^{N-2} \varepsilon_\theta^i \left(\theta^i - \theta_o^i\right)^2 \\
& + \sum_{i=1}^{N-3} \varepsilon_\phi^i \Big\{ \left[1 - \cos\left(\phi^i - \phi_o^i\right)\right] \\
& + \frac{1}{2}\left[1 - \cos\left(3\left(\phi^i - \phi_o^i\right)\right)\right] \Big\} \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} 4\varepsilon_C^{ij} \left[\left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{12} - \left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{6}\right] \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon_{NN}^{ij} \left(\frac{\sigma_{NN}^{ij}}{\Delta r^{ij}}\right)^{12}
\end{aligned}
$$

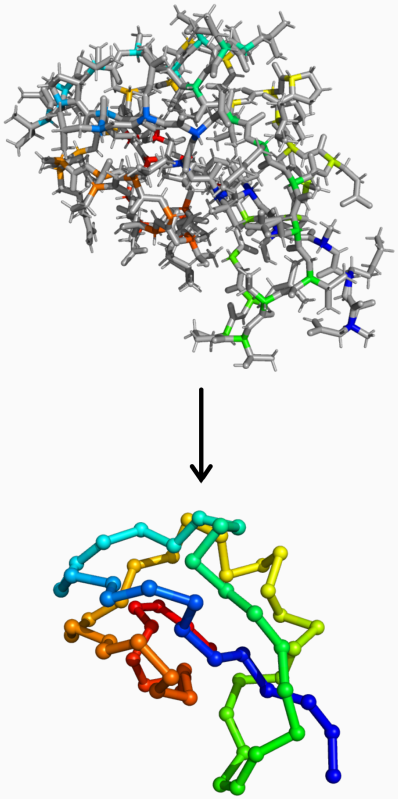
Langevin Dynamics

$$
m_i \frac{d^2\vec{r}^i}{dt^2} = -\gamma \frac{d\vec{r}^i}{dt} - \nabla V + \vec{R}^i
$$

Clementi, C. *et al.*; J. of Mol. Bio. (2000)

The model

Alpha Carbon Rep



Structure-based

$$
\begin{aligned}
V = & \sum_{i=1}^{N-1} \varepsilon_r^i \left(r^i - r_o^i\right)^2 \\
& + \sum_{i=1}^{N-2} \varepsilon_\theta^i \left(\theta^i - \theta_o^i\right)^2 \\
& + \sum_{i=1}^{N-3} \varepsilon_\phi^i \Big\{ \left[1 - \cos\left(\phi^i - \phi_o^i\right)\right] \\
& + \frac{1}{2}\left[1 - \cos\left(3\left(\phi^i - \phi_o^i\right)\right)\right] \Big\} \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} 4\varepsilon_C^{ij} \left[\left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{12} - \left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{6}\right] \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon_{NN}^{ij} \left(\frac{\sigma_{NN}^{ij}}{\Delta r^{ij}}\right)^{12}
\end{aligned}
$$

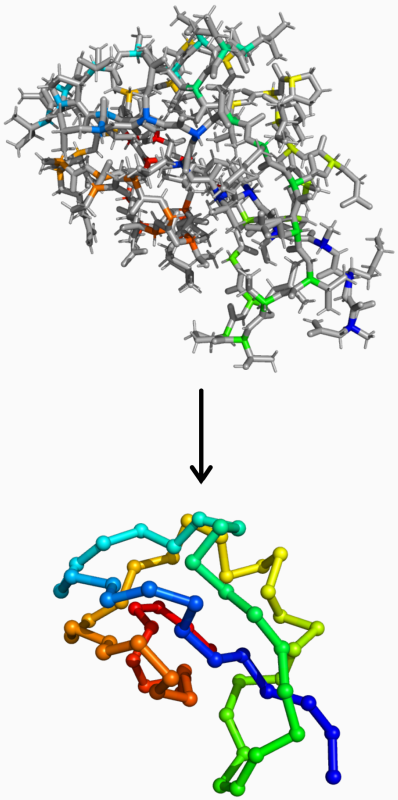
Langevin Dynamics

$$
m_i \frac{d^2\vec{r}^i}{dt^2} = -\gamma \frac{d\vec{r}^i}{dt} - \nabla V + \vec{R}^i
$$

Clementi, C. *et al.*; J. of Mol. Bio. (2000)

The model

Alpha Carbon Rep



Structure-based

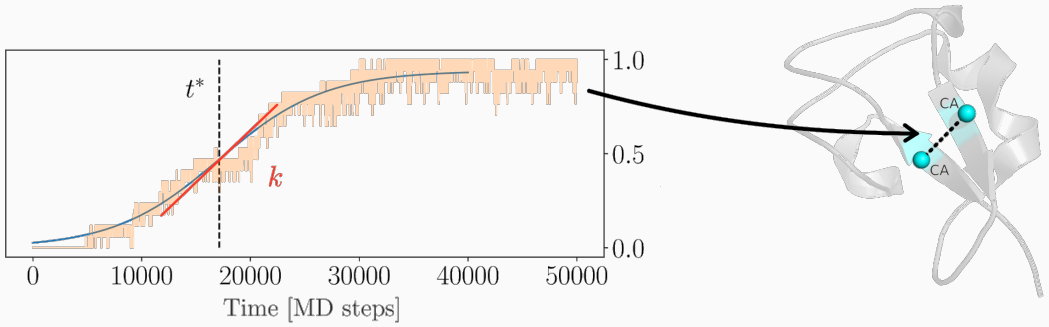
$$
\begin{aligned}
V = & \sum_{i=1}^{N-1} \varepsilon_r^i \left(r^i - r_o^i\right)^2 \\
& + \sum_{i=1}^{N-2} \varepsilon_\theta^i \left(\theta^i - \theta_o^i\right)^2 \\
& + \sum_{i=1}^{N-3} \varepsilon_\phi^i \left\{ \left[1 - \cos\left(\phi^i - \phi_o^i\right)\right] \right. \\
& \left. + \frac{1}{2}\left[1 - \cos\left(3\left(\phi^i - \phi_o^i\right)\right)\right] \right\} \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} 4\varepsilon_C^{ij} \left[\left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{12} - \left(\frac{\sigma^{ij}}{\Delta r^{ij}}\right)^{6}\right] \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon_{NN}^{ij} \left(\frac{\sigma_{NN}^{ij}}{\Delta r^{ij}}\right)^{12}
\end{aligned}
$$

Langevin Dynamics

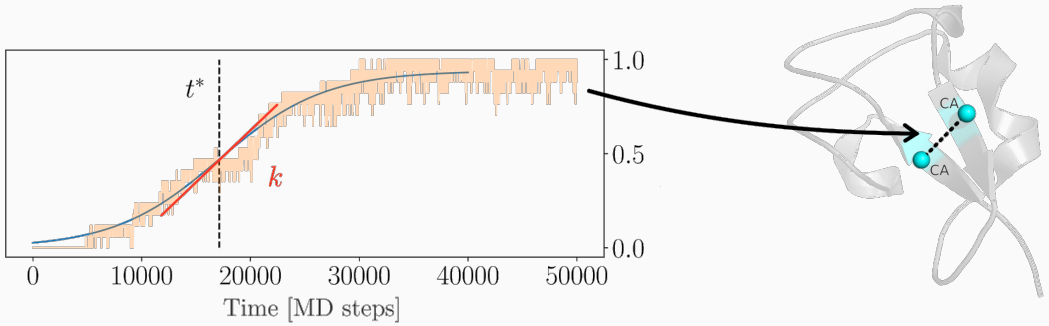
$$
m_i \frac{d^2\vec{r}^i}{dt^2} = -\gamma \frac{d\vec{r}^i}{dt} - \nabla V + \vec{R}^i
$$

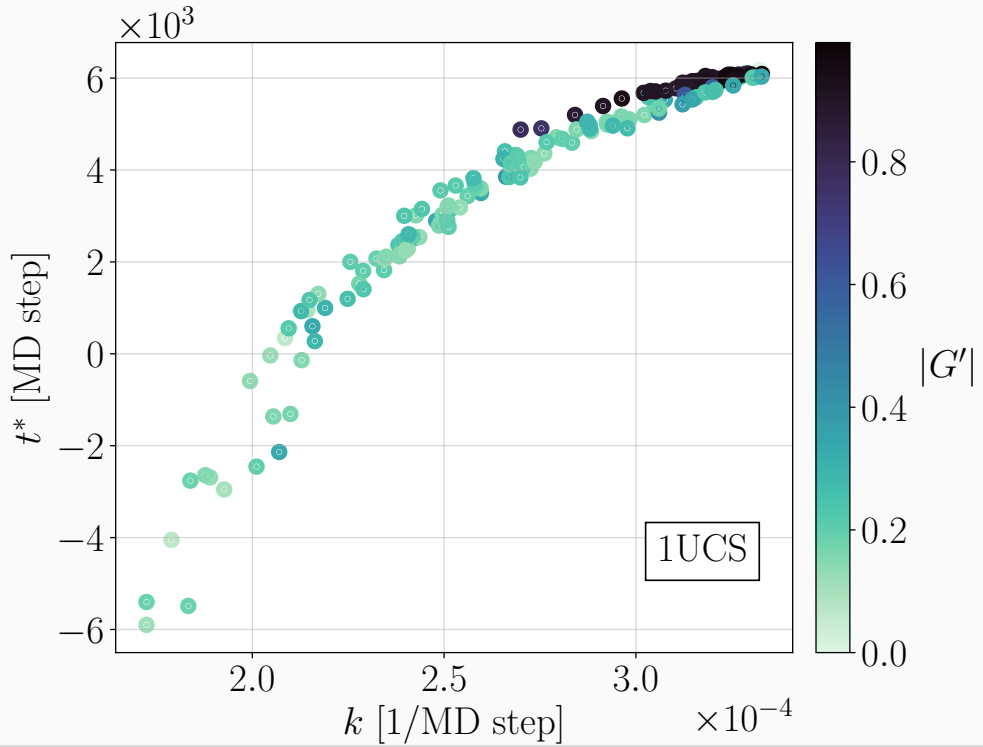
Clementi, C. *et al.*; J. of Mol. Bio. (2000)

Probability of contact formation

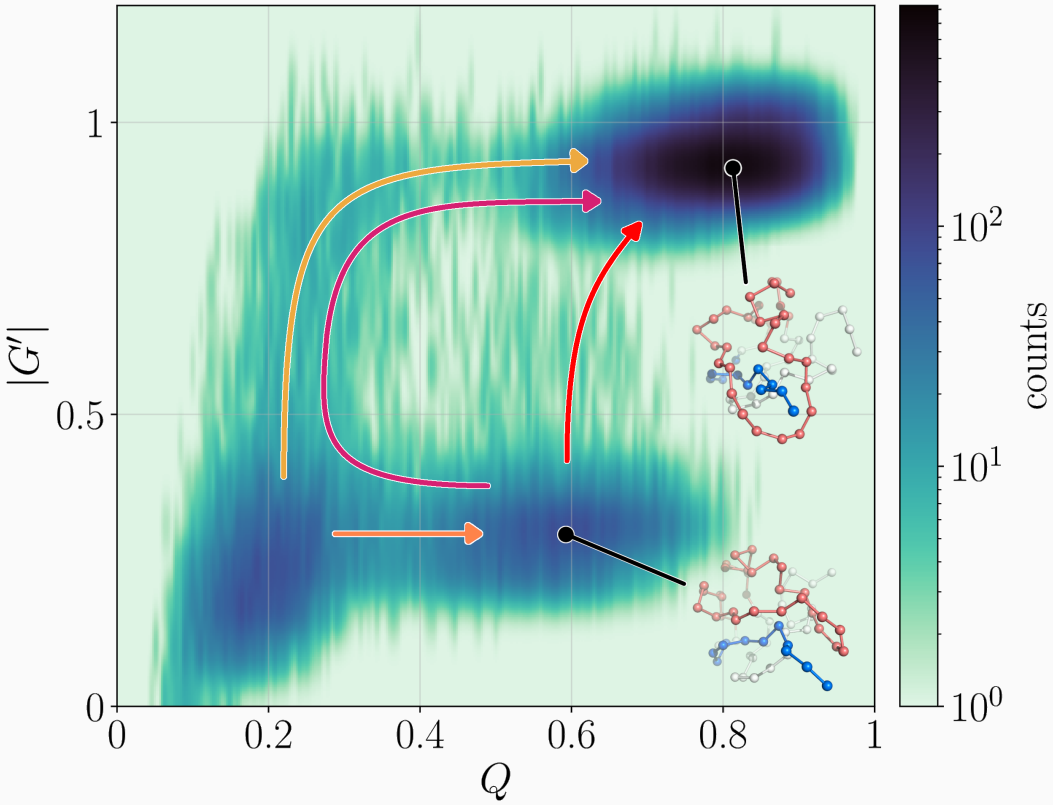


Probability of contact formation





Results Trajectories along reaction coordinates

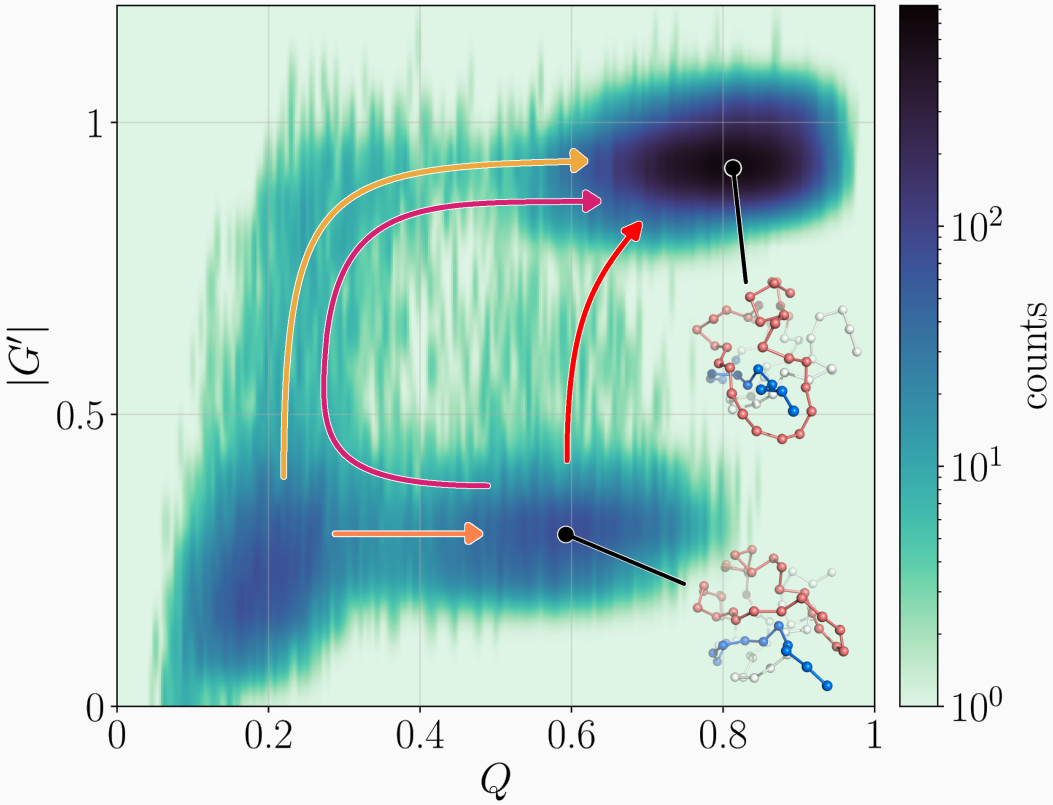


Folding Experiment

$$E_{h\nu}\phi = q + \Delta_R V \left(\frac{C_p \rho}{\beta} \right)_T$$

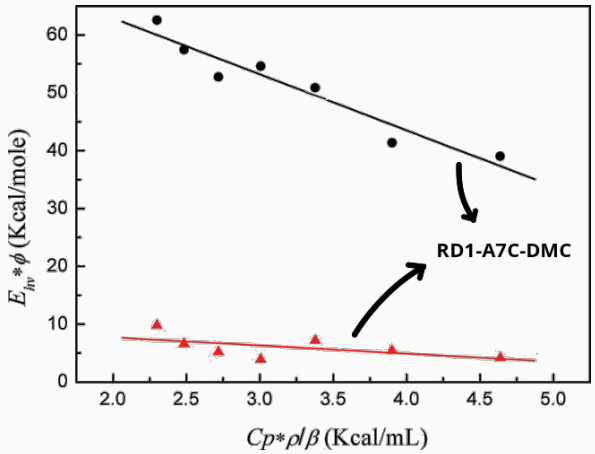
Chen, H. L. *et al.*; Proteins (2010)

Results Trajectories along reaction coordinates



Folding Experiment

$$E_{hv}\phi = q + \Delta_R V \left(\frac{C_p \rho}{\beta} \right)_T$$



Chen, H. L. *et al.*; Proteins (2010)

Outlooks The possible role of cotranslation

- Entangled loops enriched in the C-terminal side of the entanglement they are in
- high $|G'|$ contacts tend to be frustrated

Hypothesis

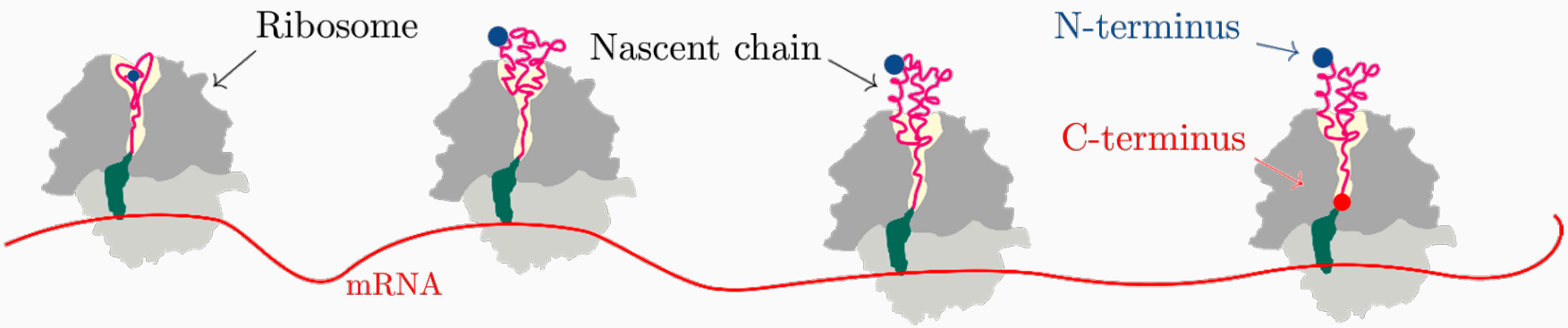
Contacts with high $|G'|$ tend to form in the later stages of the folding

Baiesi, M. *et al.*; A. Sci. Rep. (2019)

Outlooks Testing the role of cotranslation

Probing cotranslational events through:

- MD coarse-grained simulations
- Statistical models: the Wako-Saitô-Muñoz-Eaton (WSME) model

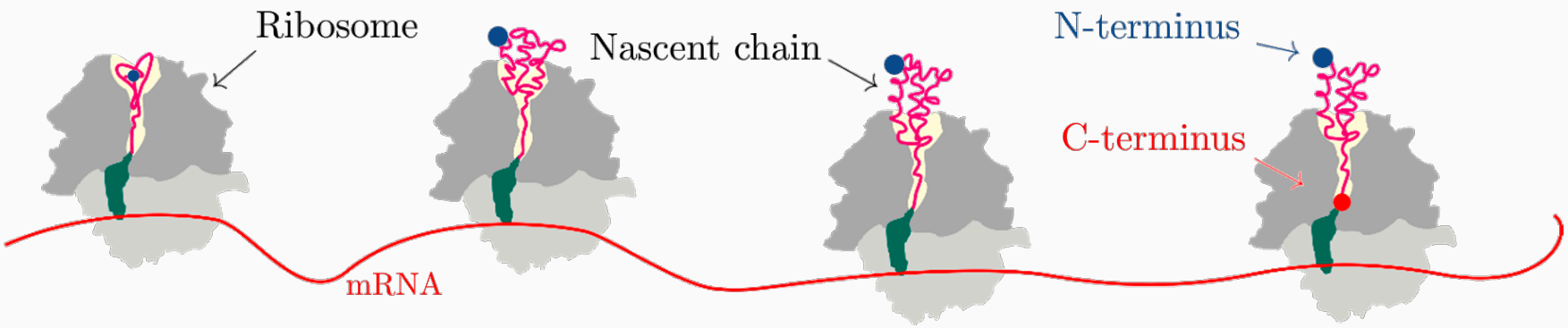


Liutkute, M. *et al.* Biomolecules (2020) | Muñoz, V. *et al.*; PNAS (1999)

Outlooks Testing the role of cotranslation

Probing cotranslational events through:

- MD coarse-grained simulations
- Statistical models: the Wako-Saitô-Muñoz-Eaton (WSME) model



Liutkute, M. *et al.* Biomolecules (2020) | Muñoz, V. *et al.*; PNAS (1999)

Appendix Example of Q and $|G'|$ timeseries