

An Information-Based Approach to Coarse-Graining

June 5, 2012, KITP Multiscale Modeling



Avi Chaimovich



Scott Carmichael

M. Scott Shell

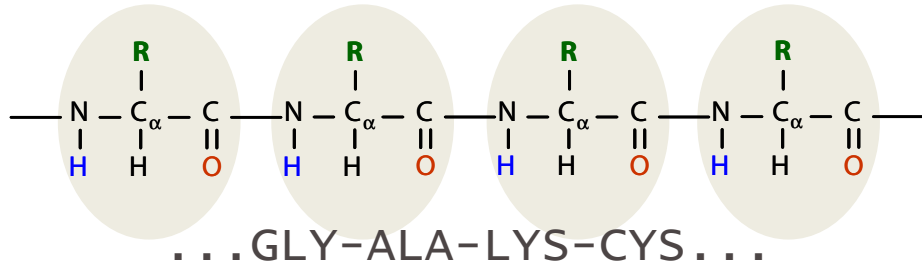
Department of Chemical Engineering
University of California Santa Barbara

Support:

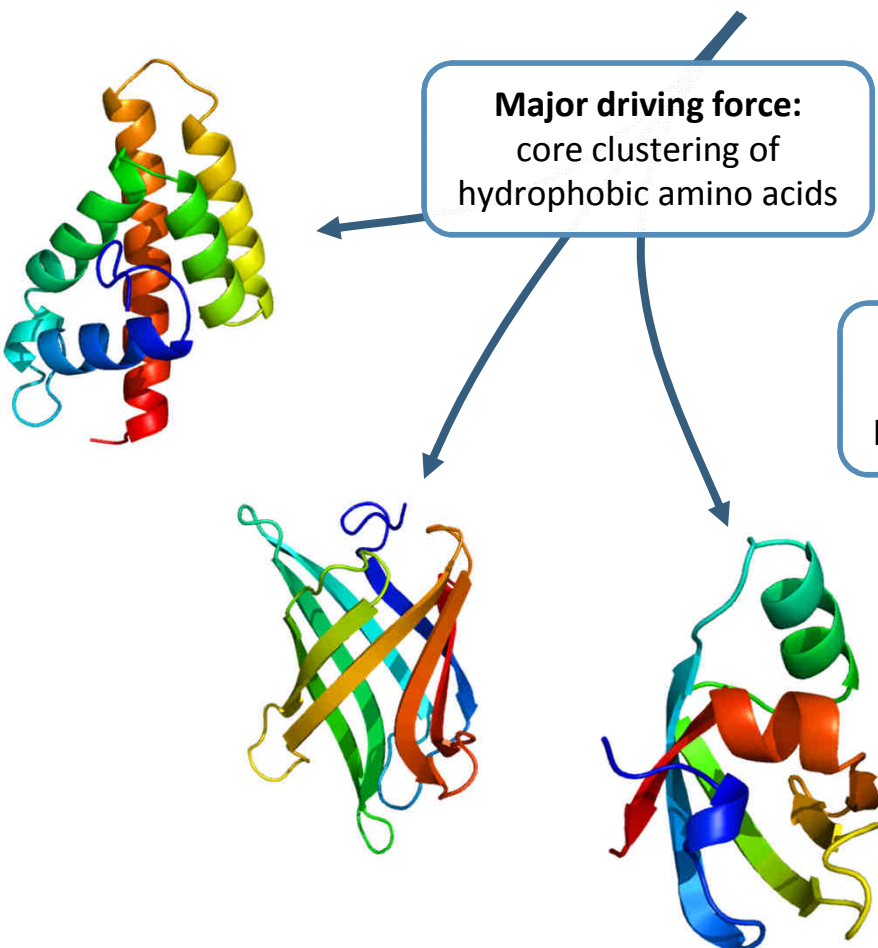
ACS Petroleum Research Fund

Dreyfus Foundation

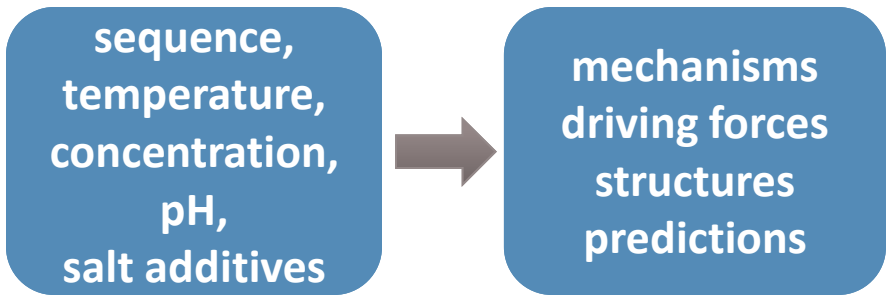
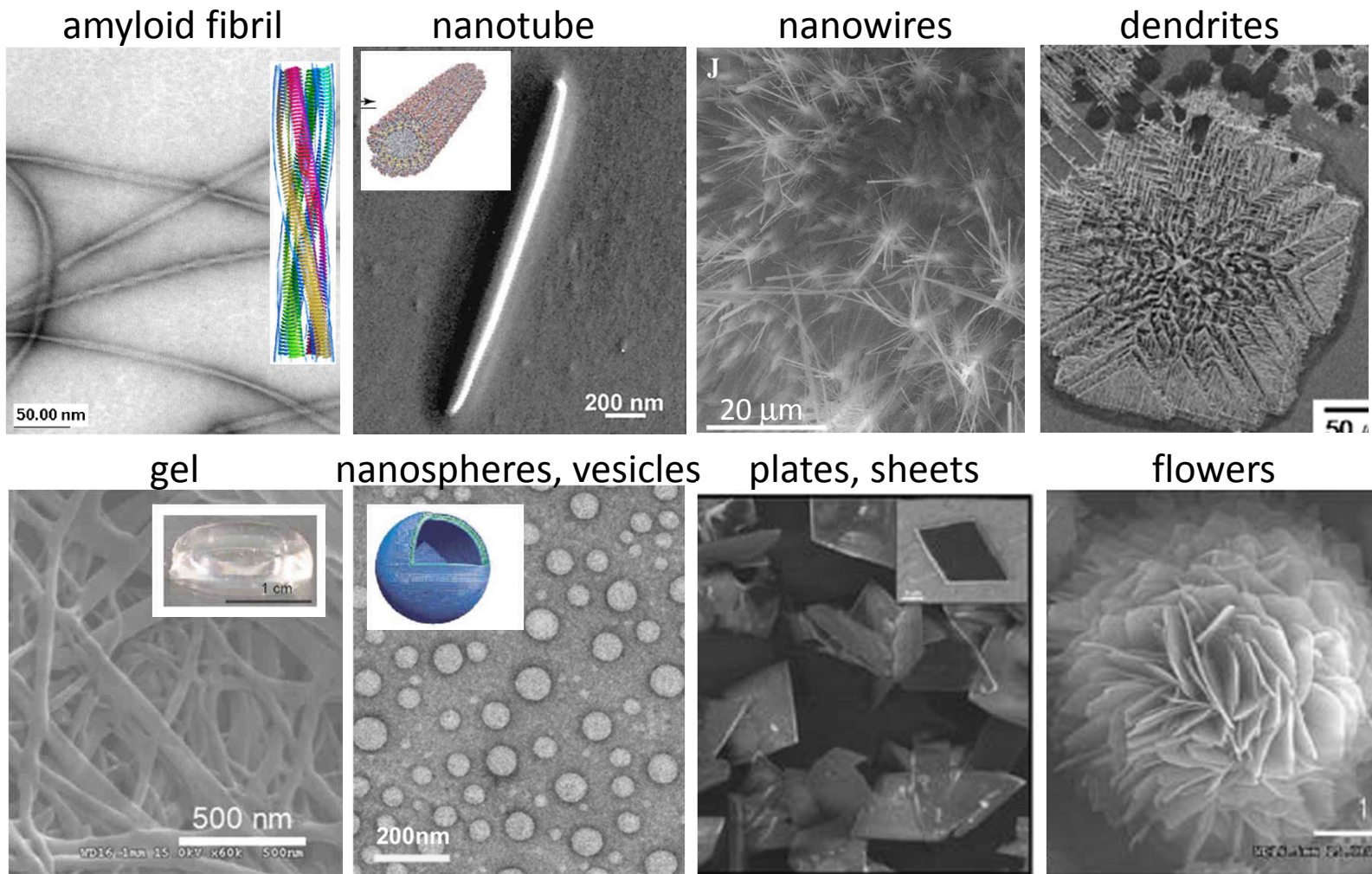
National Science Foundation



linear protein

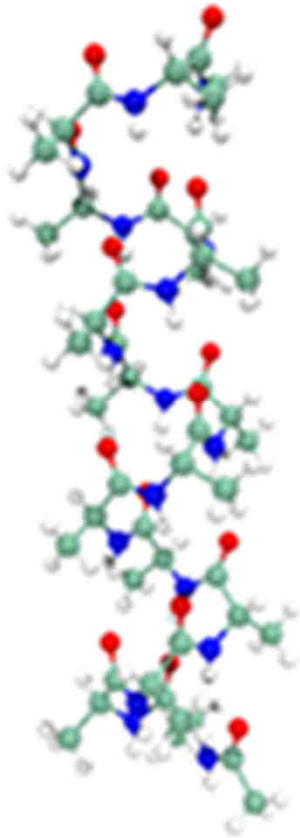


Major driving force:
inter-protein hydrophobic interactions

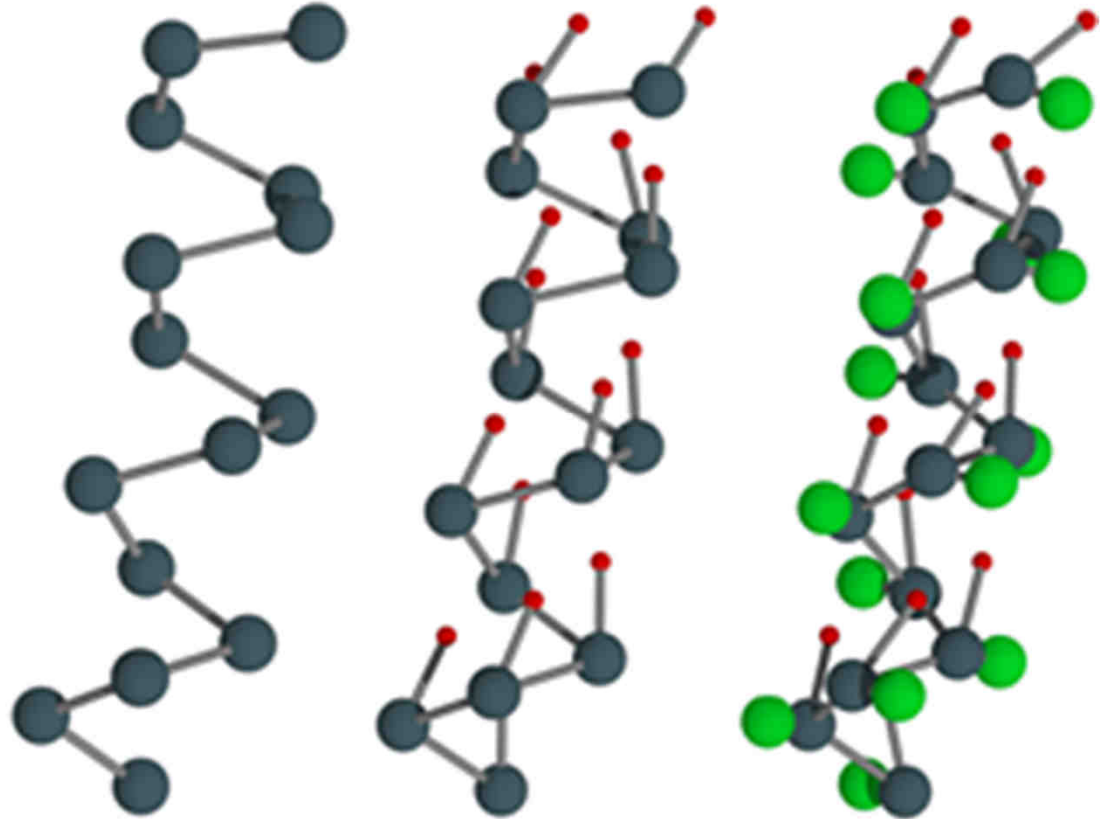


1. Tycko et al., Ann. Rev. of Phys. Chem. (2001)
2. Reches, et al. Science (2003)
3. Amdursky et al, Biomacromolecules (2011)
4. Han et al, Colloids and Biosurfaces B (2011)
5. Yan et al., Chem. Soc. Rev. (2010)
6. Yan et al., Angewandte Chem. Int. Ed. (2007)
7. Govindaraju et al, Supramolec. Chem. (2011)
8. Su et al, J. Mater. Chem. (2010)

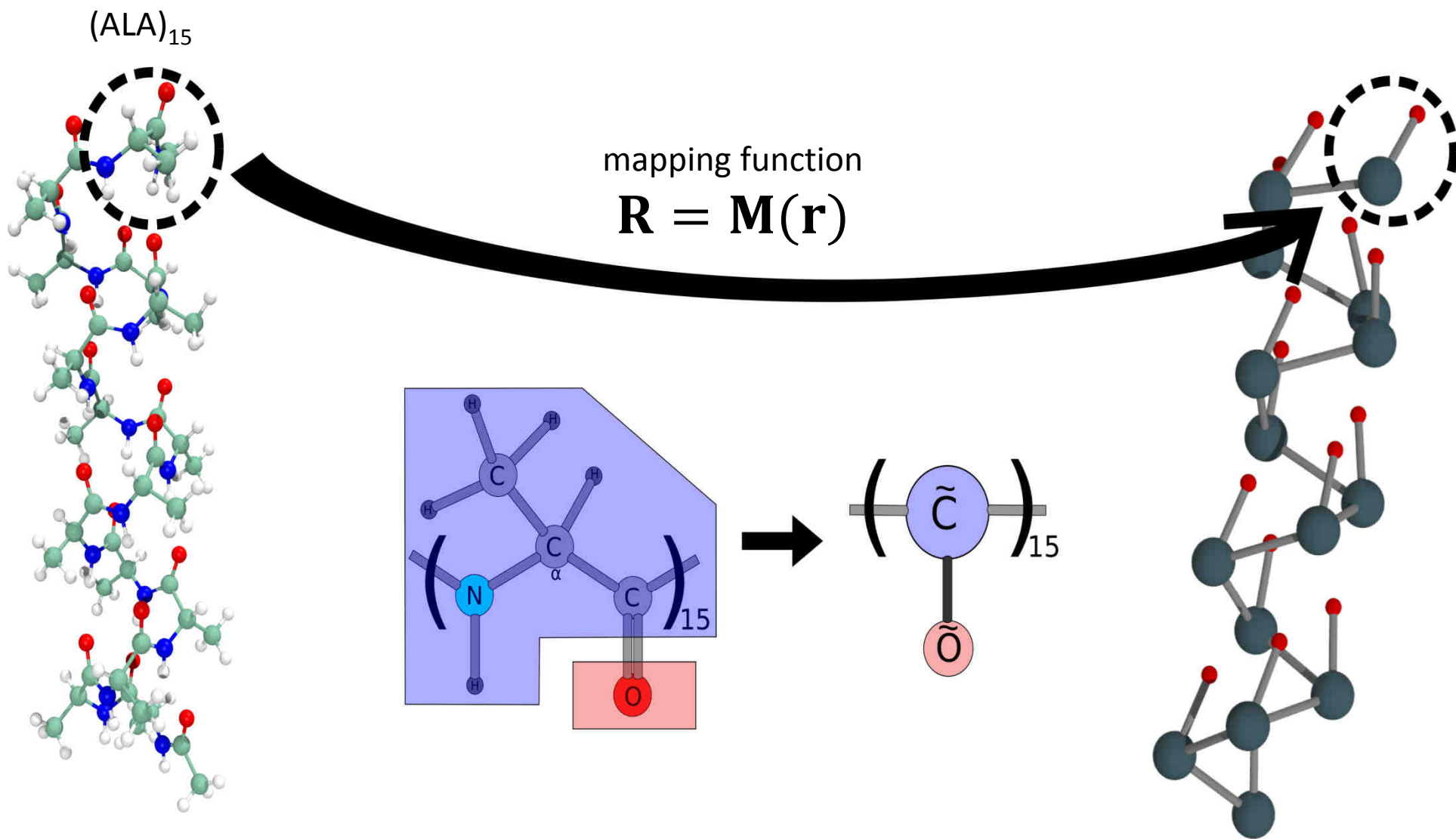
all-atom peptide model

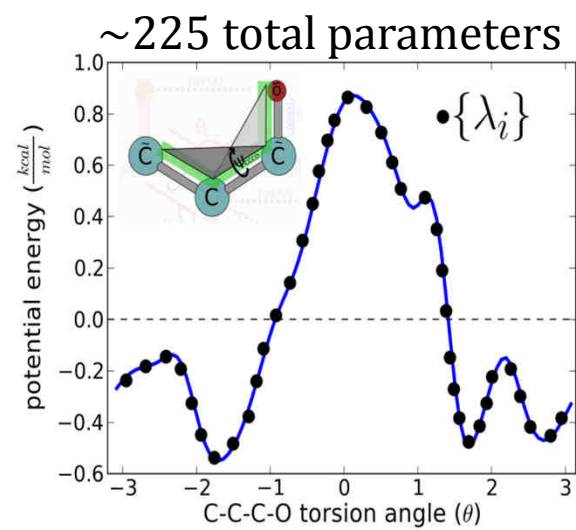
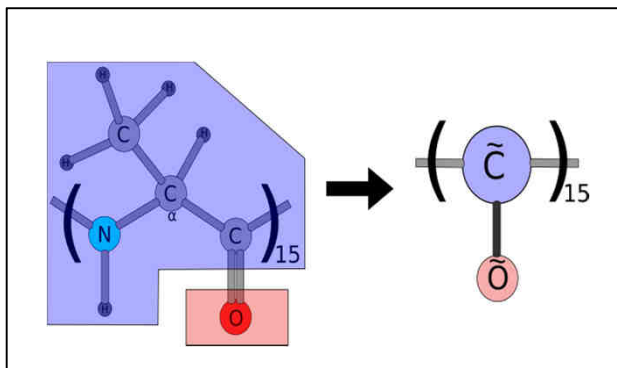
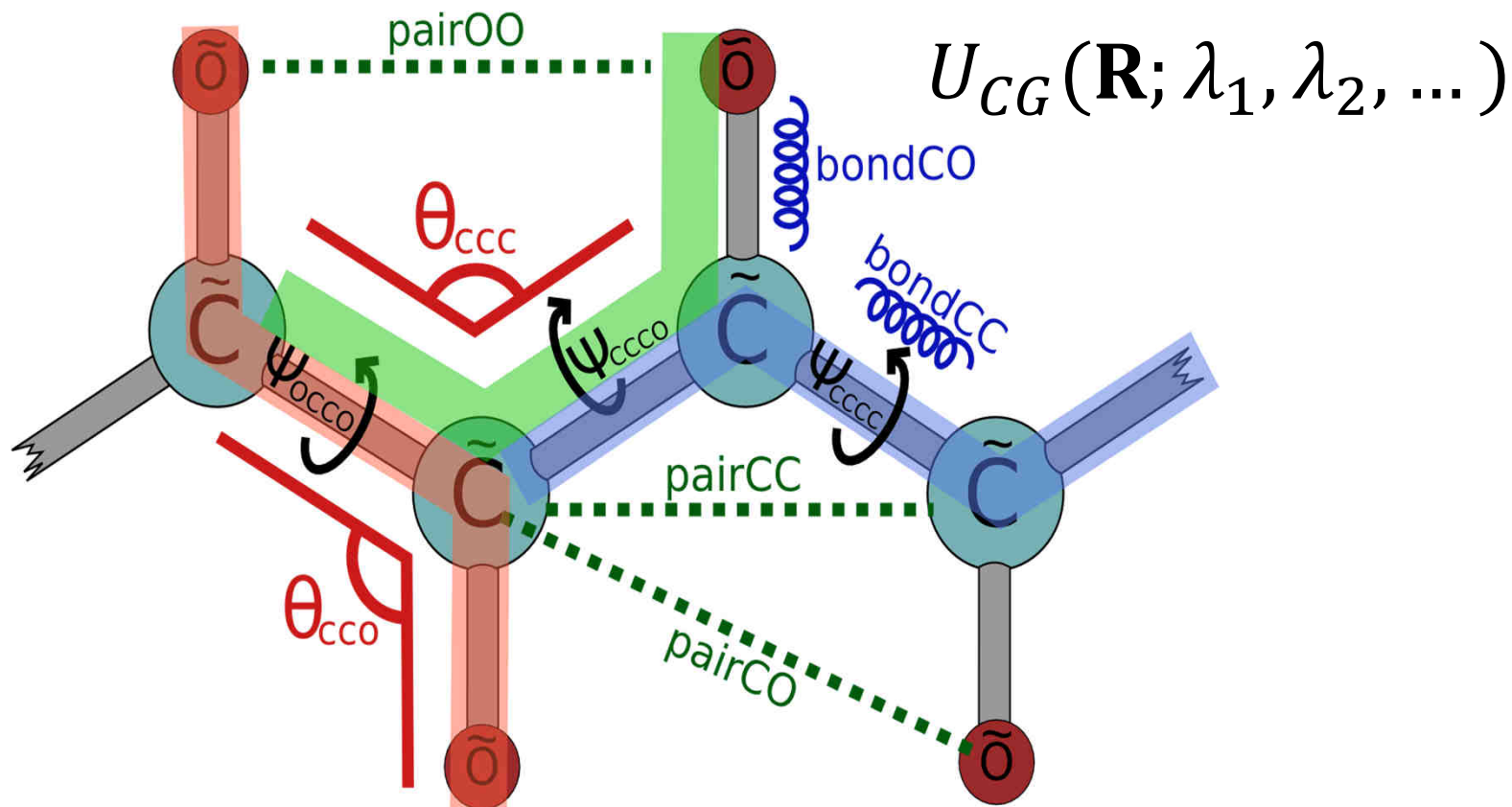


coarse-grained models of varying detail



known or "target" $U_{AA}(\mathbf{r})$





What do we want in a coarse-graining method?

practical

stable and robust for models with 1000+ parameters

relatively fast

amenable to a wide class of CG models

fundamental

prediction or control of coarse-graining errors

as free of assumptions as possible

physical insight into model design

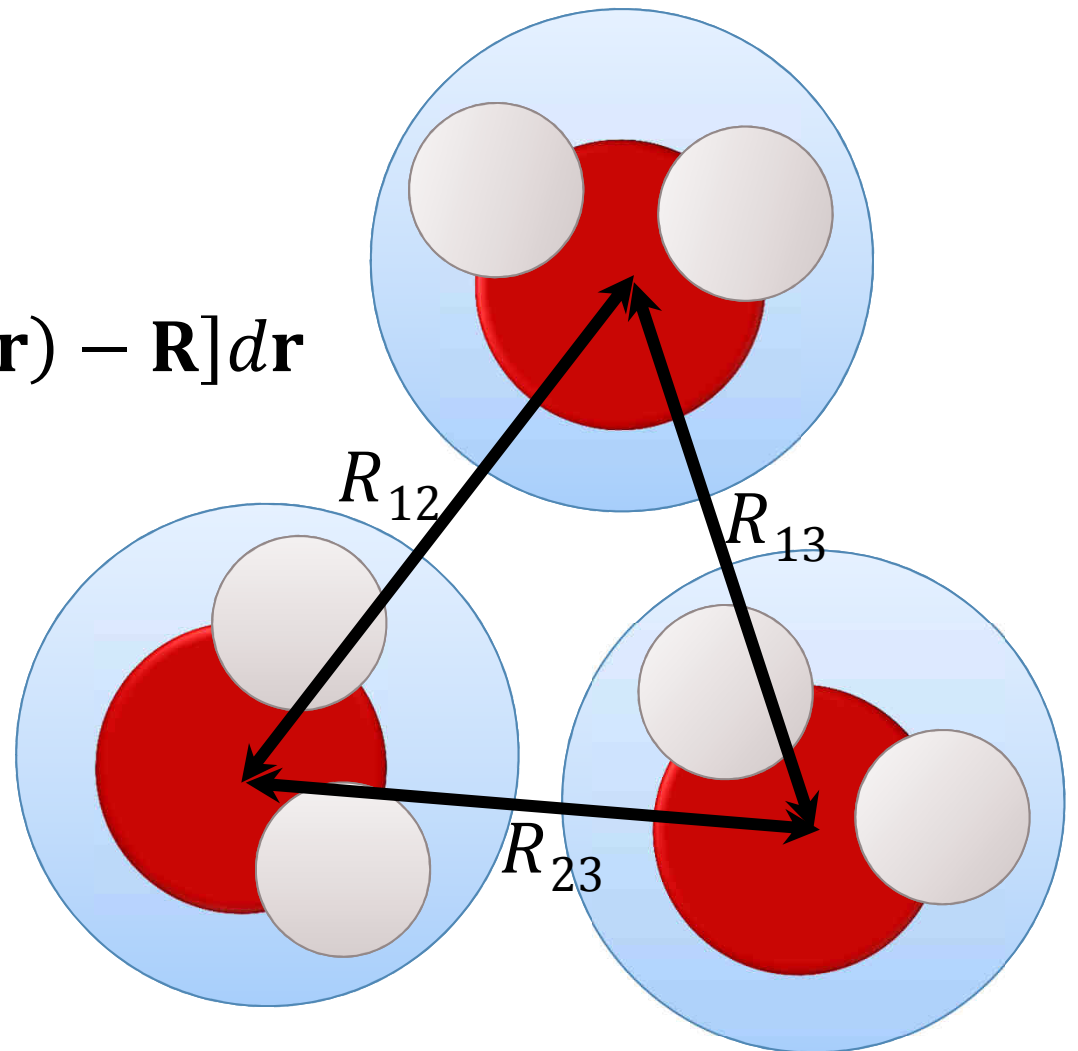
“universal” ways to compare different models

What to match?

$$P_{CG}(\mathbf{R}) = P_{AA}(\mathbf{R})$$

$$e^{-\beta U_{CG}(\mathbf{R})}$$

$$= \int e^{-\beta U_{AA}(\mathbf{r})} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r}$$



What to match?

structure

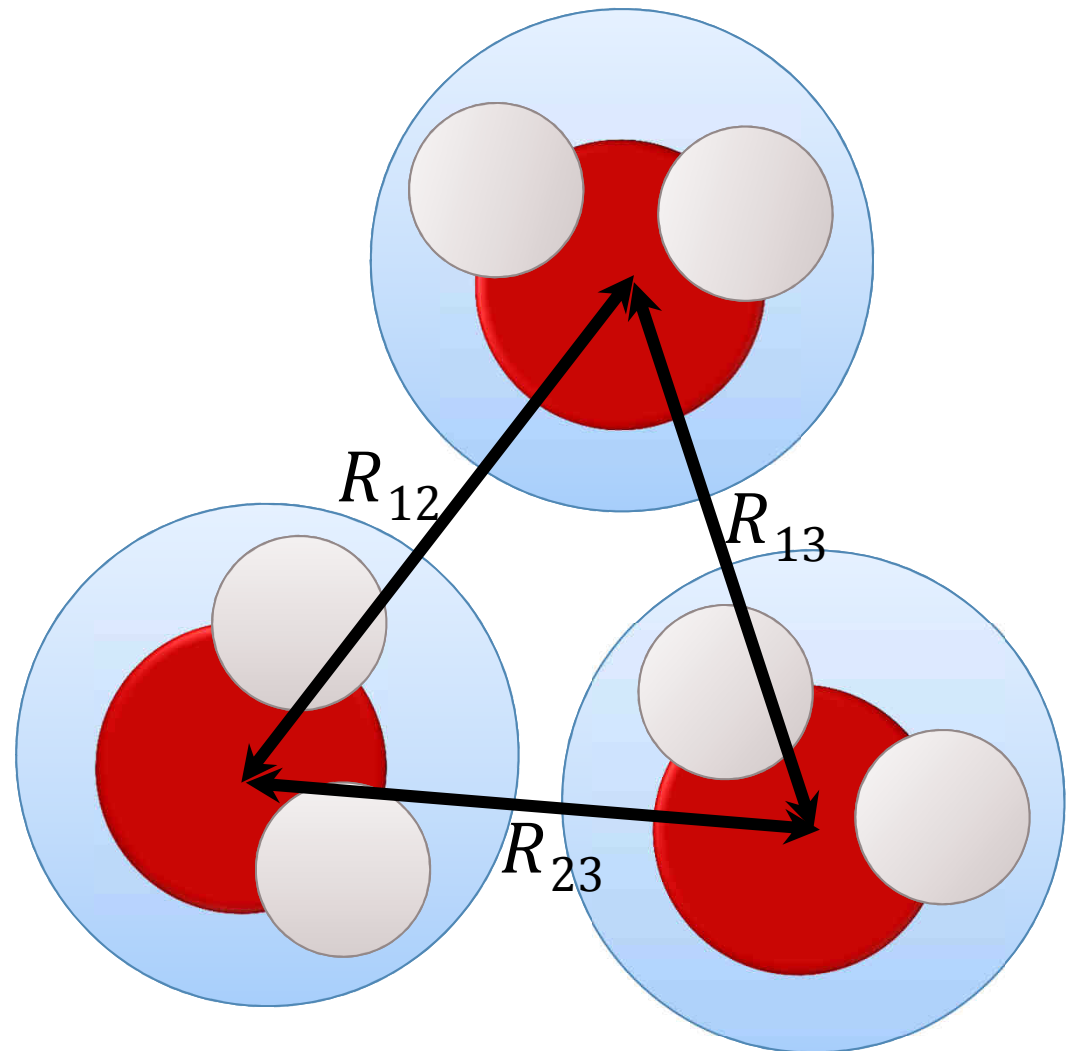
$$g_{AA}(R) = g_{CG}(R)$$

energies

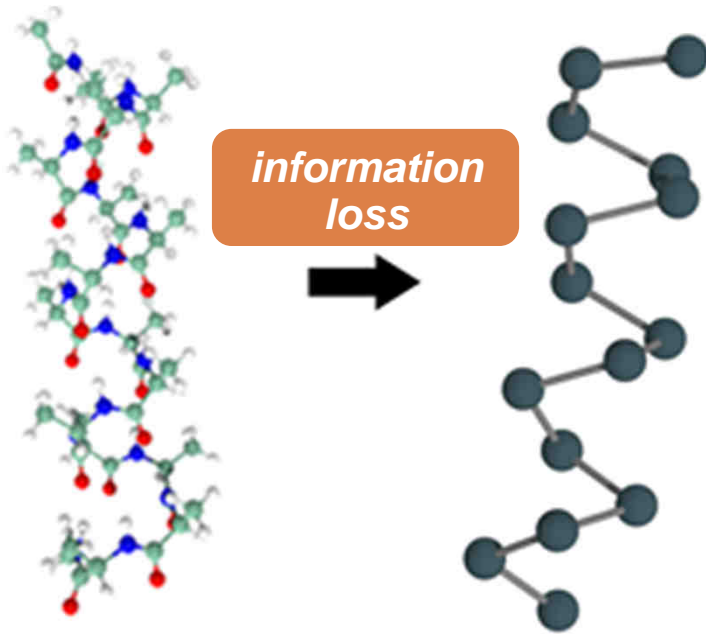
$$\langle U_{AA} \rangle = U_{CG}$$

forces

$$\langle f \rangle_{AA} = f_{CG}$$



$$S_{\text{rel}} = \sum_i p_{AA}(i) \ln \left[\frac{p_{AA}(i)}{p_{CG}(i)} \right] \geq 0$$



$p_{AA}(i)$ *all-atom* ensemble probability for configuration i , determined by U_{AA}

$p_{CG}(i)$ *coarse-grained* ensemble probability for the same configuration i , determined by U_{CG}

$$\begin{aligned}
S_{\text{rel}} &= \int p_{AA}(\mathbf{r}) \ln \left[\frac{p_{AA}(\mathbf{r})}{p_{CG}(\mathbf{r})} \right] d\mathbf{r} \\
&= \int p_{AA}(\mathbf{r}) \ln \left[\frac{p_{AA}(\mathbf{r})}{P_{CG}(\mathbf{M}(\mathbf{r}))} \right] d\mathbf{r} + S_{\text{map}} \\
&= \int P_{AA}(\mathbf{R}) \ln \left[\frac{P_{AA}(\mathbf{R})}{P_{CG}(\mathbf{R})} \right] d\mathbf{R} + S'_{\text{map}} \\
&= S_U + S_{\text{map}}
\end{aligned}$$

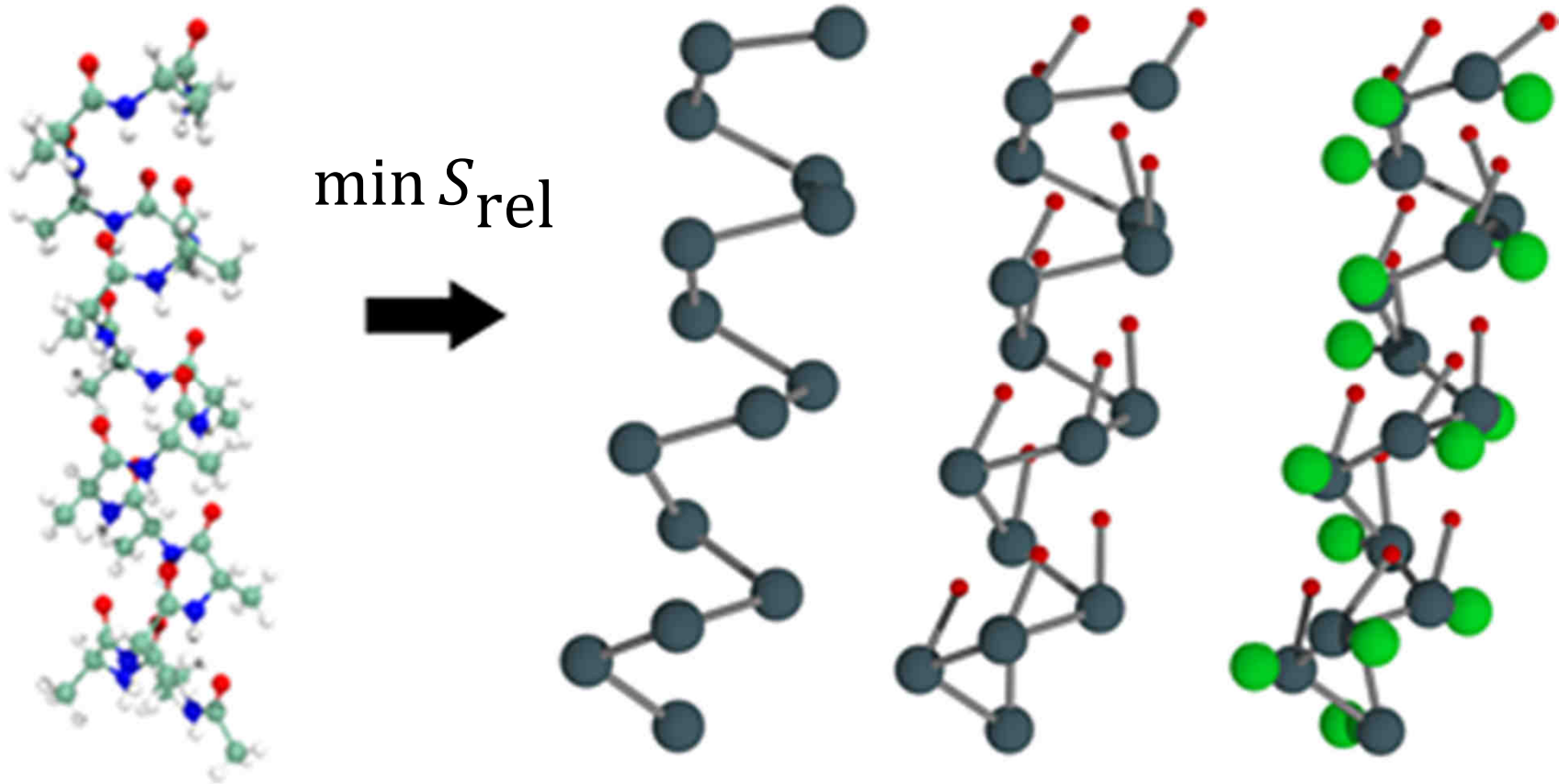
“information loss” due to
approximating interactions
with CG potential

“information loss” due to
DOF reduction; independent
of CG potential

$$p_{AA}(\mathbf{r}) = \frac{e^{-\beta U_{AA}(\mathbf{r})}}{Z_{AA}}$$

$$P_{CG}(\mathbf{R}) = \frac{e^{-\beta U_{CG}(\mathbf{R})}}{Z_{CG}}$$

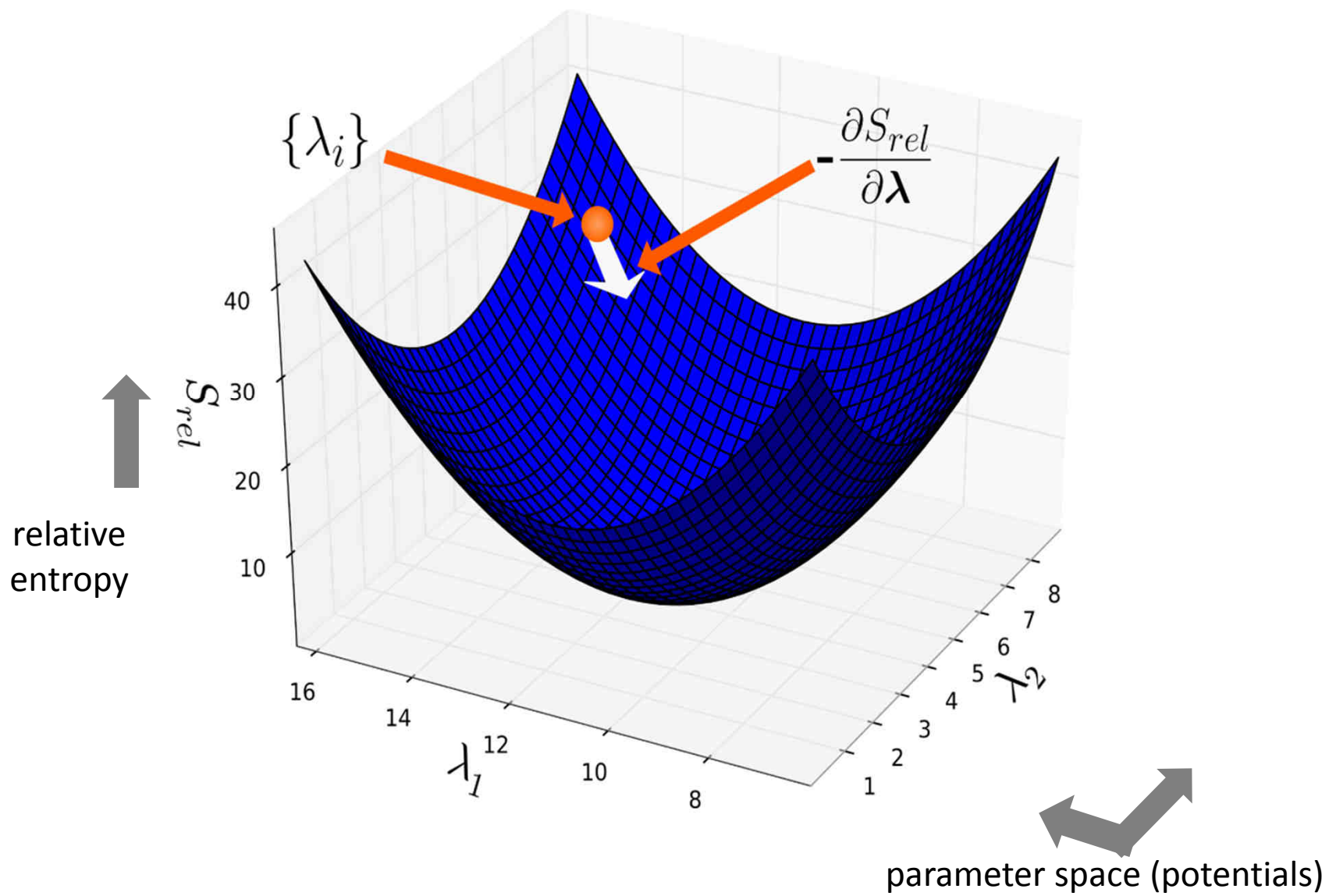
$$S_{\text{rel}} = \beta \langle U_{CG} - U_{AA} \rangle_{AA} - \beta (A_{CG} - A_{AA}) + S_{\text{map}}$$



$$S_{\text{rel}} = \underbrace{\beta \langle U_{CG} - U_{AA} \rangle_{AA} - \beta (A_{CG} - A_{AA})}_{U_{CG}(\mathbf{R}; \lambda_1, \lambda_2, \dots)} + \underbrace{S_{\text{map}}}_{\mathbf{R} = \mathbf{M}(\mathbf{r})}$$

$U_{CG}(\mathbf{R}; \lambda_1, \lambda_2, \dots)$

$\mathbf{R} = \mathbf{M}(\mathbf{r})$



Newton-Raphson iteration to minimize S_{rel}

$$\lambda^{k+1} = \lambda^k - \frac{\left(\frac{\partial S_{\text{rel}}}{\partial \lambda}\right)}{\left(\frac{\partial S_{\text{rel}}^2}{\partial \lambda^2}\right)}$$

from single reference
atomistic simulation
(e.g., saved trajectory)

$$= \lambda^k - \frac{\left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{CG} - \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{AA}}{\left\langle \frac{\partial^2 U_{CG}}{\partial \lambda^2} \right\rangle_{CG} + \beta \left\langle \left(\frac{\partial U_{CG}}{\partial \lambda} \right)^2 \right\rangle_{CG} - \beta \left\langle \left(\frac{\partial U_{CG}}{\partial \lambda} \right)^2 \right\rangle_{CG} - \left\langle \frac{\partial^2 U_{CG}}{\partial \lambda^2} \right\rangle_{AA}}$$

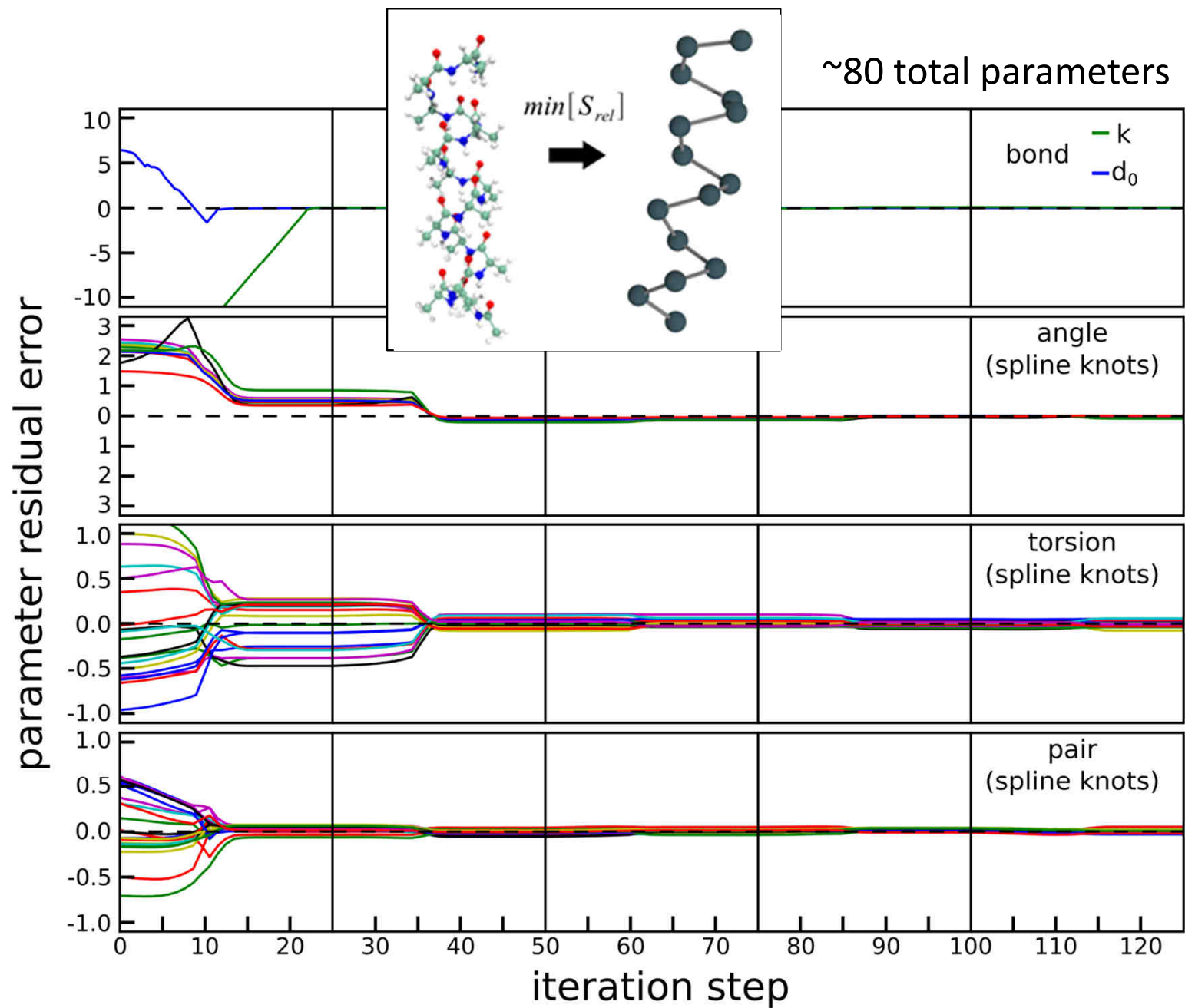
evaluated at each iteration with a trial CG simulation
as force field parameters converge

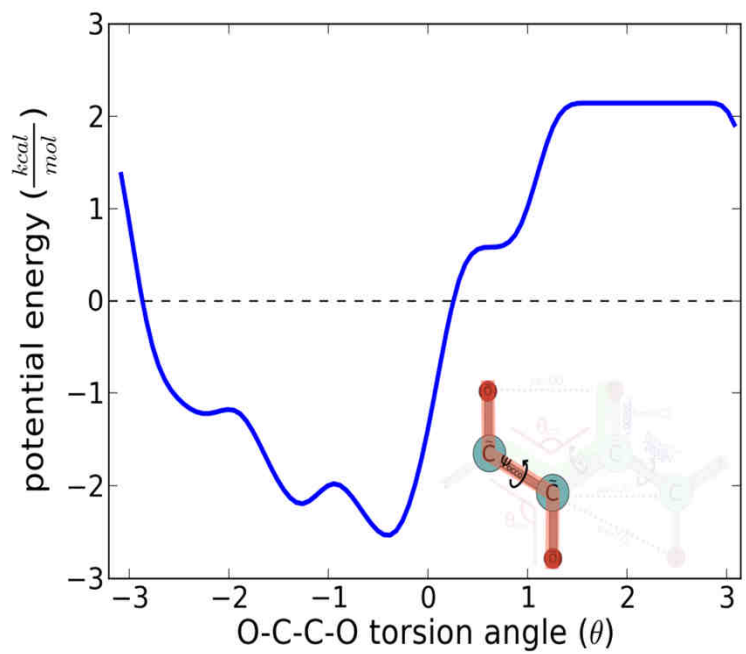
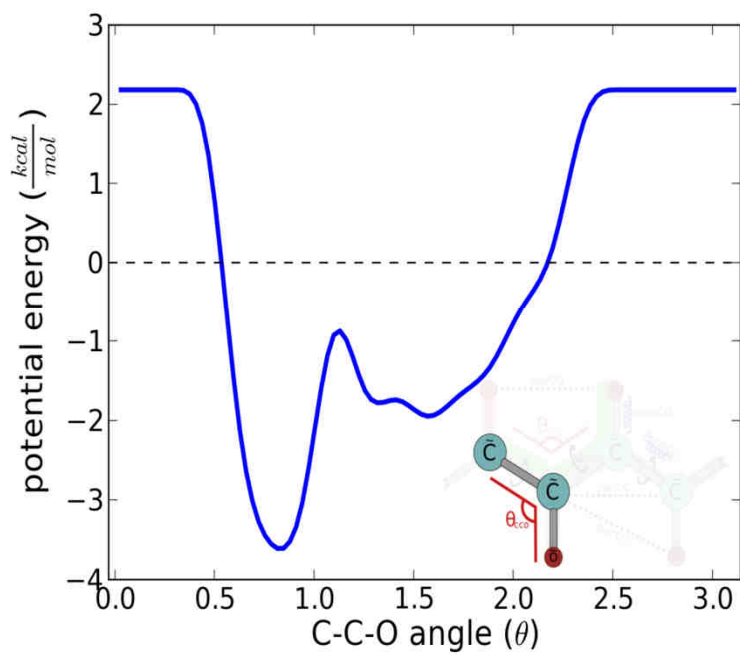
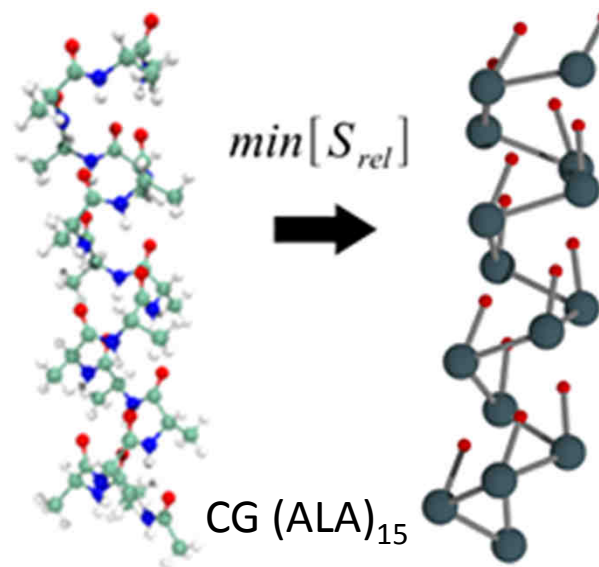
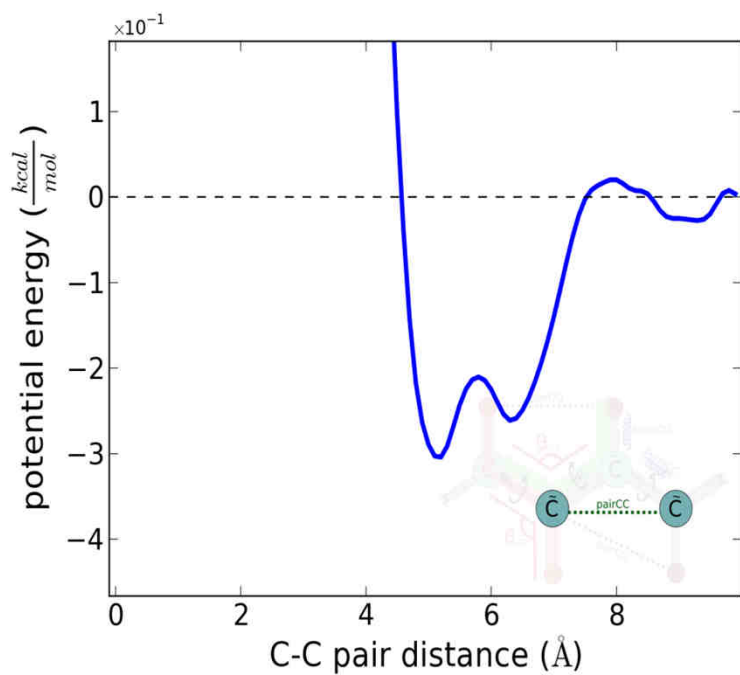
*No need for a new CG simulation at each step.
Reweight old one instead!*

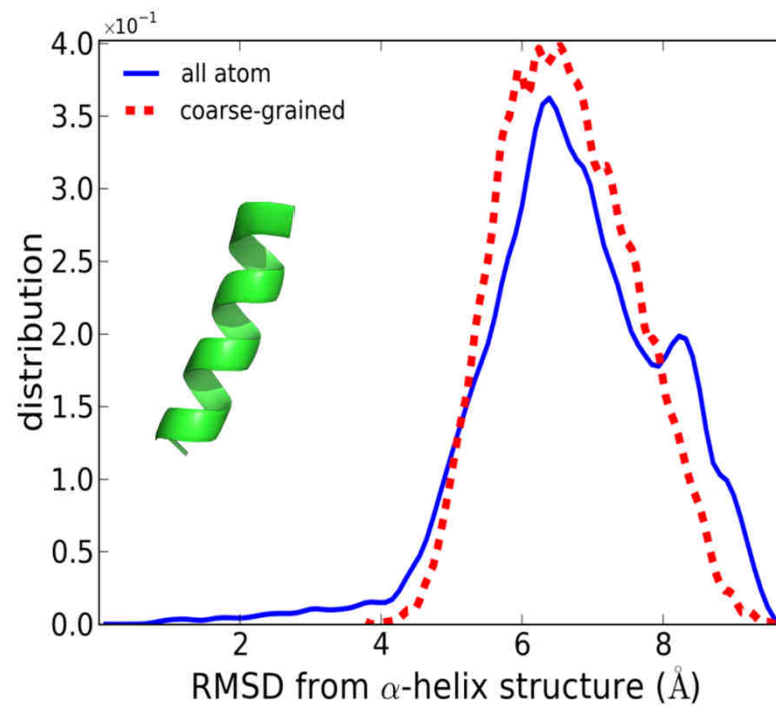
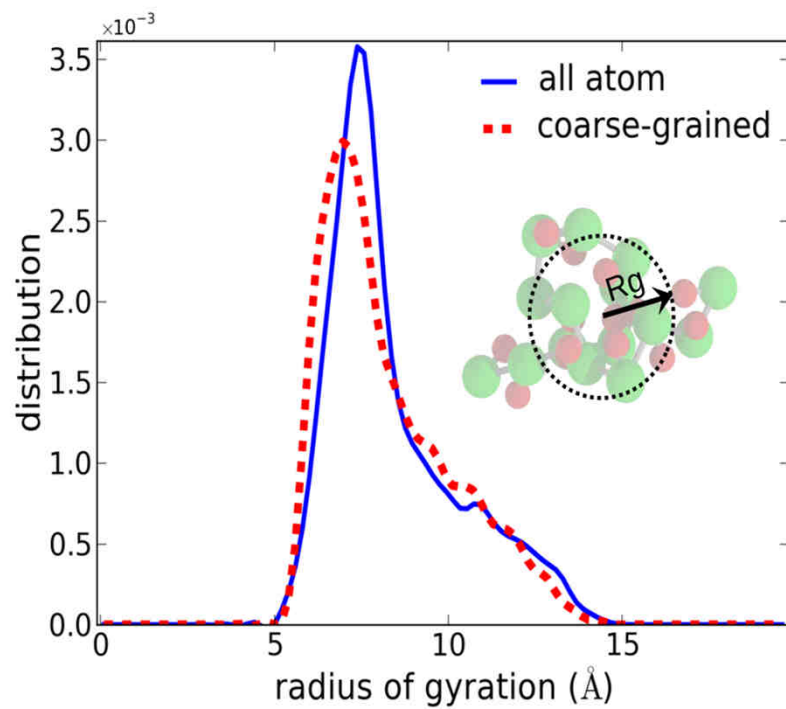
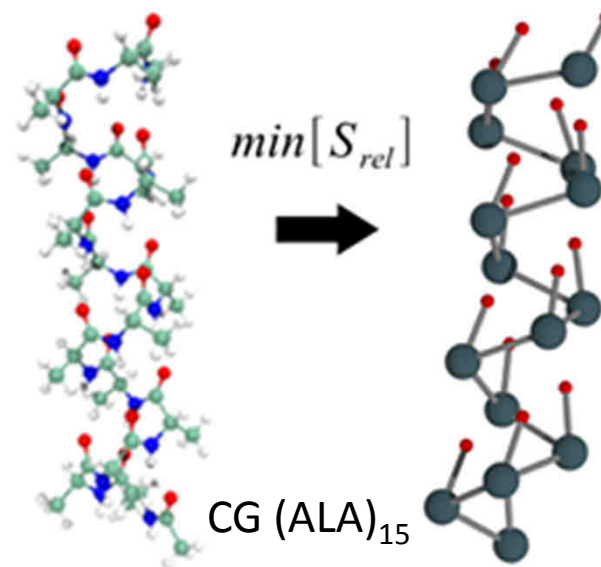
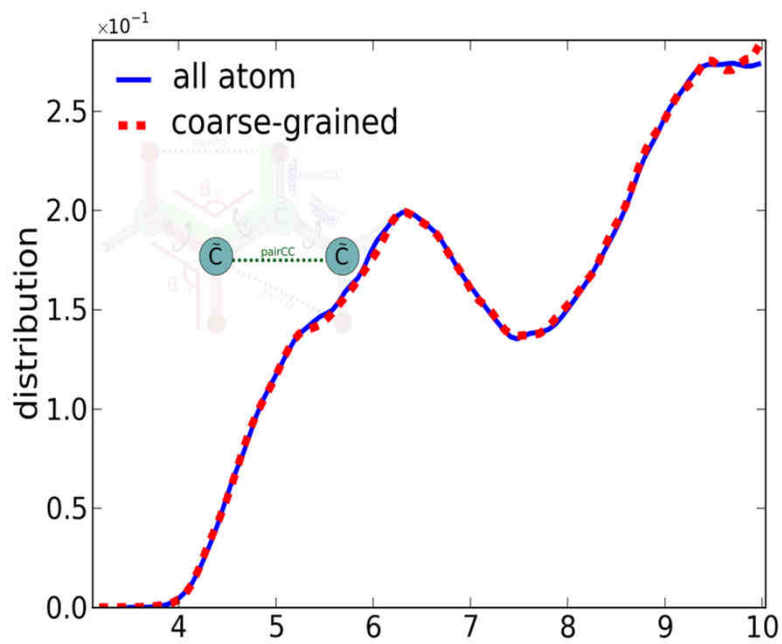
$$\langle X \rangle_{CG,\lambda} = \frac{\langle wX \rangle_{CG,\lambda_0}}{\langle w \rangle_{CG,\lambda_0}}$$

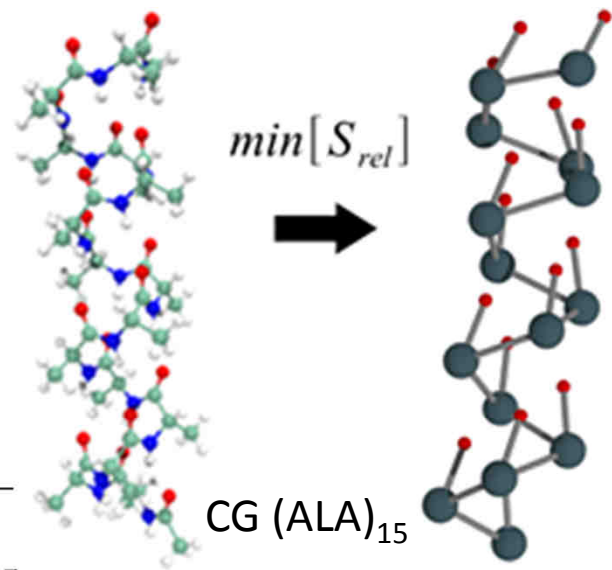
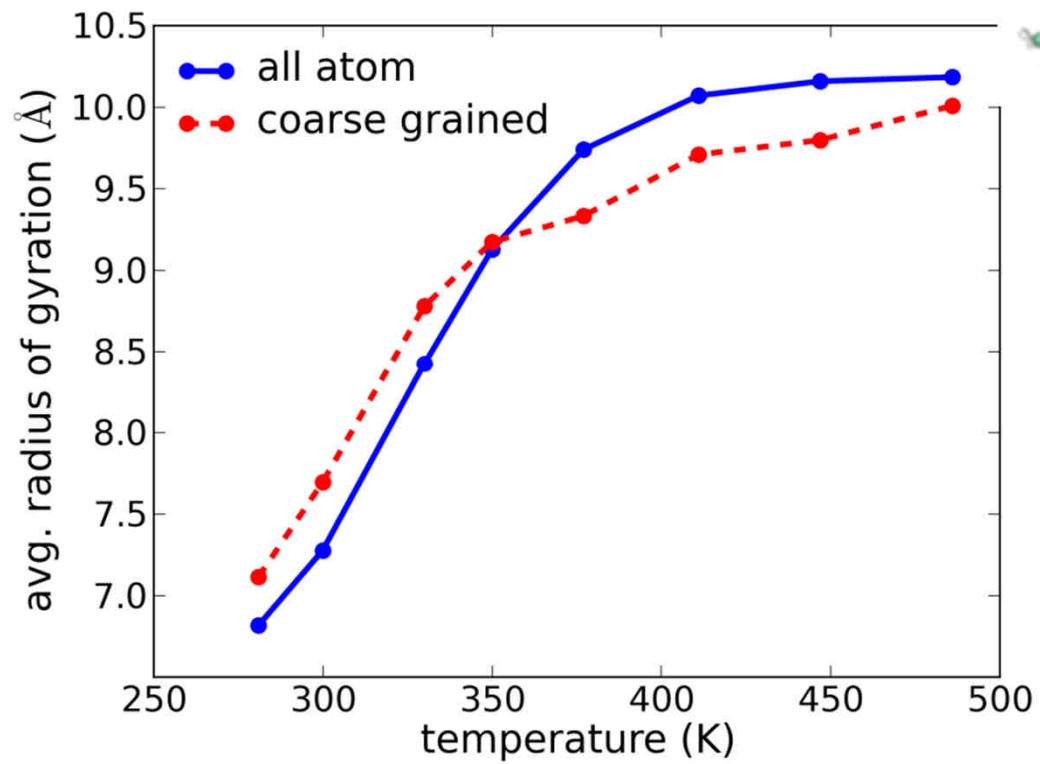
$$w \equiv e^{\beta(U_{CG,\lambda_0} - U_{CG,\lambda})} = e^{\beta\Delta U_{CG}}$$

$$\Delta S_{\text{rel}} = -\beta \langle \Delta U_{CG} \rangle_{AA} + \ln \langle w \rangle_{CG,\lambda_0}$$







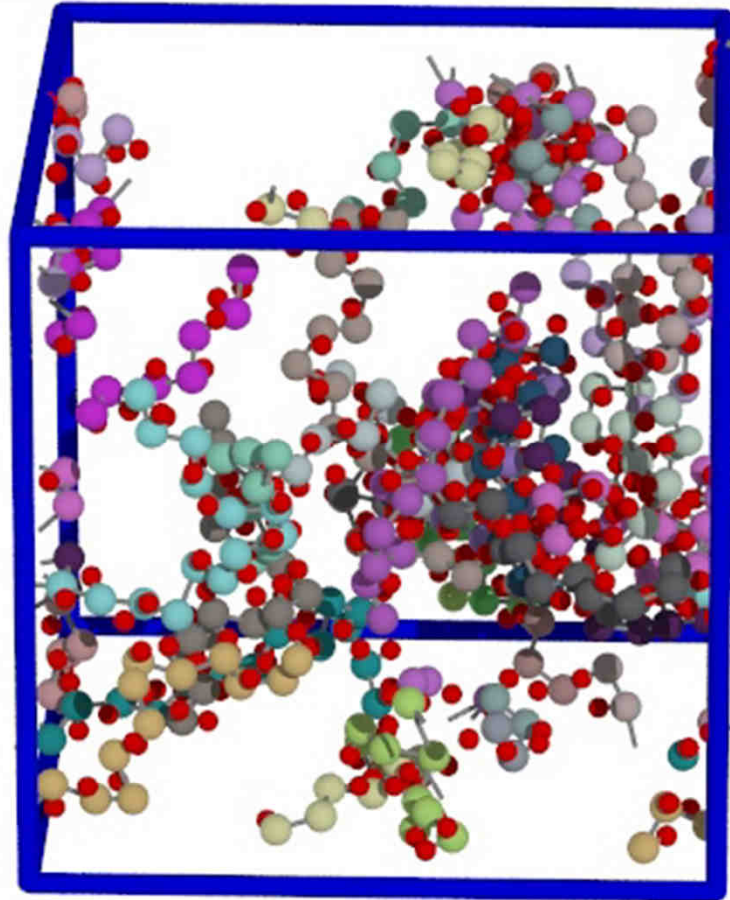


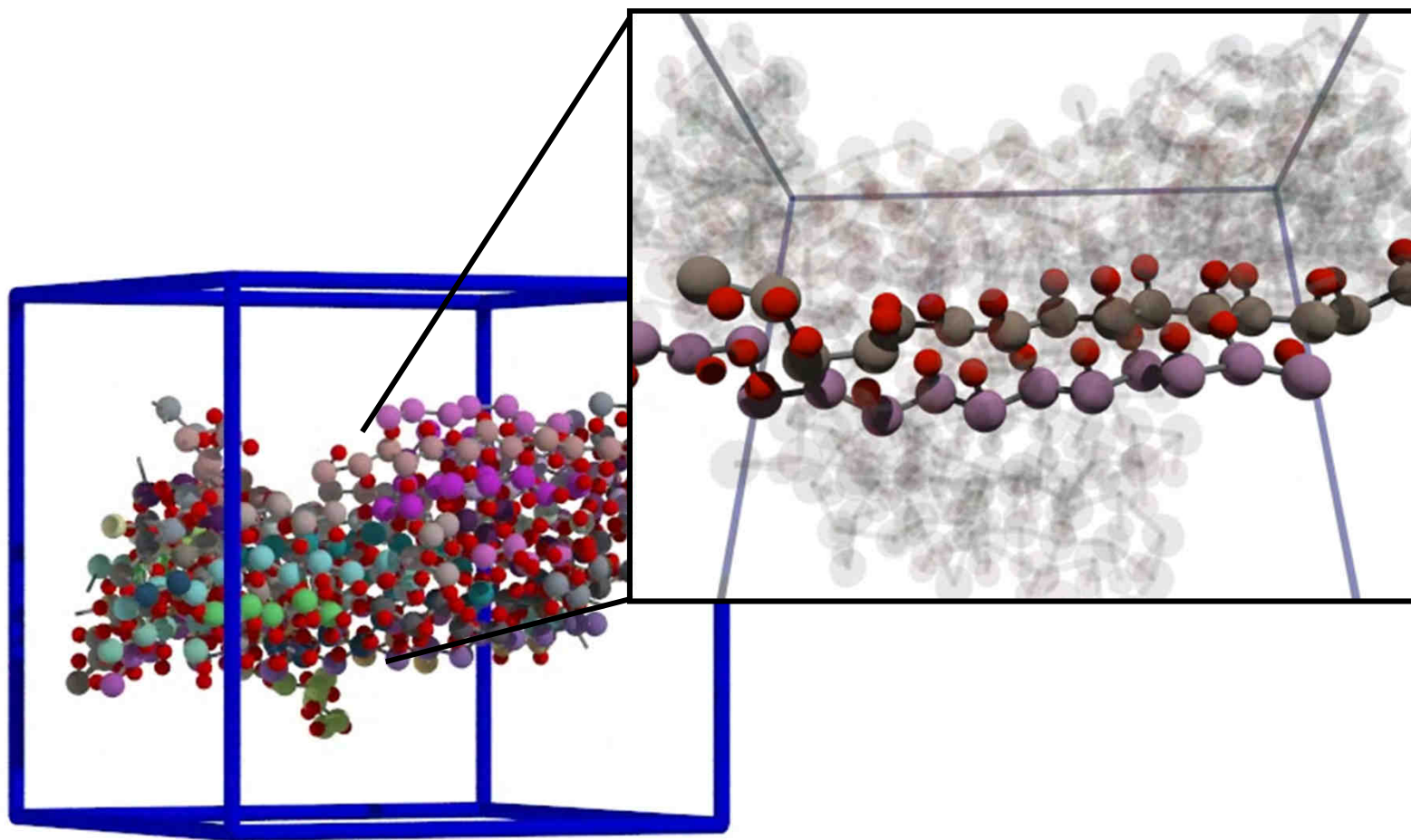
25 chains of $(ALA)_{15}$

number of molecules = 25

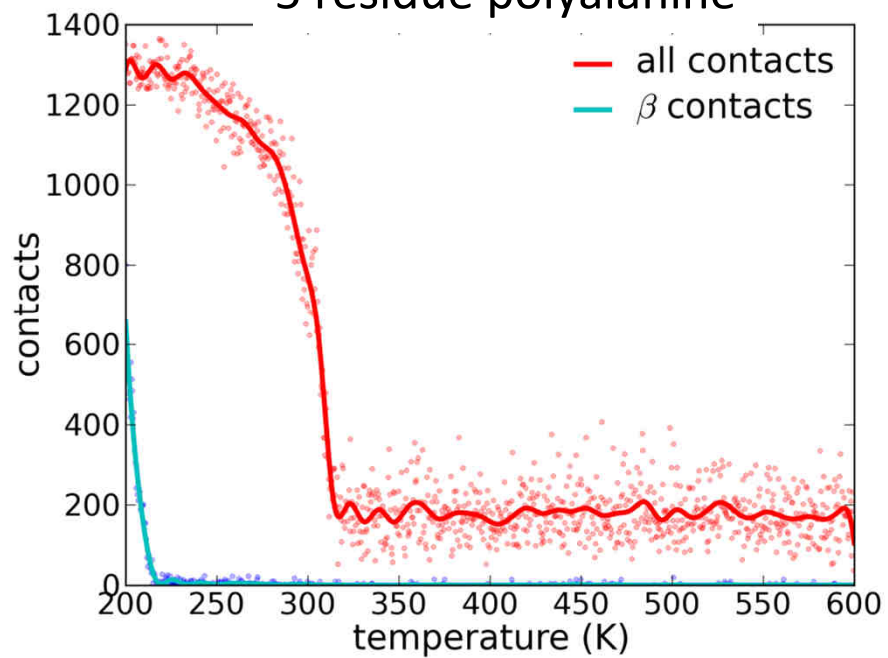
number density = 0.0002

$T = 300K$

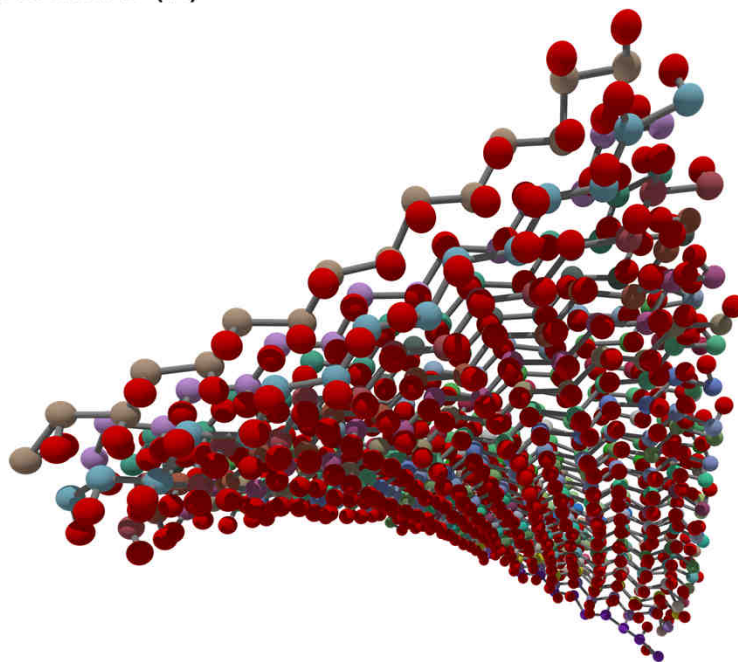
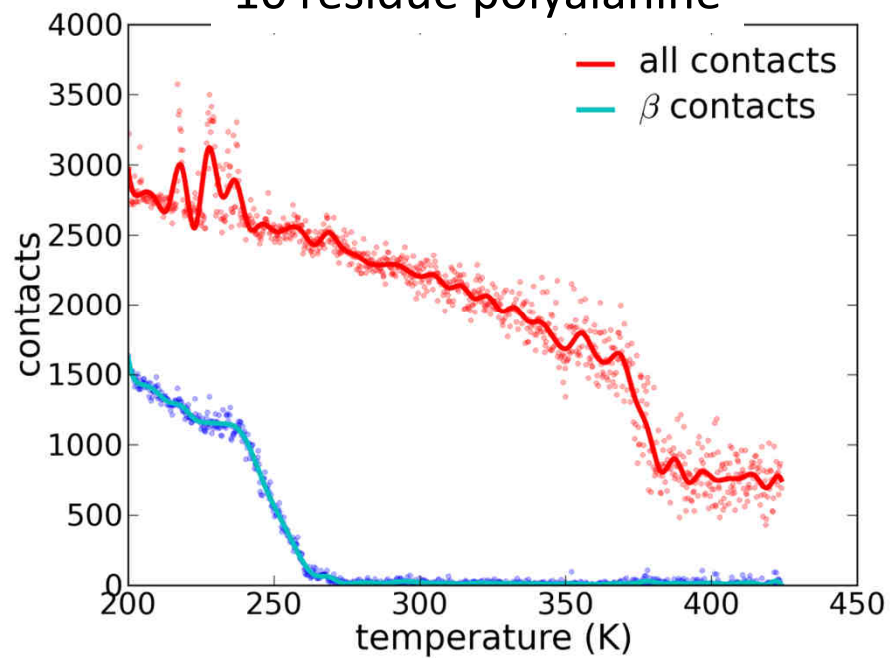


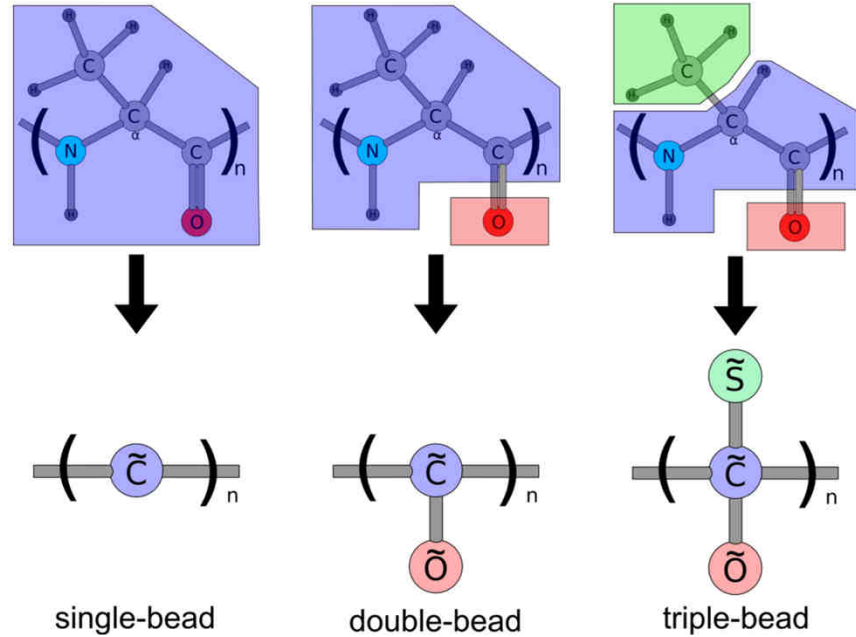


5 residue polyaniline

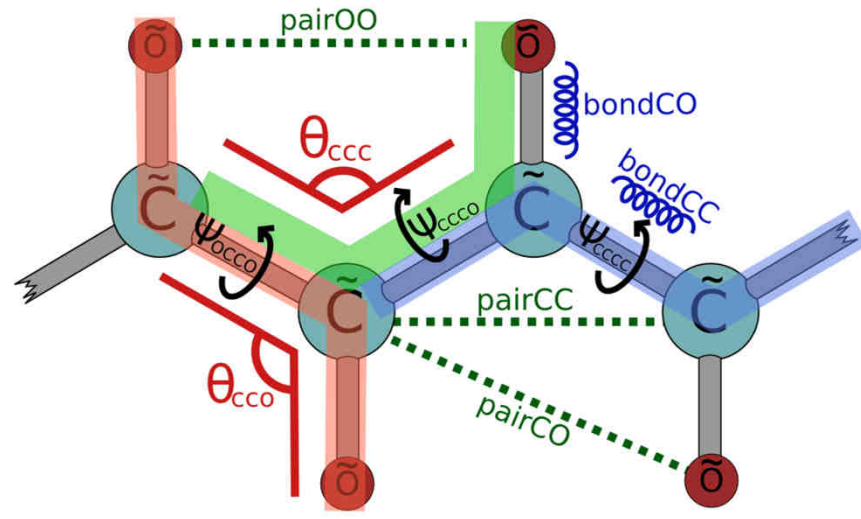


10 residue polyaniline



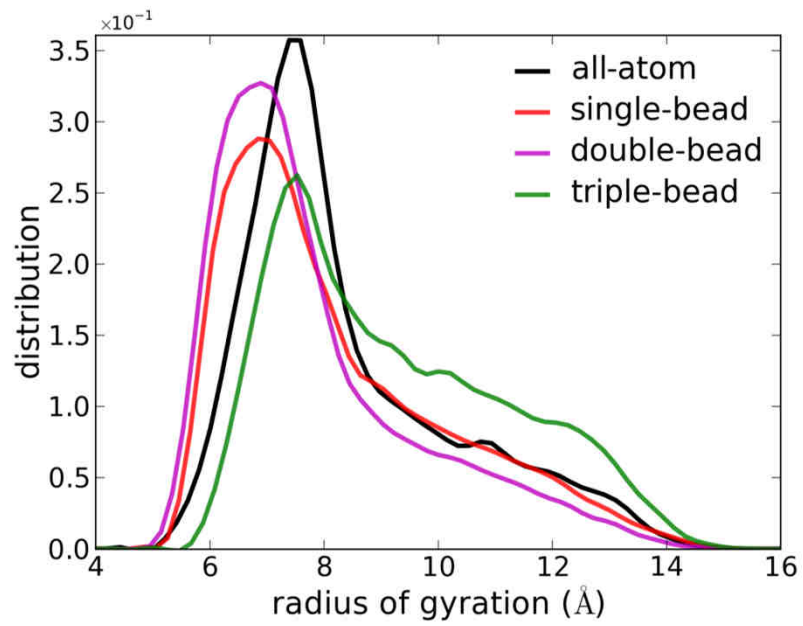


(a)

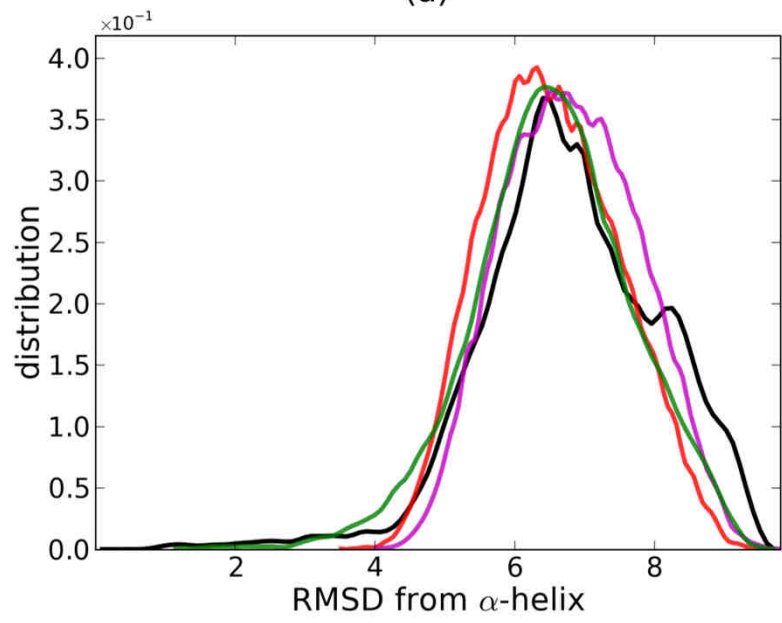


(b)

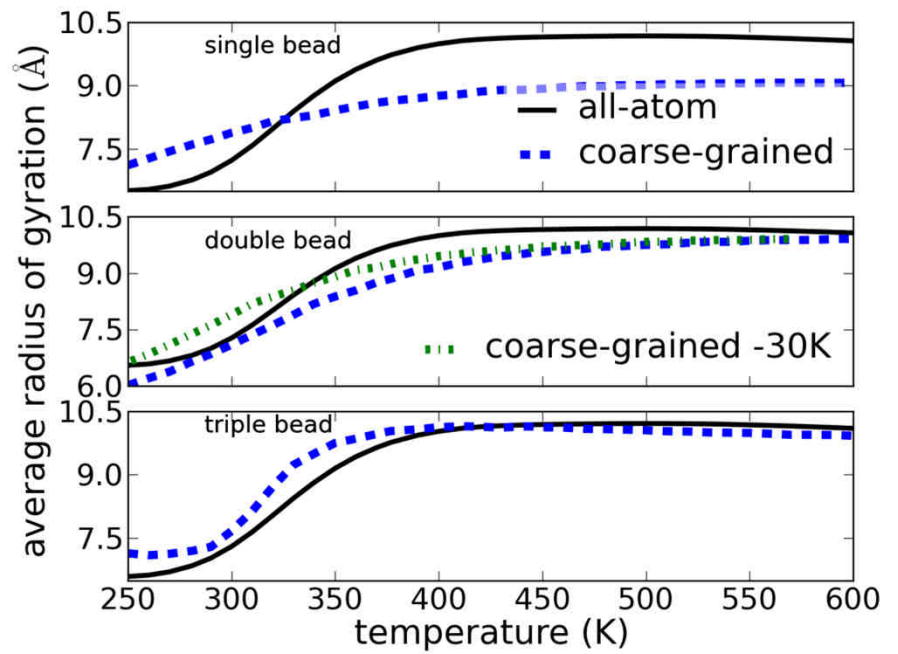
82-444
parameters



(a)



(b)



Does the relative entropy teach us anything new?

What to match?

structure

$$\frac{\delta S_{\text{rel}}}{\delta [u_{CG, \text{pair}}(R)]} = 0 \quad \rightarrow \quad g_{AA}(R) = g_{CG}(R)$$

energies

$$S_{\text{rel}} = \text{var}_{AA}(\beta U_{AA} - \beta U_{CG}) + \mathcal{O}(\beta^3)$$

forces

$$\frac{\delta S_{\text{rel}}}{\delta U_{CG}} = 0 \quad \rightarrow \quad U_{CG} = PMF_{AA} \quad \rightarrow \quad \langle f \rangle_{AA} = f_{CG}$$

see also: Rudzinski and Noid, JCP 135, 214101 (2011)

$TS_{\text{rel}} \rightarrow$ nonequilibrium work associated with coarse-graining

Measures of trajectory ensemble disparity in nonequilibrium statistical dynamics

Gavin E Crooks and David A Sivak

Physical Biosci
Berkeley, CA 94
E-mail: GECroc

Received 15 Apr
Accepted 12 Ma
Published 3 Jun

PRL **108**, 150601 (2012)

PHYSICAL REVIEW LETTERS

week ending
13 APRIL 2012

Near-Equilibrium Measurements of Nonequilibrium Free Energy

David A. Sivak and Gavin E. Crooks

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(Received 14 December 2009; revised manuscript received 5 January 2012; published 9 April 2012)

A central endeavor of thermodynamics is the measurement of free energy changes. Regrettably, although we can measure the free energy of a system in thermodynamic equilibrium, typically all we can say about the free energy of a nonequilibrium ensemble is that it is larger than that of the same system at equilibrium. Herein, we derive a formally exact expression for the probability distribution of a driven system, which involves path ensemble averages of the work over trajectories of the time-reversed system. From this we find a simple near-equilibrium approximation for the free energy in terms of an excess mean time-reversed work, which can be experimentally measured on real systems. With analysis and computer simulation, we demonstrate the accuracy of our approximations for several simple models.

DOI: [10.1103/PhysRevLett.108.150601](https://doi.org/10.1103/PhysRevLett.108.150601)

PACS numbers: 05.70.Ln, 05.40.-a, 89.70.Cf

$$S_2 \sim \int dR_1 dR_2 P_{ref}(R_1, R_2, \Delta t) \ln \left[\frac{P_{ref}(R_1, R_2, \Delta t)}{P_{model}(R_1, R_2, \Delta t)} \right]$$

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 10538–10545

www.rsc.org/pccp

PAPER

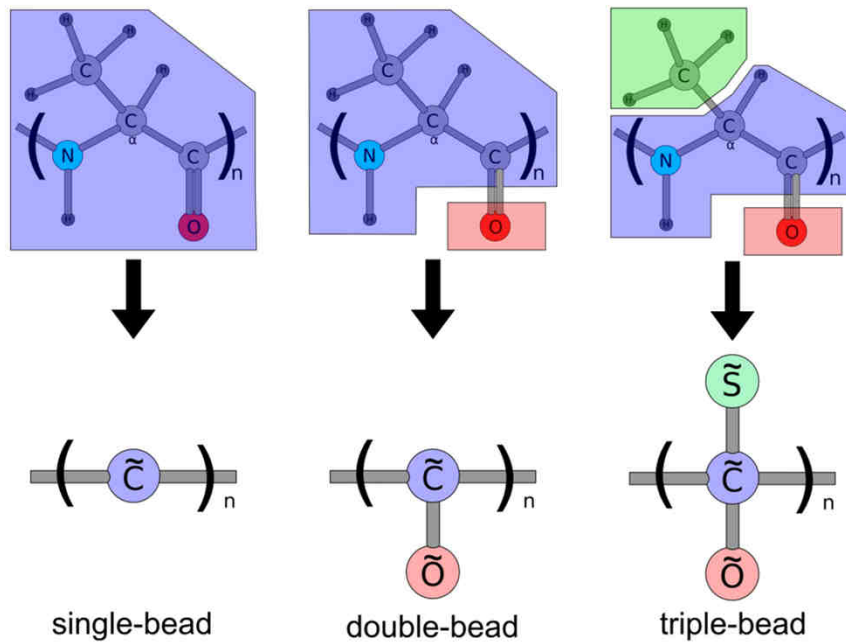
Obtaining fully dynamic coarse-grained models from MD

Pep Español^{*ab} and Ignacio Zúñiga^a

Received 9th December 2010, Accepted 1st March 2011

DOI: 10.1039/c0cp02826f

We present a general method to obtain parametrised models for the drift and diffusion terms of the Fokker–Planck equation of a coarse-grained description of molecular systems. The method is based on the minimisation of the relative entropy defined in terms of the two-time joint probability and thus captures the full dynamics of the coarse-grained description. In addition, we show an alternative Bayesian argument that starts from the path probability of a diffusion process which allows one to obtain the best parametrised model that fits an actual observed path of the coarse-grained variables. Both approaches lead to exactly the same optimisation function giving strong support to the methodology. We provide an heuristic argument that explains how both approaches are connected.



$$S_{\text{rel}} = S_U + S_{\text{map}}$$

“information loss” due to approximating interactions with CG potential

“information loss” due to DOF reduction; independent of CG potential

$$\sim - \int P_{AA}(\mathbf{R}) \ln P_{AA}(\mathbf{R}) d\mathbf{R}$$

Constraints – Lagrange multipliers

$$S_{\text{rel}} - \alpha(\langle X \rangle_{AA} - \langle X \rangle_{CG})$$

Controlling errors

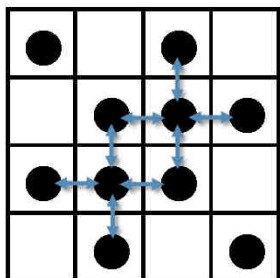
$$U_{CG}(\mathbf{R}) = \lambda X(\mathbf{R}) + \dots$$

$$\rightarrow \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{CG} = \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{AA} \rightarrow \langle X \rangle_{CG} = \langle X \rangle_{AA}$$

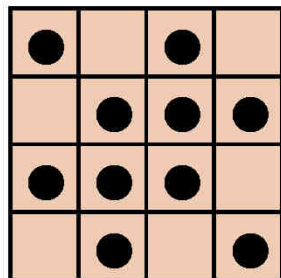
Predicting errors

$$\langle X \rangle_{CG} - \langle X \rangle_{AA} \propto S_{\text{rel}}$$

atomistic
2D lattice gas, pairwise



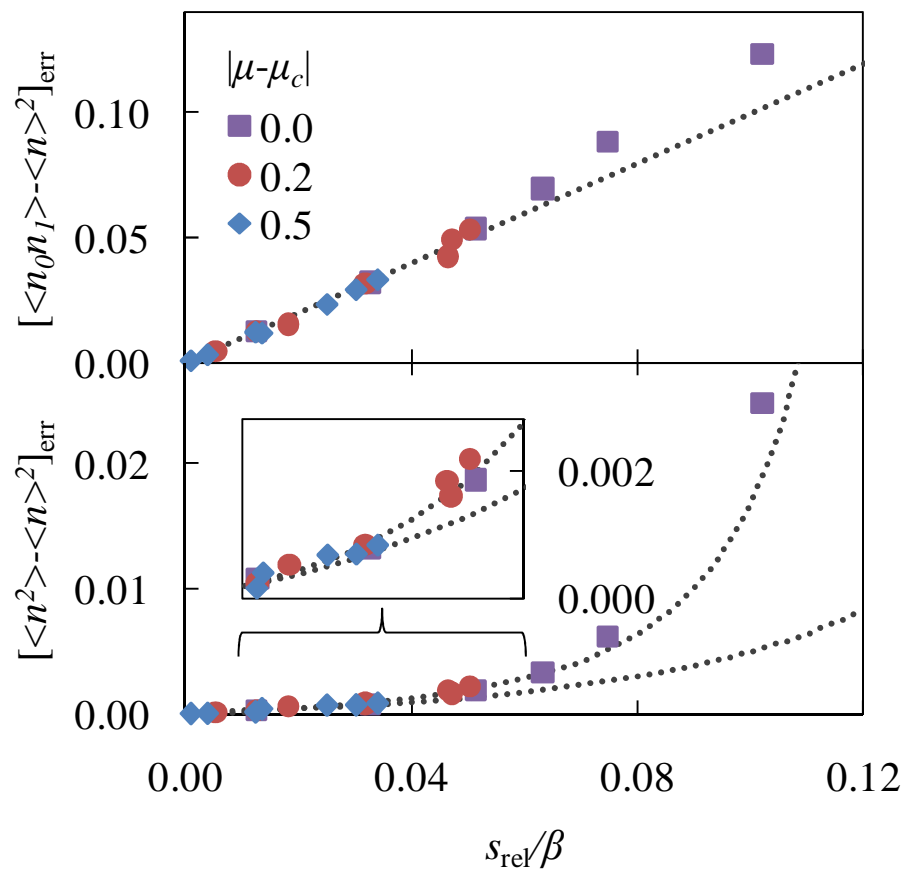
coarse-grained
2D lattice gas, mean-field



nearest-neighbor

bulk

errors in particle fluctuations

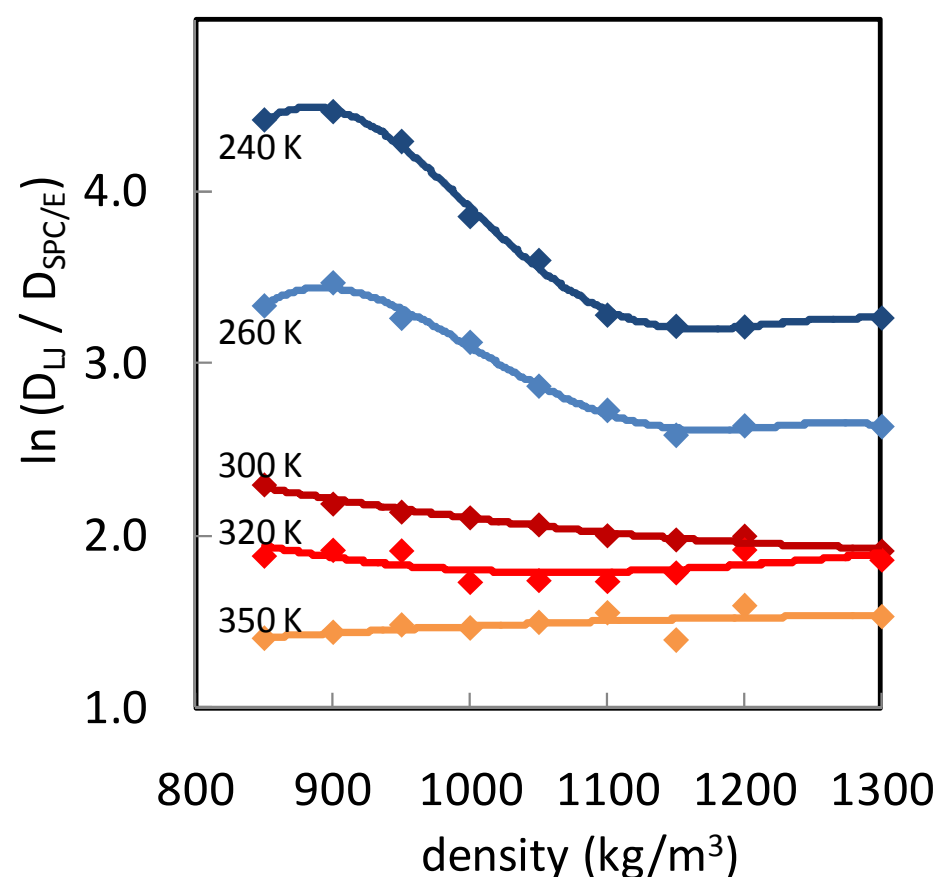
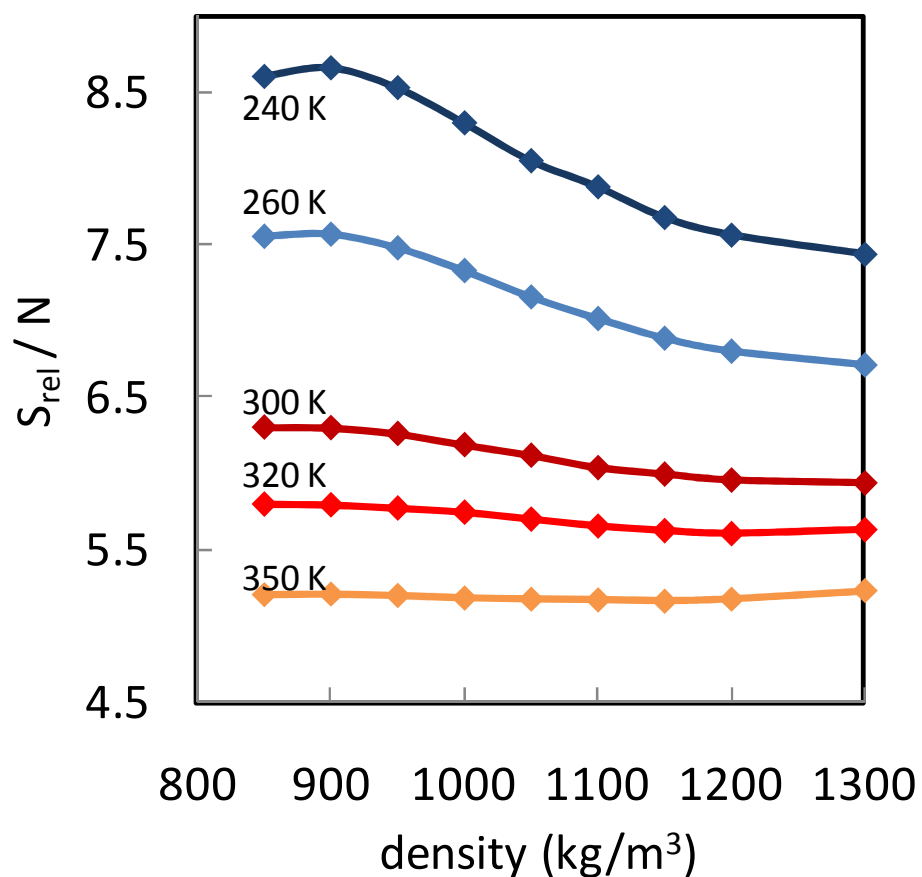




atomistic
SPC/E water

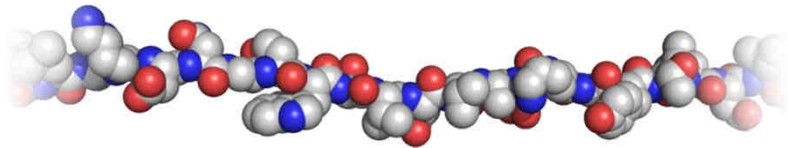


coarse-grained
Lennard-Jones sphere



Protein structure prediction

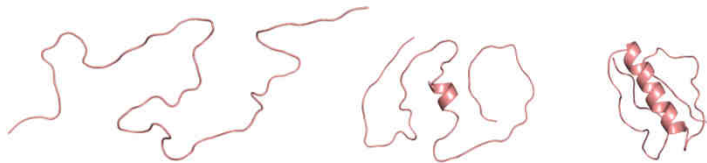
AAGHWWKGPVGEWTLMTYVAVWKHI
unknown sequence



atomically detailed representation

Protein Data Bank

database of known structures



simulated conformational folding
process guided by atomic energetics

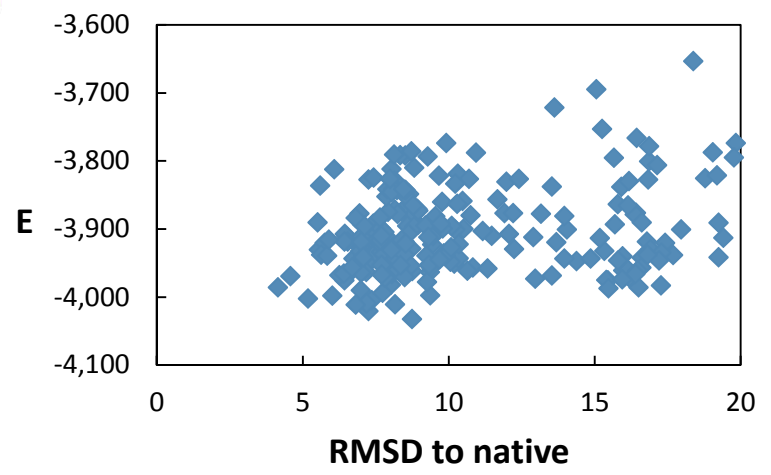
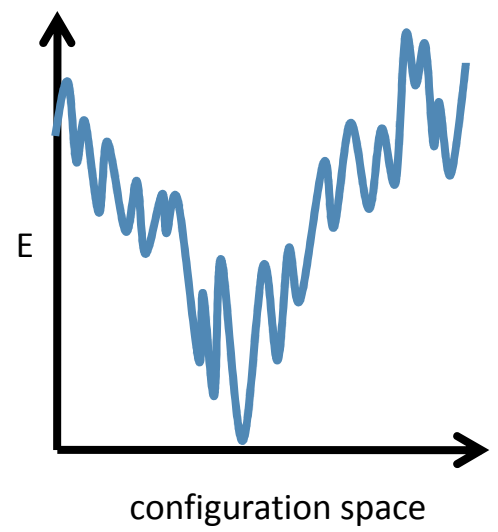
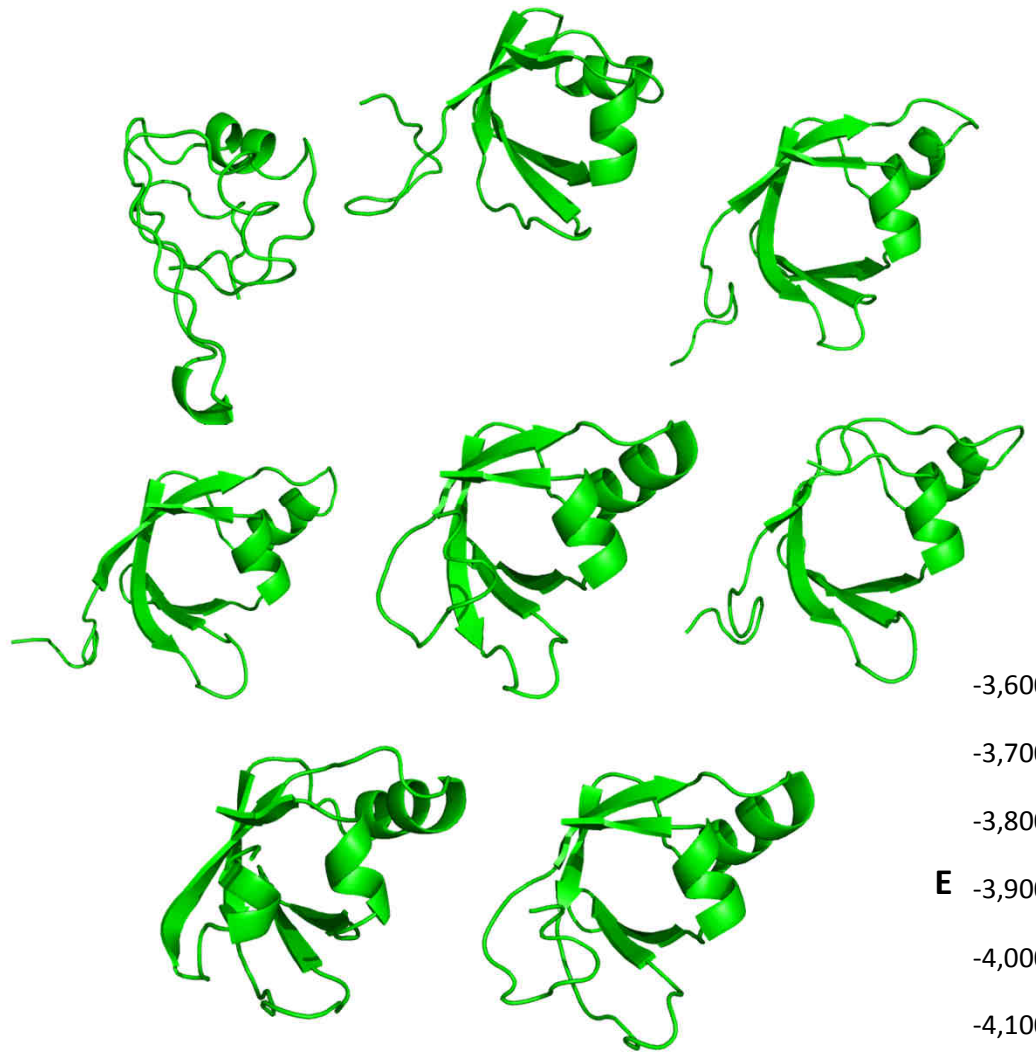
AIGHWWLRGPGAEWTLCTYVAHI
LAGHWFSGGVGEWTIMTYAAWLVEHI
AGHWWKAAVGEYITYEKVADAVWHI

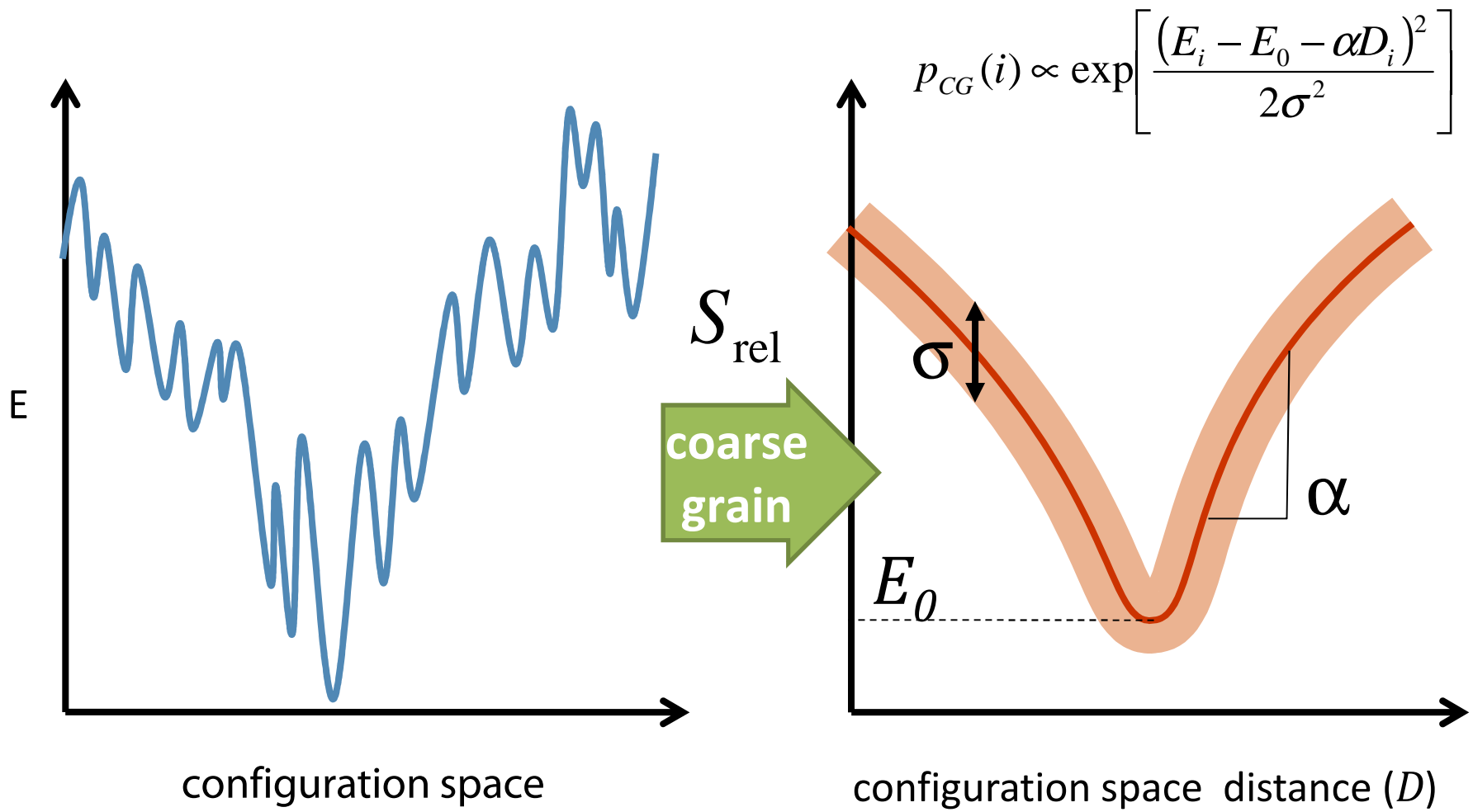
similar sequences and their structures

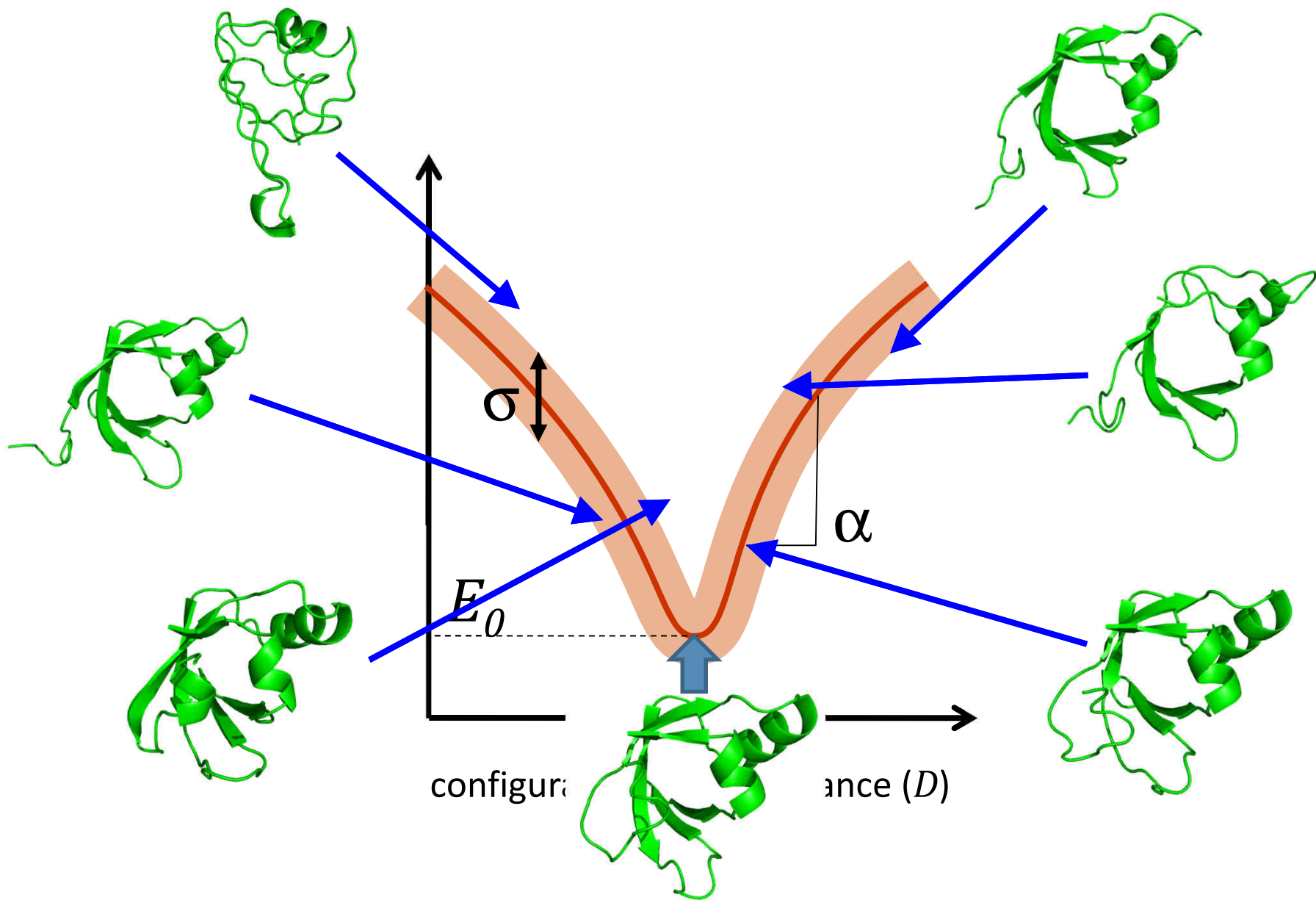


predicted structure

bioinformatics-produced
candidate structures

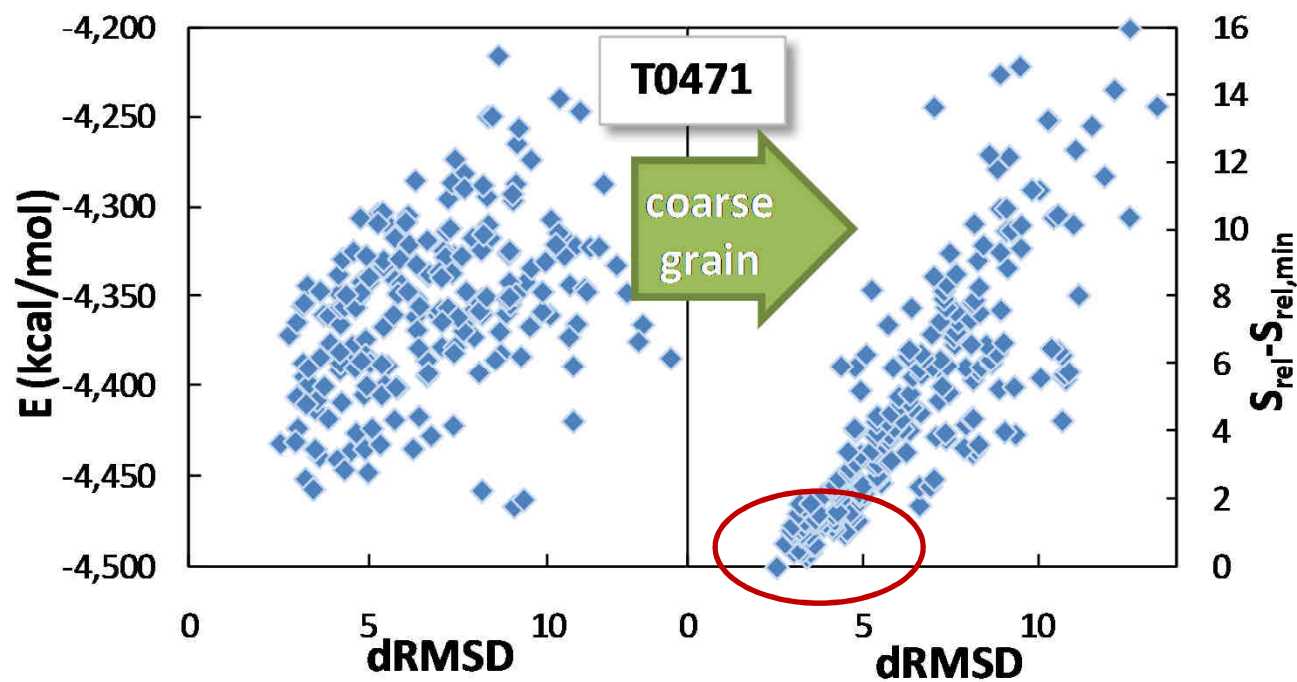


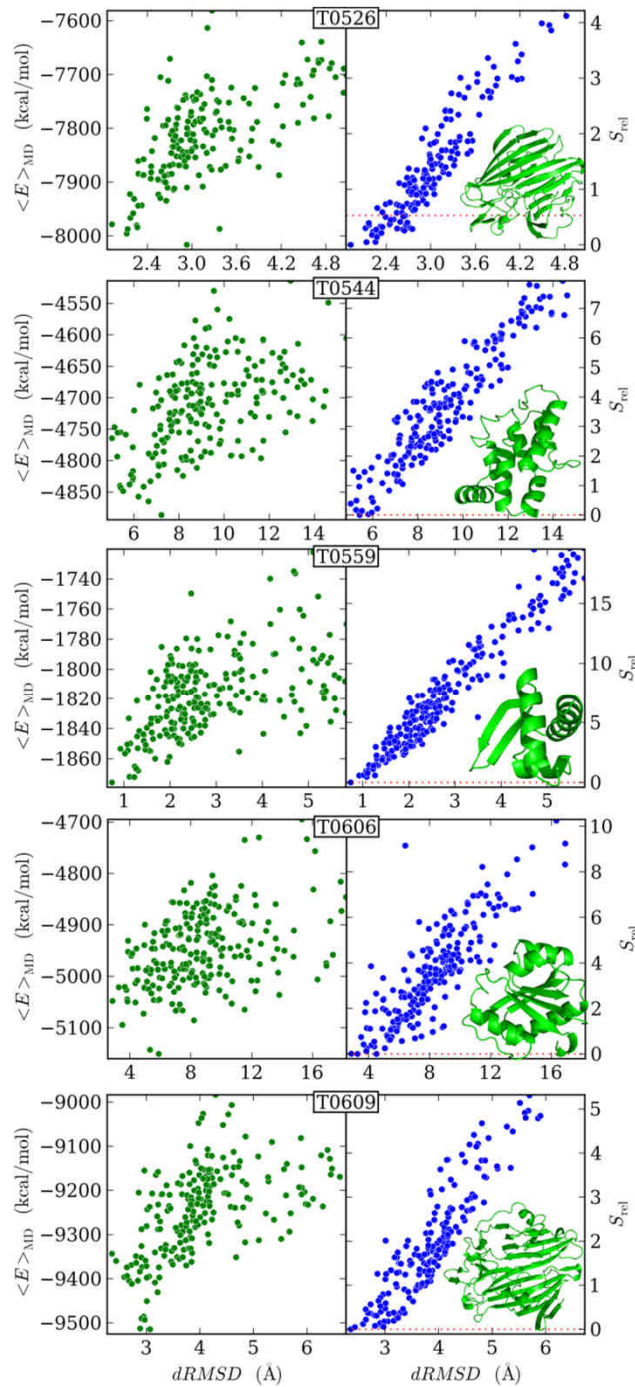




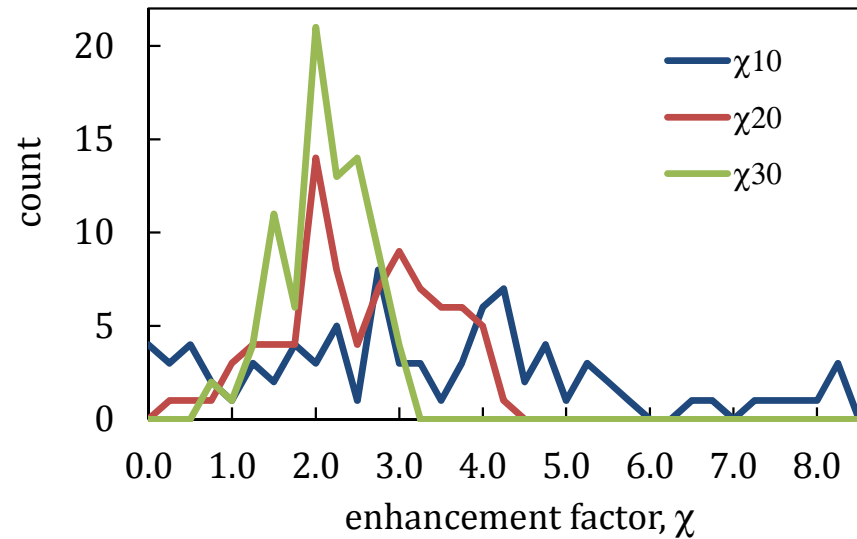


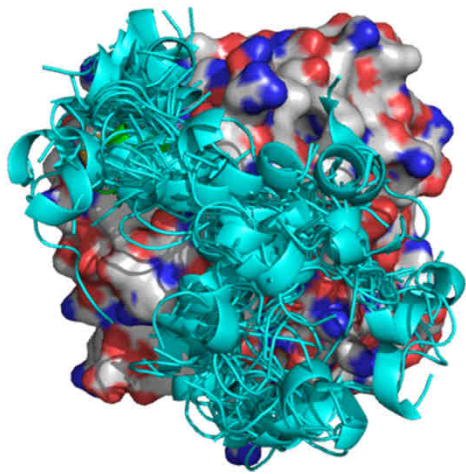
CASP target T0471



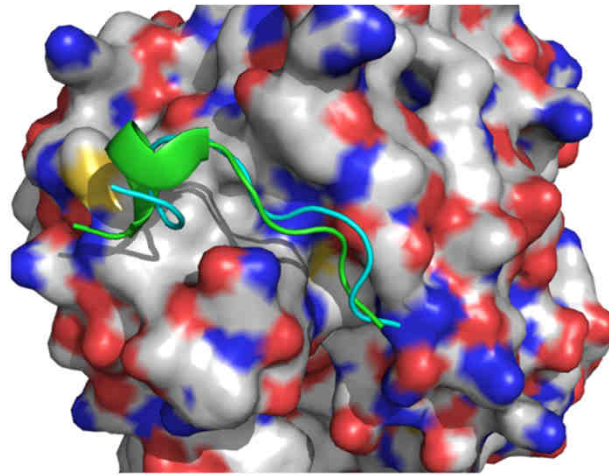


Distribution of enhancement over random selection for 86 proteins

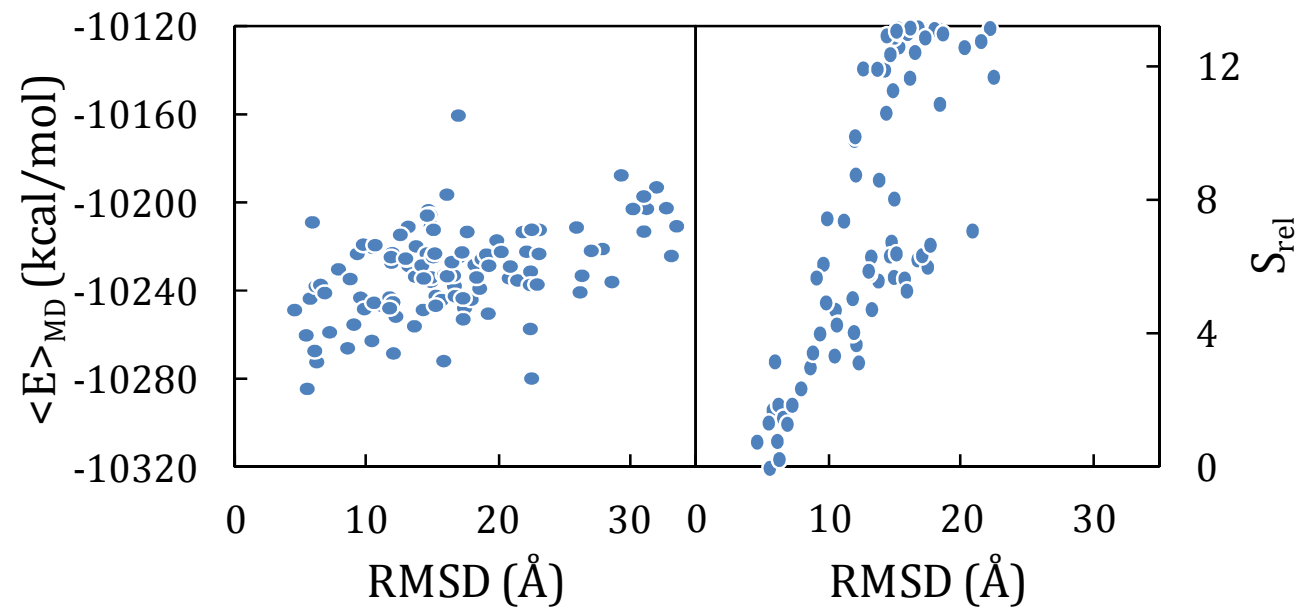




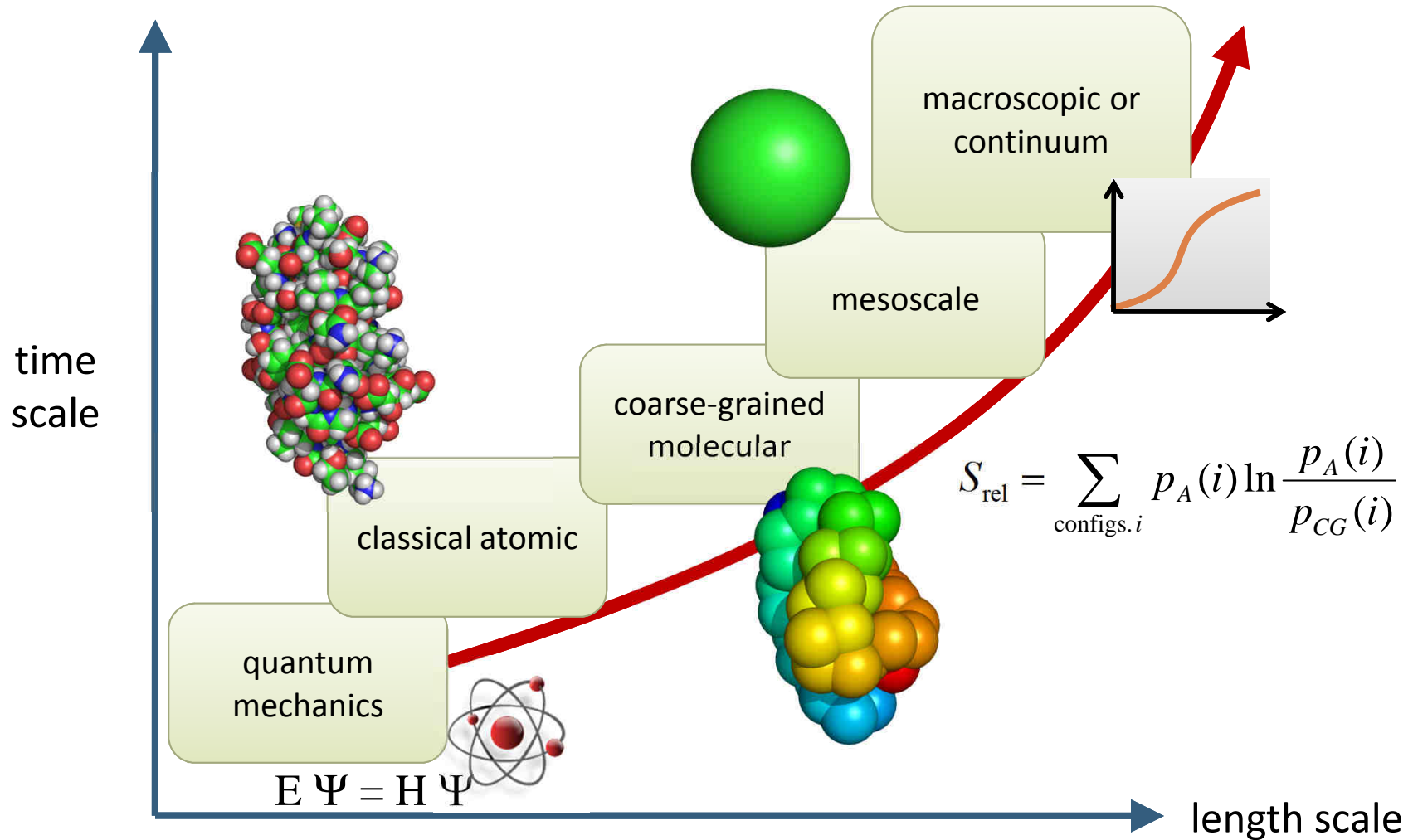
all docked
peptide poses



minimum S_{rel}
(2nd best pose)



Conclusions



The relative entropy provides a systematic and flexible strategy for moving to coarse-grained models and large-scale behavior.