Optimising the spatial structure of BLN protein models by means of thermal cycling

F. Günther[1,2,3], A. Möbius[2], and M. Schreiber[3]

[1] Helmholtz-Zentrum Dresden-Rossendorf

[2] Leibniz Institute for Solid State and Materials Research Dresden

[3] Technical University Chemnitz

CompPhys13, Leipzig, 28.11.2013

Outline

1. Motivation & background
2. BLN model
3. Thermal cycling algorithm
4. Remarks on implementation
5. Results
6. Conclusions

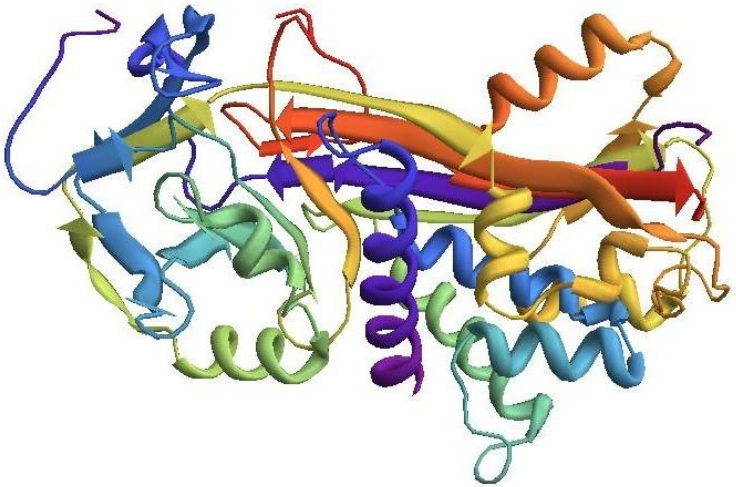
1. Motivation and background

MERMLPLLALGLLAAGFCPAVLCH
PNSPLDEENLTQENQDRGTHVDLG
LASANVDFAFSLYKQLVLKAPDK…

- proteins have a vast area of functions within living organisms
- 3D structure determines biochemical activity
- thermodynamic hypothesis:
  functional fold ⇔ minimal free energy

What ?

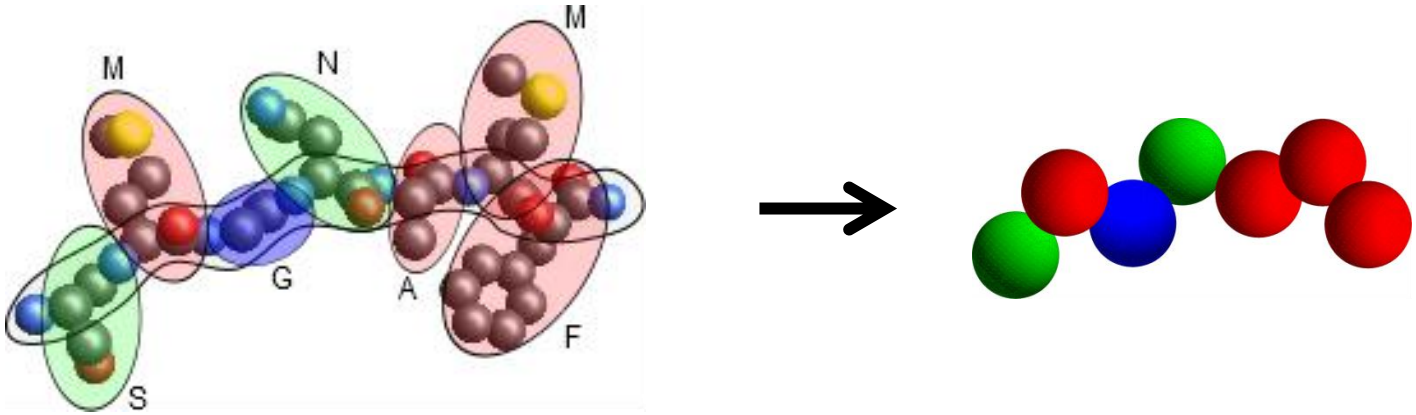
How ?



Backbone-only models

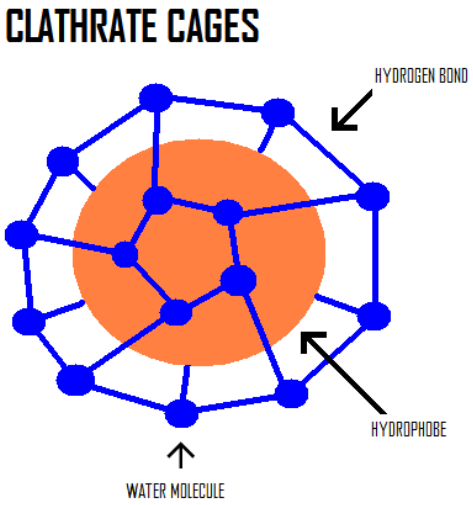
Approximations:

- rigid relative positions of the amino acid atoms
- averaging over all atom interactions of different amino acids, influence of solvent, finite temperature
  → effective potential
- sort amino acids into classes
- consider amino acids as isotropic beads without volume



➔ Only the shape of the backbone is considered

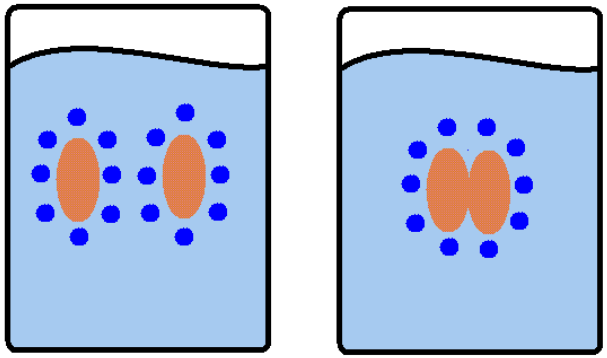
Hydrophobic force



non-polar substrate in polar solvent:

- breaking of existing hydrogen bonds at surface
- formation of new hydrogen bonds (ice-like structure)

more order → less entropy → higher free energy



bringing “drops” together reduces surface

less order → more entropy → lower free energy

→ effective attractive potentials

We focus on finding low free energy protein configurations by means of heuristic optimisation.

2. BLN model

- three kinds of monomers: hydrophobic (B), hydrophilic (L), neutral (N)

$$E = \frac{K_\theta}{2}\sum_i (\theta_i - \theta_0)^2 +$$

$$\epsilon \sum_i [A_i(1+\cos\varphi_i) + B_i(1+\cos 3\varphi_i)] +$$

$$4\epsilon \sum_i \sum_j C_{ij}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - D_{ij}\left(\frac{\sigma}{r_{ij}}\right)^{6}\right] +$$

$$\frac{K_R}{2}\sum_i (R_i - R_0)^2$$

θ… bond angle

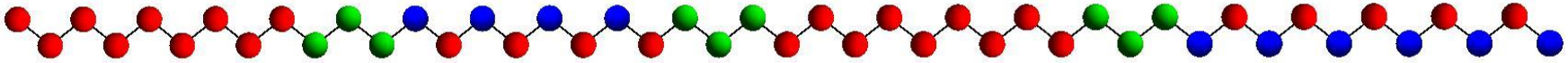
φ… dihedral angle

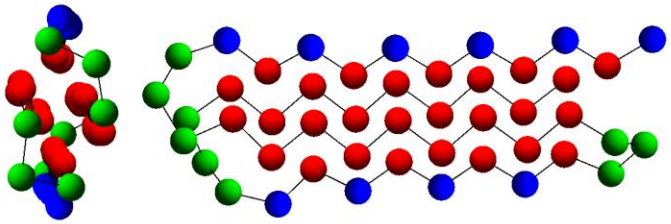
- original model[1]: $R_i = R_0$, corresponding to $K_R = \infty$
- extend configuration space by substituting springs for rigid bonds[2] ($K_R = 231.2$)

[1] J.D Honeycutt, D. Thirumalai, *Biopolymers* **32**, p. 695, 1992

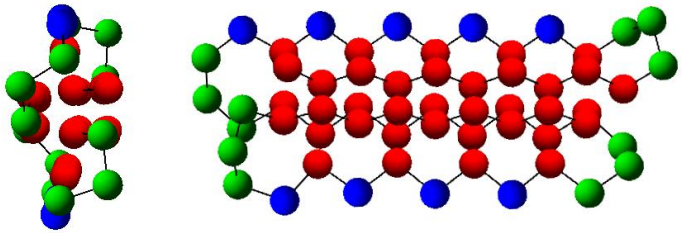
[2] R.S. Berry, N. Elmaci, J.P. Rose, B. Vekhter, *Proc. Natl. Acad. Sci. USA* **94**, p. 9520, 1997

Considered sequences

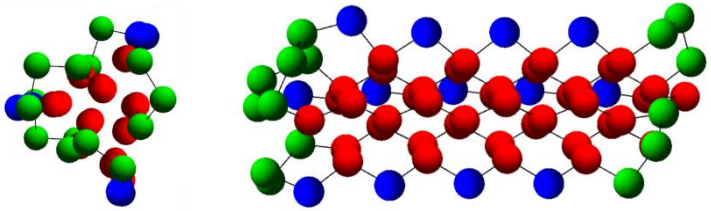




46-bead sequence



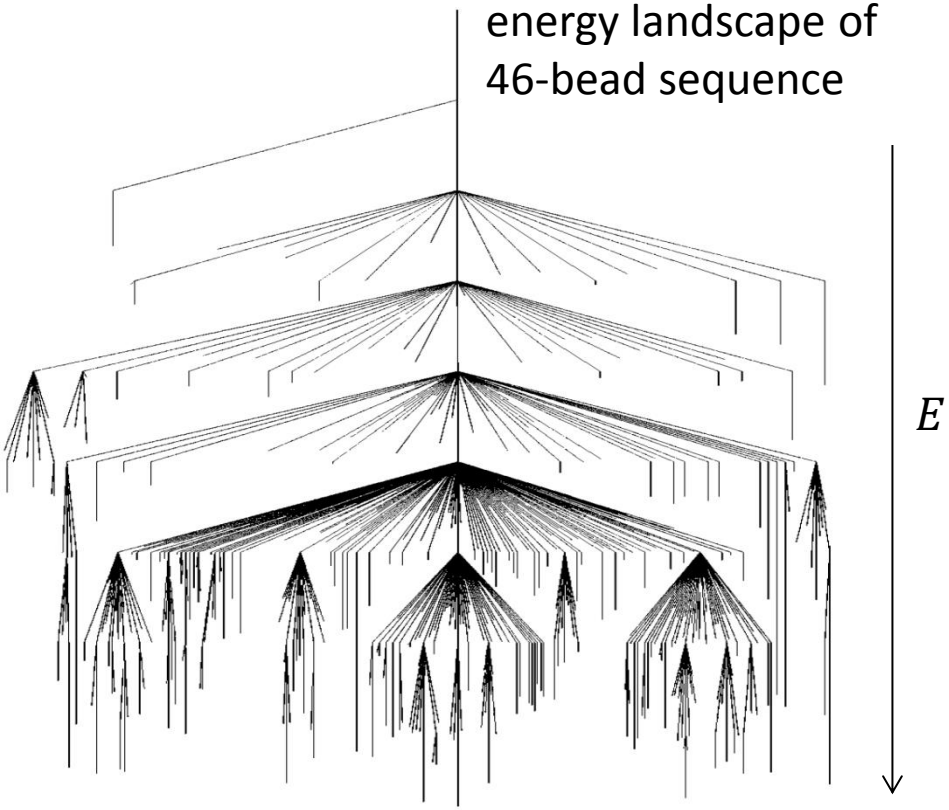
58-bead sequence



69-bead sequence

large number of local minima separated by high barriers

energy landscape of 46-bead sequence



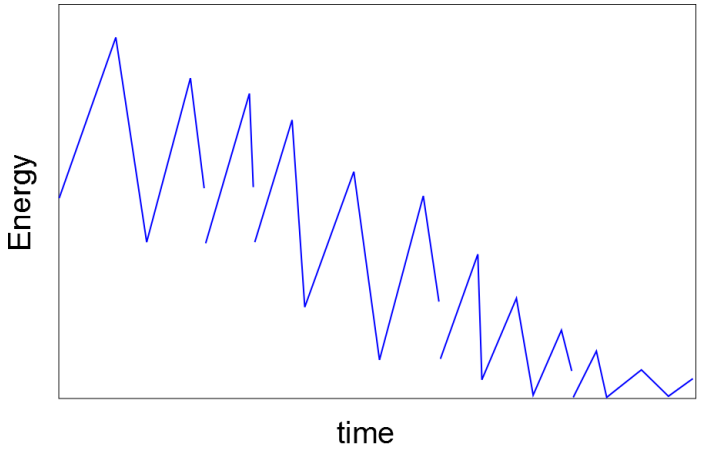
M.A. Miller & D. Wales, *JCP* **111** (14), p. 6610, 1999

3. Thermal cycling algorithm

perform certain number of cycles of three steps

1. disturb current state (heating)
2. search for minimum (quenching)
3. compare initial and quenched state (selection)

reduce amplitude of distortion



Basic feature: **Incomplete heating** to retain gains of previous cycles.

Additional features: Complex moves can be incorporated in quenching.

Consideration of ensembles (genetic local search[2])

i. to reduce risk of getting trapped in "high" minimum, and

ii. to focus on "sensitive" energy regions.

Relatives: thermal bouncing[3] and basin hopping[4]

[1] A. Möbius et all., *Phys. Rev. Lett.* **79**, p. 4297,1997

[2] B. Freisleben, P. Merz, *Proc. IEEE ICEC'96*, p. 616, 1996

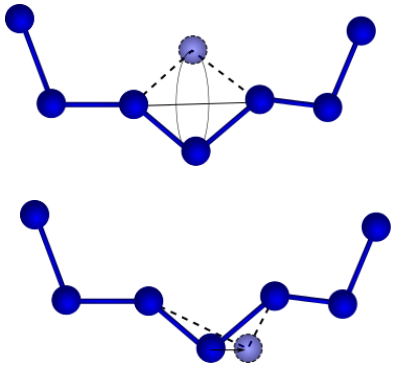
[3] J. Schneider et al., *Phys. Rev E* **58** (4), p. 5085, 1998

[4] D.J. Wales, J.P.K. Doye, *J. Phys. Chem. A* **101**, p. 5111, 1997

4. Remarks on implementation

heating: Metropolis procedure

- two possibilities for move class depending on considered bonds:
  - rotate a node around connection of adjacent nodes; used for rigid bonds
  - move a node in one direction; only for spring-like bonds



- if acceptance rate in heating is smaller than 10% step size is decreased (50% for simulated annealing)

quench: local minimisation

- based on the LBFGS code by Liu and Nocedal[1] which makes use of analytically obtained gradients

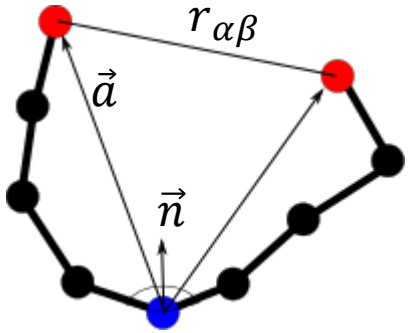
[1] D. Liu, J. Nocedal, *Mathematical Programming B* **45**, p. 503, 1989

Calculation of the gradient considering rigid bonds (RB)

- coordinates: $\theta_i \in [0, \pi]$ and $\varphi_i \in [0,2\pi)$

$$\frac{\partial E_\theta}{\partial \theta_i} = K_\theta(\theta_i - \theta_e)$$

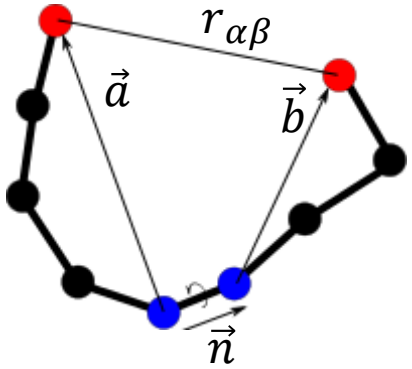
$$\frac{\partial E_\varphi}{\partial \varphi_i} = -\epsilon[A_i \sin\varphi_i + 3B_i \sin 3\varphi_i]$$



$$\frac{\partial E_r}{\partial \mu_k} = \sum_\alpha \sum_\beta \frac{\partial E_{r_{\alpha\beta}}}{\partial r_{\alpha\beta}} \frac{\partial r_{\alpha\beta}}{\partial \mu_k}$$

$$\frac{\partial E_{r_{\alpha\beta}}}{\partial r_{\alpha\beta}} = -4\varepsilon\sigma^{-1} C_{ij} \left[12 \left(\frac{\sigma}{r_{ij}} \right)^{13} - 6 D_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{7} \right]$$

$$\frac{\partial r_{\alpha\beta}}{\partial \mu_k} = \frac{\vec{b} - \vec{a}}{|\vec{b} - \vec{a}|} . (\vec{n} \times \vec{b})$$



→ **boundary conditions are problematic**

→ **calculation effort is proportional to 3rd power of system size**

→ **singularities at $\theta_i = \pi$**

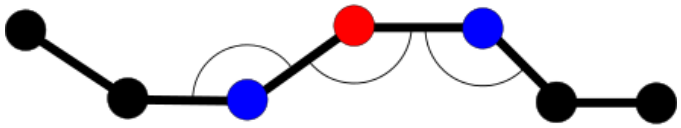
Calculation of the gradient considering spring-like bonds (SLB)

- generalized coordinates: $\vec{r}_i \in \mathbb{R}^3$

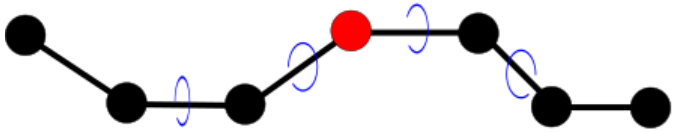
$$\frac{\partial E_r}{\partial \vec{r}_i} = -4\varepsilon\sigma^{-2} \sum_j C_{ij} \left[12 \left(\frac{\sigma}{r_{ij}} \right)^{14} - 6D_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{8} \right] (\vec{r}_i - \vec{r}_j)$$

$$\frac{\partial E_R}{\partial \vec{r}_i} = K_R \left(\left(1 - \frac{R_e}{R_i} \right) (\vec{r}_i - \vec{r}_{i-1}) + \left(1 - \frac{R_e}{R_{i+1}} \right) (\vec{r}_{i+1} - \vec{r}_i) \right)$$

$$\frac{\partial E_\theta}{\partial \vec{r}_i} = \sum_{j=i}^{i+2} \frac{\partial E_\theta}{\partial \theta_j} \frac{\partial \theta_j}{\partial \vec{r}_i}$$



$$\frac{\partial E_\varphi}{\partial \vec{r}_i} = \sum_{j=i}^{i+3} \frac{\partial E_\theta}{\partial \varphi_j} \frac{\partial \varphi_j}{\partial \vec{r}_i}$$

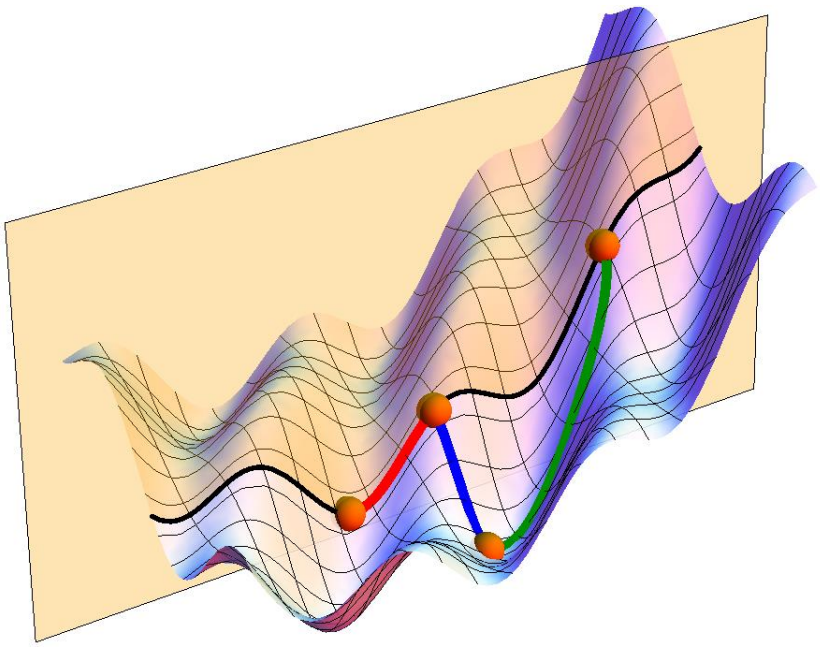


→ **no boundary conditions to be considered**

→ **calculation effort is proportional to 2nd power of system size**

→ **enlarged number of degrees of freedom**

Hybrid procedure to study original model



1. start with rigid bonds
2. preliminary optimisation considering spring-like bonds
   (roughly 95% of total simulation)
3. projection of best state to original bond length
4. single final minimisation using rigid bonds

5. Results

CPU time used for 10^4 minimisations

	46-bead	58-bead	69-bead
τ_{RB}	5.8 h	19.0 h	1.4 d
τ_{SLB}	8.1 min	11.7 min	15.2 min
$\tau_{\mathrm{RB}}/\tau_{\mathrm{SLB}}$	42.8	97.4	132.6

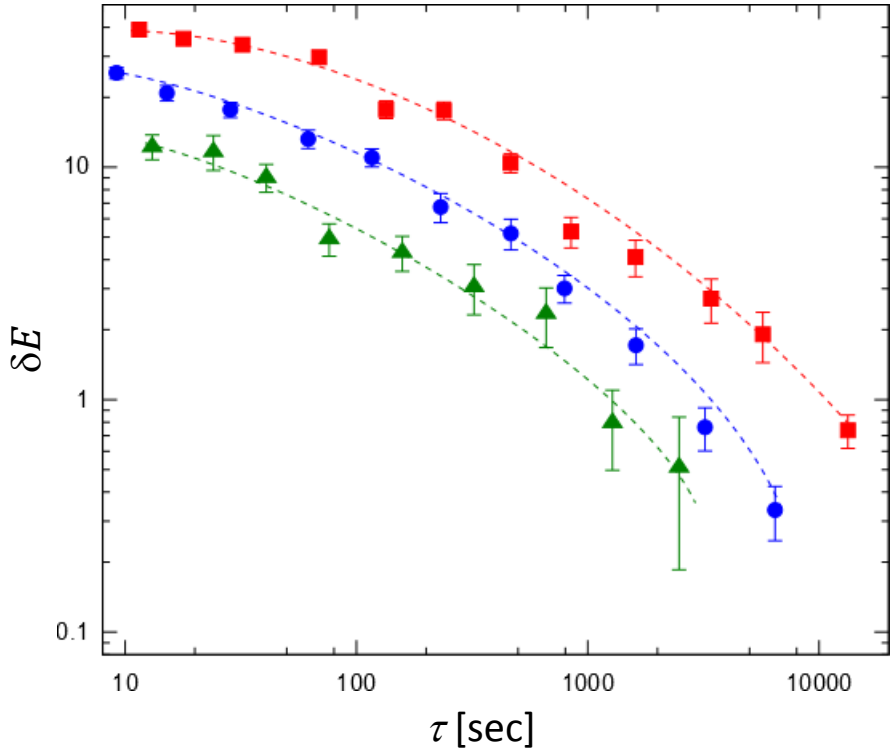
→ **SLB minimisations are much faster than RB minimisations**

→ **ratio $\tau_{\mathrm{RB}}/\tau_{\mathrm{SLB}}$ becomes larger with increasing chain length**

→ **Be careful: total result of RB minimisation can be better**

What have we reached?

Comparison of individual algorithms for the 46-bead chain with rigid bonds:



simulated annealing, with additional minimisation of best state found

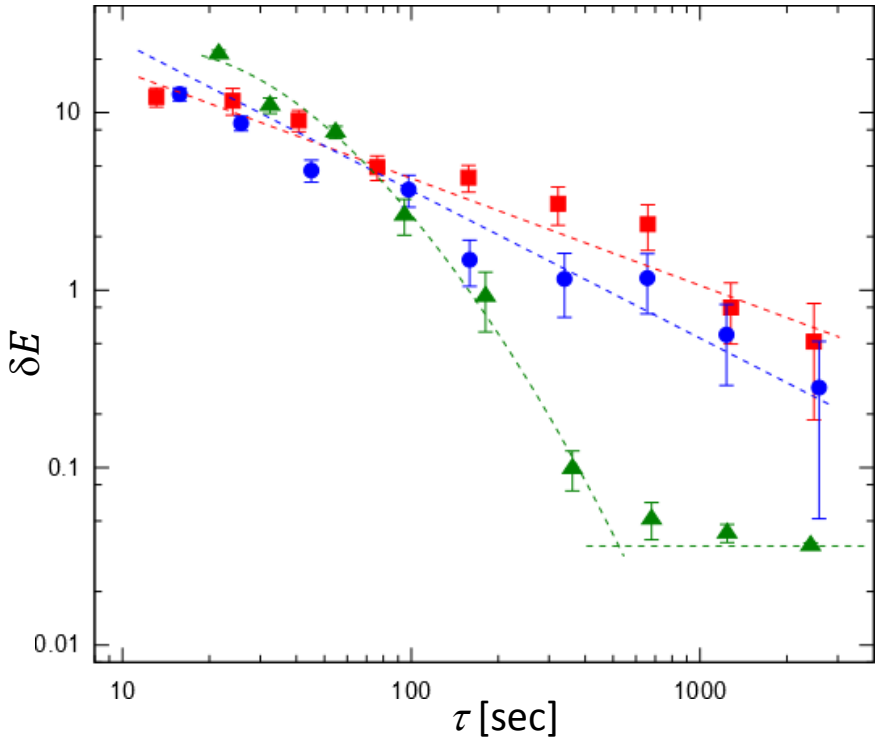
multi-start local search

thermal cycling

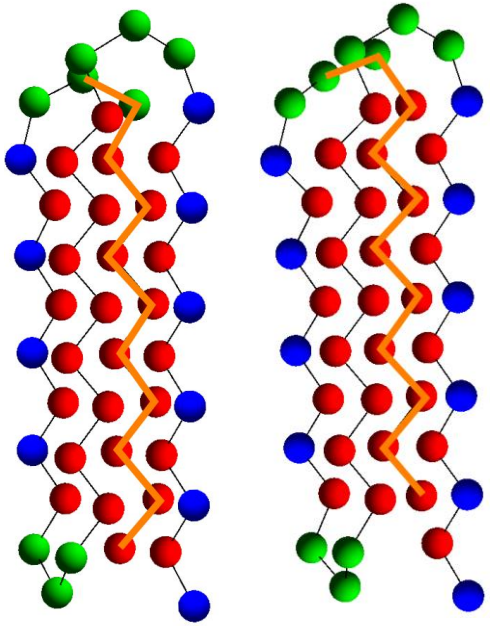
→ **Simple multi-start local search beats simulated annealing.**

→ **Thermal cycling is most appropriate.**

To what extent can ensemble consideration help?



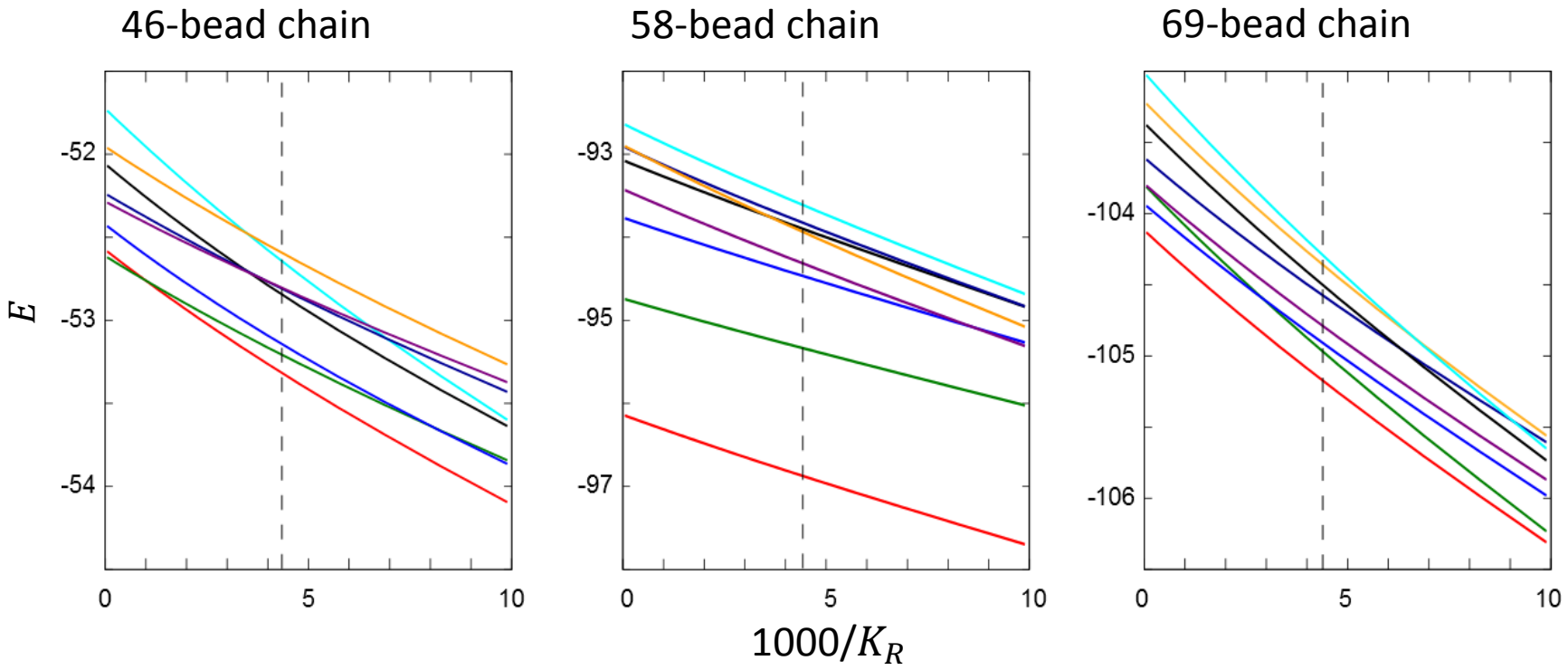
The performance of thermal cycling of single sample is improved by considering an ensemble of 10 or 30 states.



→ **Ensemble approach improves computation.**

→ **Calculations with large numbers of cycles get stuck in a local minimum above the ground state.**

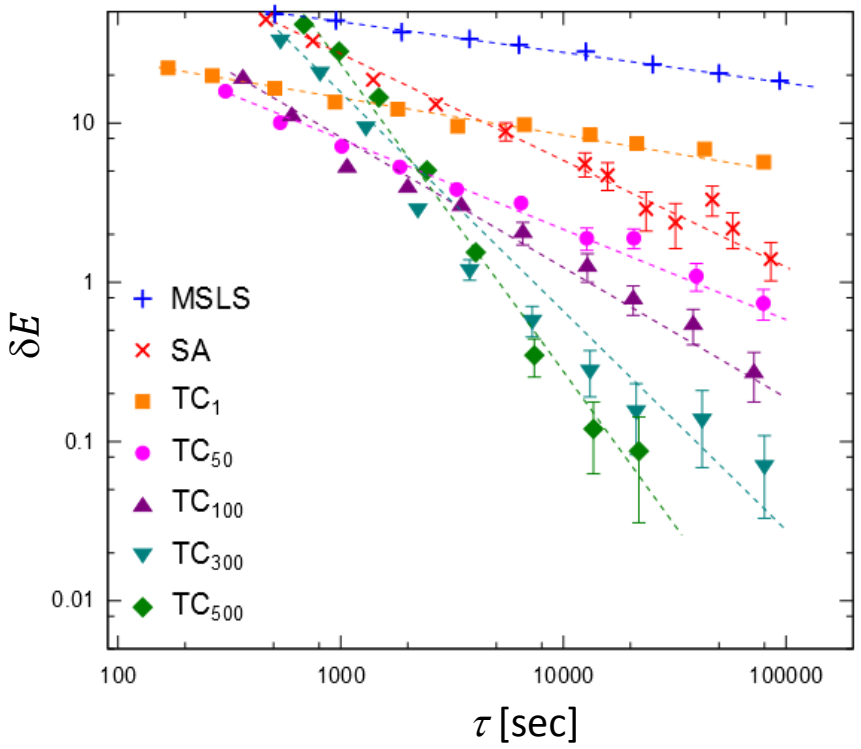
Sequence of lowest levels changes with varying spring constant

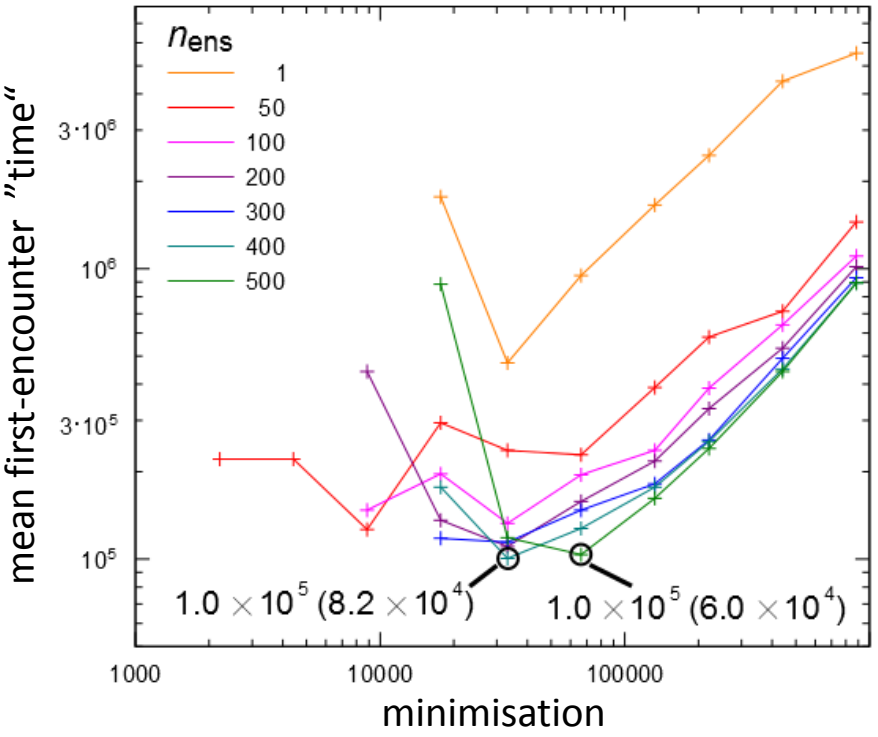


→ **Although the additional harmonic potential only constrains the separation of consecutive beads, it has significant effects. Even the sequence of levels is modified. This explains differences to Berry et al.[1]**

[1] R.S. Berry, N. Elmaci, J.P. Rose, B. Vekhter, *Proc. Natl. Acad. Sci. USA* **94**, p. 9520, 1997

Efficiency of our approach for the 69-bead SLB model





value for basin hopping, M.T. Oakley et al., J. Phys. Chem. B **115** (2011) 11525:

2.6×10^4 (2.3×10^4)

→ **TC is superior to MSLS and SA.**

→ **Required number of minimisations is currently only by factor of four larger than in best literature study.**

6. Conclusions

- TC is a very appropriate algorithm for the BLN model.
- Extension of the model by softening the rigid bonds simplifies treatment,
- but level crossings appear.

Open tasks to be investigated in future studies:

- Apply optimised schedules.
- Focus on sensible regions of sequence in heating.
- Extend selection criterion of TC for ensemble by niching restrictions.
- Include crossovers of different ensemble states.

6. Conclusions

- TC is a very appropriate algorithm for the BLN model.
- Extension of the model by softening the rigid bonds simplifies treatment,
- but level crossings appear.

Open tasks to be investigated in future studies:

- Apply optimised schedules.
- Focus on sensible regions of sequence in heating.
- Extend selection criterion of TC for ensemble by niching restrictions.
- Include crossovers of different ensemble states.

Many thanks to

Professor Jeroen van den Brink, Ulrike Nitzsche and Philipp Cain,

and to you for your attention.