

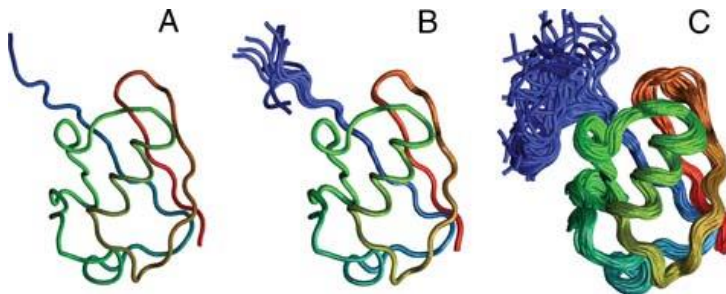
Optimization of Force Field Parameters Using Structural Information

Francis

Dec 1, 2017

Molecular dynamics simulations

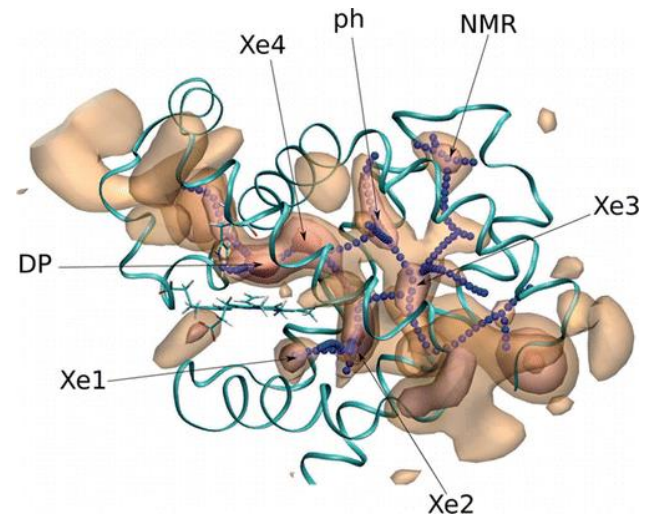
- Energy refinement, protein dynamics, protein-ligand binding



(A) Crystallographic structure obtained by X-ray diffraction (pdb code 1UBQ).

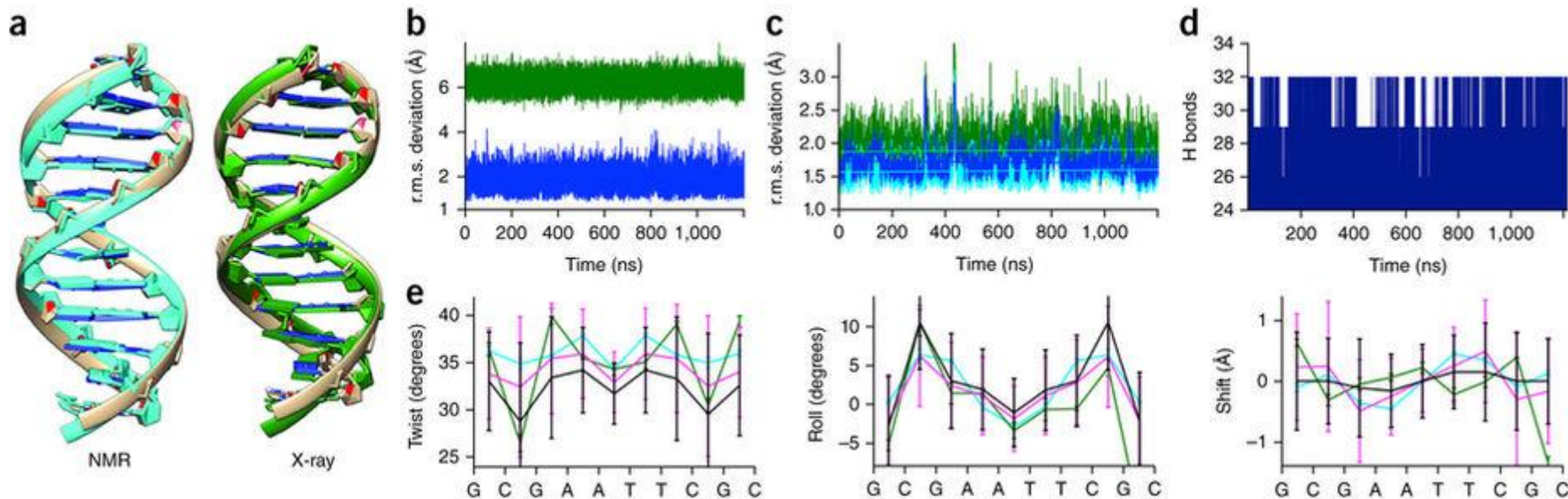
(B) Static average(s) structure obtained by solution nuclear magnetic resonance spectroscopy (pdb code 1D3Z).

(C) Molecular dynamics simulation (1 μ s with CHARMM force field)



Pathway of CO entering myoglobin

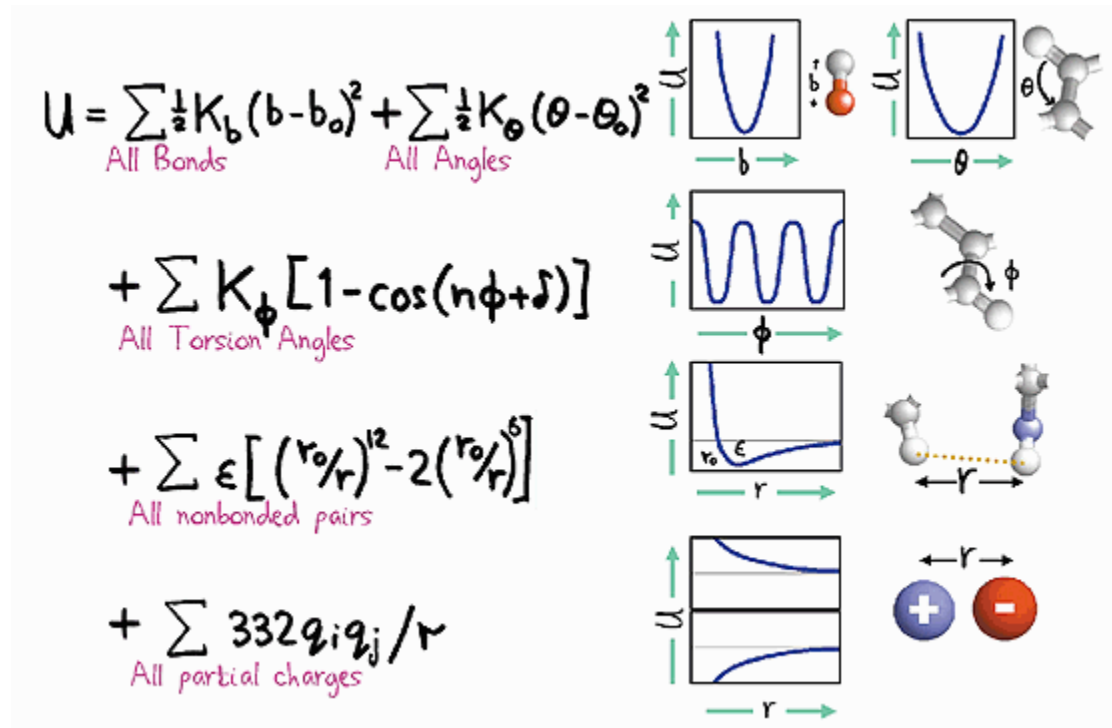
MD results are affected by the accuracy of force field



- (A) Comparison of the MD average structure (light brown) with NMR and X-ray.
- (B) RMSD from B-DNA (blue) and A-DNA (green)
- (C) RMSD from X-ray (green) and NMR (blue) structures
- (D) Total number of hydrogen bonds
- (E) Comparison of average values of helical rotational parameters (twist, roll and shift) per base-pair step coming from NMR (cyan), X-ray (green), 1-μs parmbosc0 trajectory (black) and 1.2-μs parmbosc1 trajectory (magenta) data

Functional forms of force fields

- Energy functions in CFF



Reference data for force field development

- Energy surface (QM)
- Spectroscopy and crystallography
(for equilibrium bond lengths and angles, vibration frequencies)
- Condensed-phase properties
 - Density, enthalpy of vaporization, solvation free energy
- Crystal structure
- NMR measurements
 - NOE: Distance between ^1H nuclei i and j ($<6 \text{ \AA}$)
 - ^3J coupling: Dihedral angle between bond vectors i and j
 - RDC: Angle between atoms i and j and the molecular frame
 - Chemical shift: chemical environment of nucleus i



Turning statistical physics models into materials design engines

Marc Z. Miskina, Gurdaman Khaira, Juan J. de Pablo, and Heinrich M. Jaeger

PNAS, 2016

Optimization equations

$$\Delta\rho(x|\lambda_i) \sim \rho(x|\lambda_i) [f(x) - \langle f(x) \rangle] \quad \text{Eq. 1}$$

ρ : probability distribution of x

λ_i : tunable parameters

x : configuration

$f(x)$: reward function

Choice of $f(x)$:

rank of the configuration x in the current ensemble
by some quality function $g(x)$

Optimization equations

Continuous time

$$\frac{\partial \rho(x|\lambda_i)}{\partial t} \sim \rho(x|\lambda_i) [f(x) - \langle f(x) \rangle]$$

$$\frac{\partial \lambda_i}{\partial t} \frac{\partial \log \rho}{\partial \lambda_i} \sim f(x) - \langle f(x) \rangle$$

Approximate equation:

$$\frac{\partial \lambda_i}{\partial t} = \left\langle \partial_{\lambda_i} \log \rho \partial_{\lambda_j} \log \rho \right\rangle^{-1} \left\langle [f(x) - \langle f(x) \rangle] \partial_{\lambda_j} \log \rho \right\rangle \quad \text{Eq. 2}$$

Fisher information

$f(X; \theta)$: probability density of X given θ

$$\mathcal{I}(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \mid \theta \right]$$

The information $f(X; \theta)$ contains that can be used to estimate θ

Matrix form:

$$[\mathcal{I}(\theta)]_{ij} = E \left[\left(\frac{\partial}{\partial \theta_i} \log f(X; \theta) \right) \left(\frac{\partial}{\partial \theta_j} \log f(X; \theta) \right) \mid \theta \right]$$

Properties of Eq. 2

- Invariant to invertible reparameterization
 - The choice of unit, normalization, rotation
- Invariant to changes in the function the characterizes the design goal
- “Deterministic” evolution of parameters

Compare Eq. 2 with other optimization algorithms

Goal: minimize $S = (\langle A(x|\lambda_i) \rangle - A^*)^2$

$$\frac{\partial}{\partial \lambda_i} \langle A(x|\lambda) \rangle = \int \frac{\partial}{\partial \lambda_i} \rho(x|\lambda) A(x|\lambda) dx = -\beta \left(\left\langle A \frac{\partial E}{\partial \lambda_i} \right\rangle - \langle A \rangle \left\langle \frac{\partial E}{\partial \lambda_i} \right\rangle \right)$$

Gradient descent:

$$\Delta \lambda_i \propto -\frac{\partial S}{\partial \lambda_i} = -2(\langle A(x|\lambda) \rangle - A^*) \frac{\partial \langle A(x|\lambda) \rangle}{\partial \lambda_i}$$

Gauss-Newton:

$$\mathbf{r} = \vec{A}^* - \langle \vec{A}(x|\lambda) \rangle$$

$$\Delta \lambda = -(\mathbf{J}_r^T \mathbf{J}_r)^{-1} \mathbf{J}_r^T \mathbf{r}(\lambda)$$

$$(\mathbf{J}_r)_{ij} = \frac{\partial r_i}{\partial \lambda_j} = -\frac{\partial}{\partial \lambda_j} \langle \vec{A}(x|\lambda) \rangle$$

Design a Ising model that maximizes magnetization

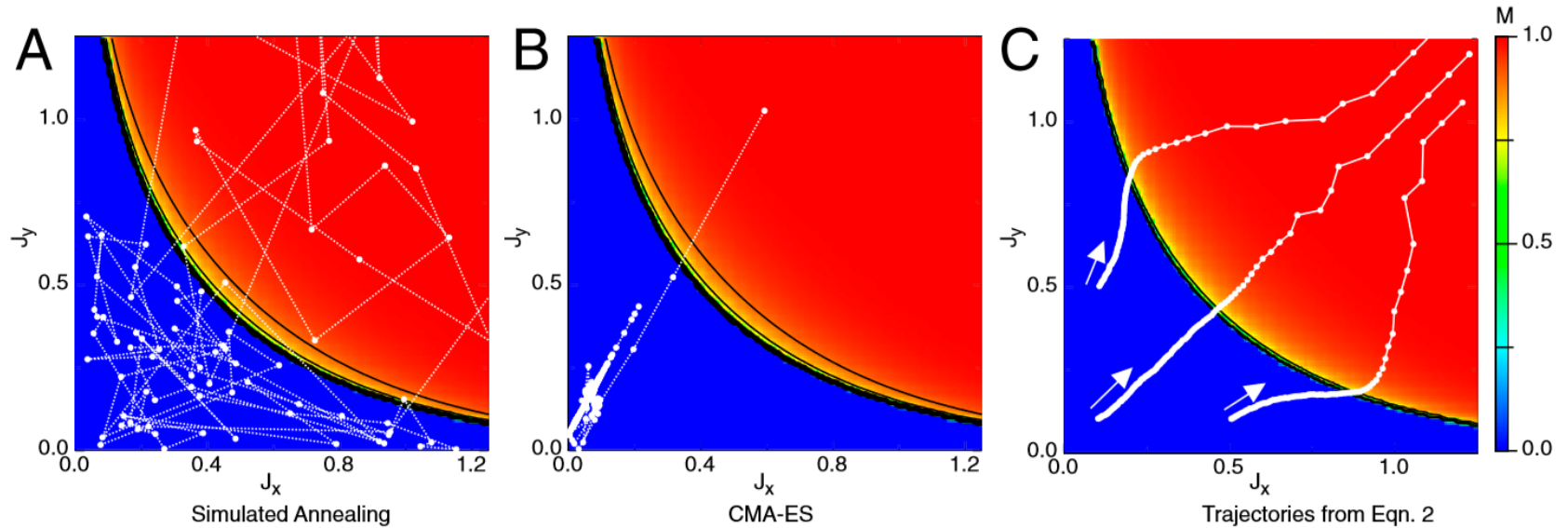
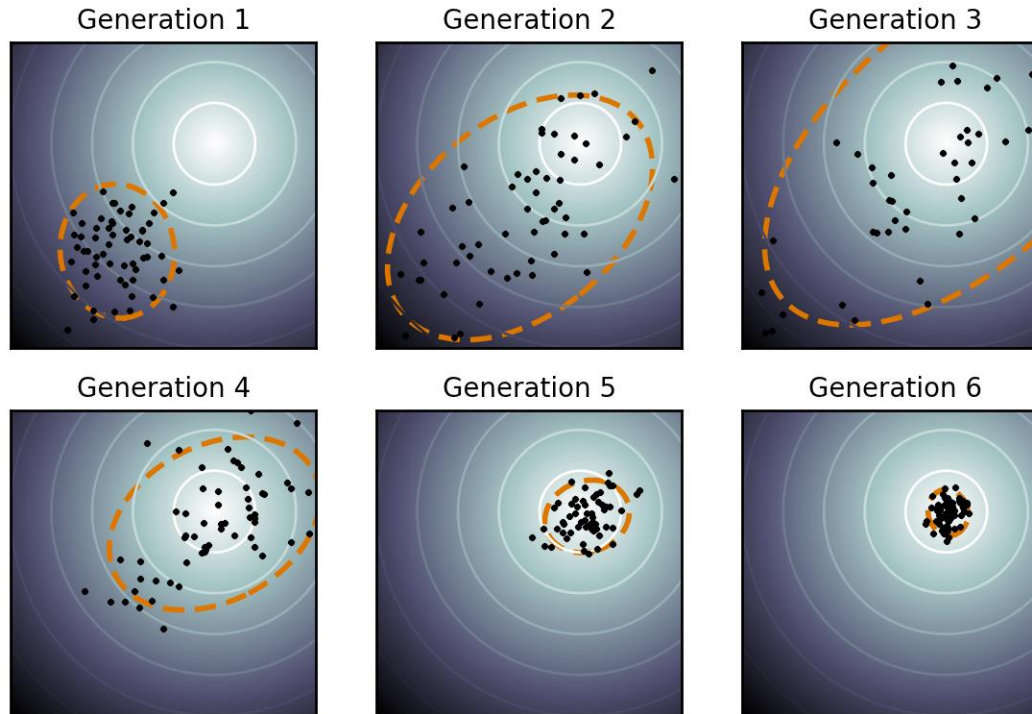


Fig. 1. Phase transitions and flat landscapes. If a materials design problem features a phase transition as part of the search space, black-box optimizers can struggle or fail due to inefficient use of simulation data. For example, if optimizing the magnetization, M , of a 2D Ising model by changing the coupling constants along the x and y directions (J_x/kT and J_y/kT , respectively), black-box methods like simulated annealing (A) and the CMA-ES (B) walk randomly if initialized in the zero magnetization state. Our method (C) interprets fluctuations in solution quality against model components as encoded by Eq. 2 and thus can navigate to the magnetized state regardless of initialization.

Design a Ising model that maximizes magnetization

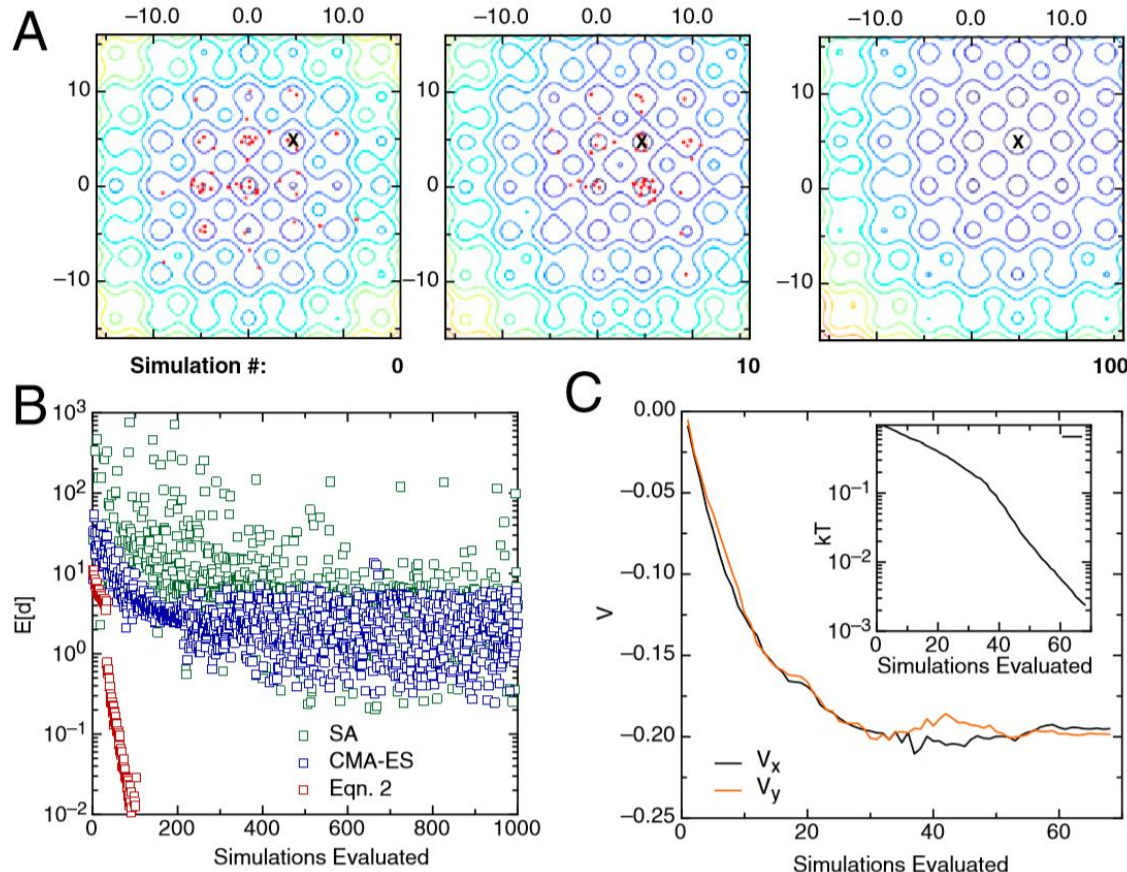


Covariance Matrix Adaption Evolution Strategy (CMA-ES)

Trap a particle in a well

$$H = h_s + v_{x_1}x_1 + v_{x_2}x_2; h_s = \sum_i \left(-\cos \frac{2\pi}{5} x_i + \frac{x_i^2}{25} \right)$$

Optimize kT, v_{x_1}, v_{x_2} to trap the particle at (5,5)

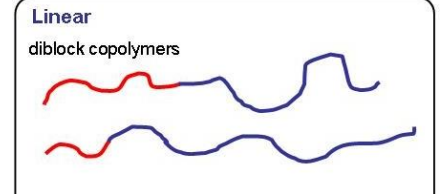
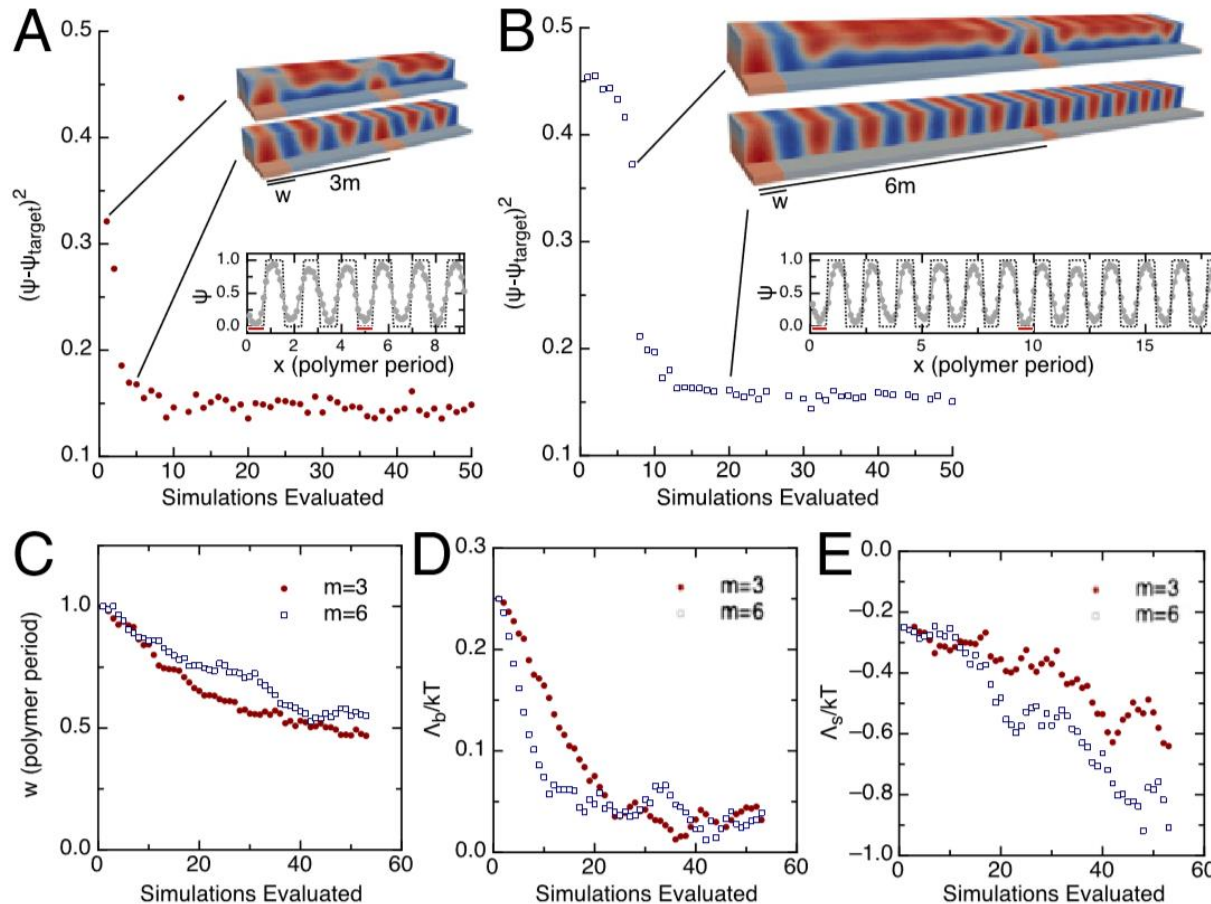


Sampled particle locations after 0, 10, 100 iterations (A). Evolution of distance to the target position with different optimization methods (B). How temperature and field parameters change at each iteration (C).

Directed self-assembly

Optimize w, Λ_b, Λ_s (stripe width, background and stripe interactions)

Goal: The target stripe replicates m times between the two guiding stripes



Optimizing directed self-assembly of diblock copolymer with $m = 3$ (A) and $m = 6$ (B).

Inserts shows the order parameter $\Psi(x) = \frac{n_a}{n_a + n_b}$.

Plots of parameters generated by Eq. 2 (C-E).

Summary

- An algorithm is designed that combines ideas from evolution algorithm (fitness) and quasi-Newton methods
- The algorithm is invariant to reparameterization and rank-preserving changes in the design goal (quality function)
- The algorithm is suitable for optimizing models towards goals defined on some order parameters