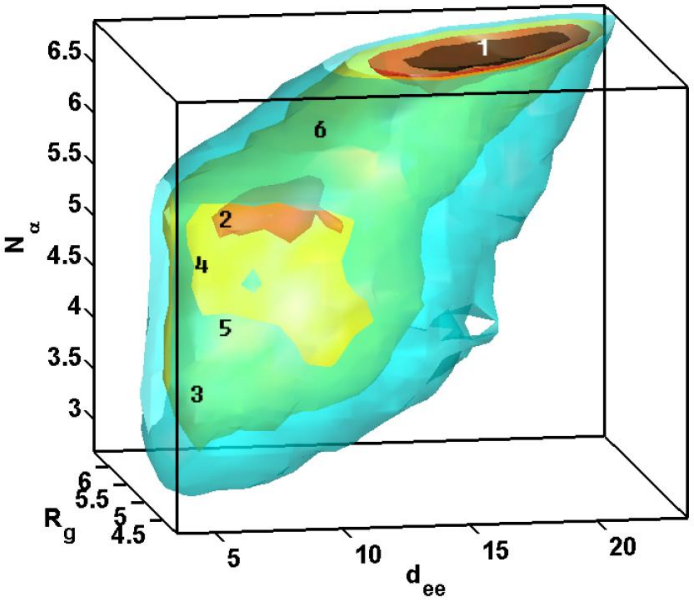
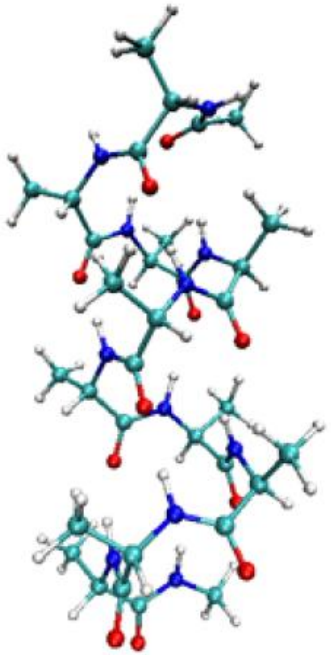
Exploring the free energy landscape of complex molecules and crystalline polymorphs

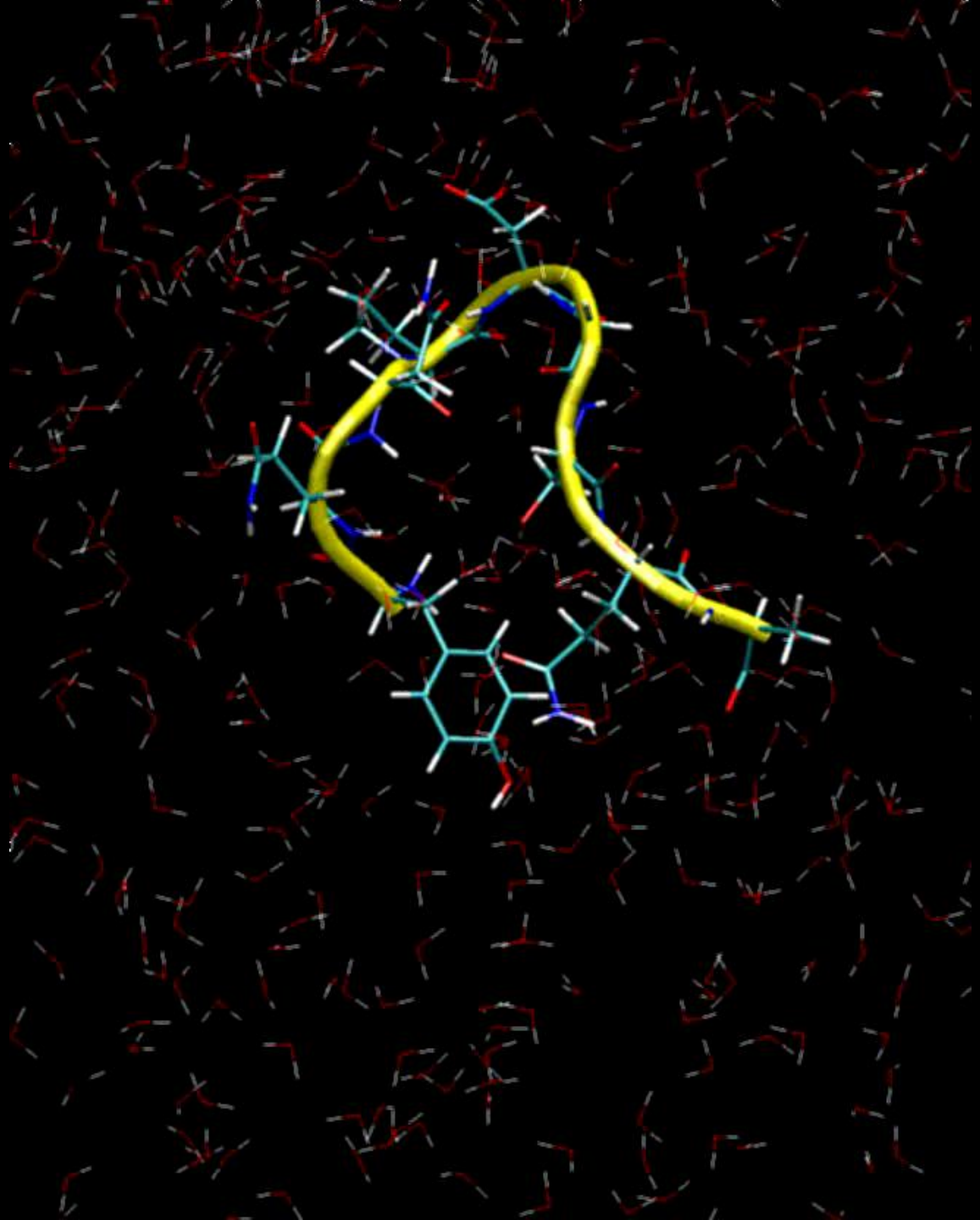




Mark E. Tuckerman
Dept. of Chemistry and Courant Institute of Mathematical Sciences
New York University, 100 Washington Square East, NY 10003

NSF
The Camille & Henry Dreyfus Foundation, Inc.
New York University

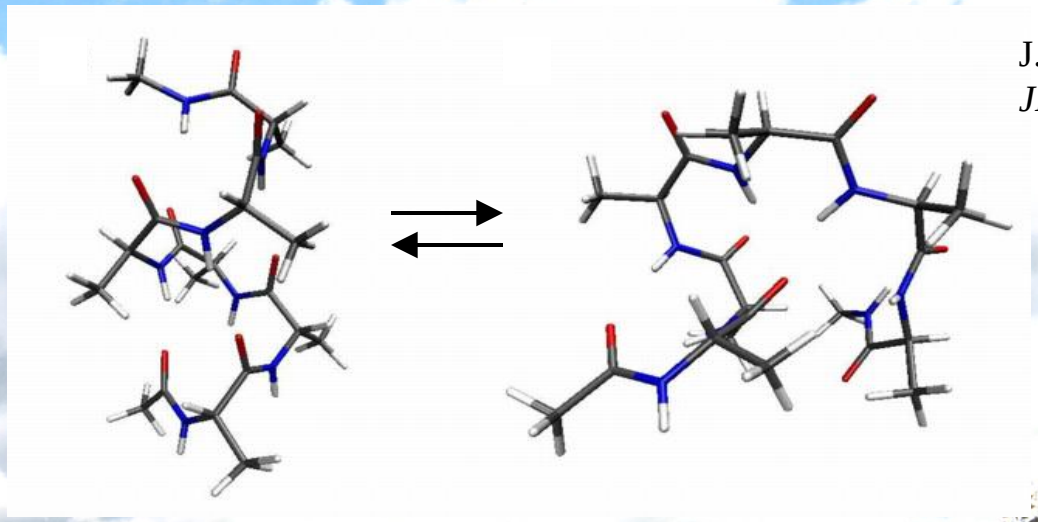
Sequence:
YQPDGSQA



I.-C. Lin and MET
J. Phys. Chem. B
(2010)

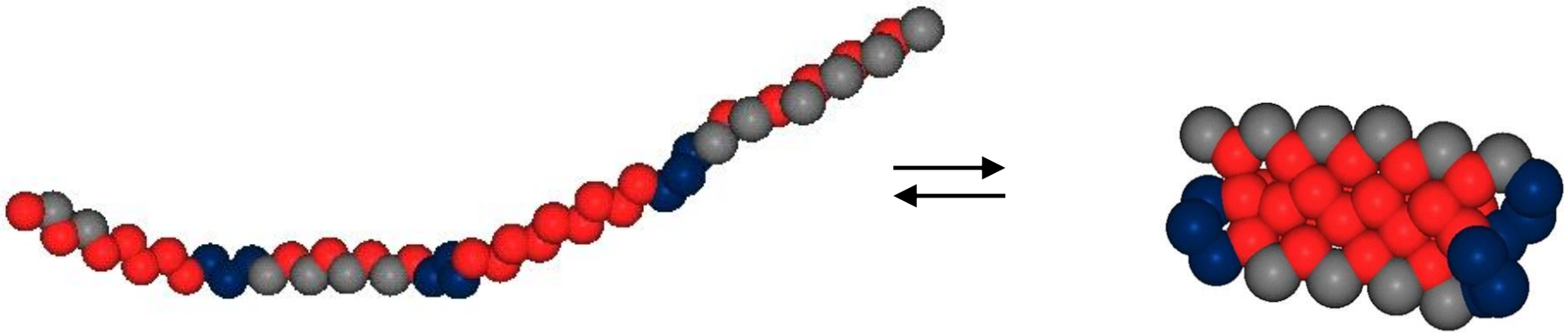
Solvent
mass scaling
0.1

Side-chain
mass scaling
0.6

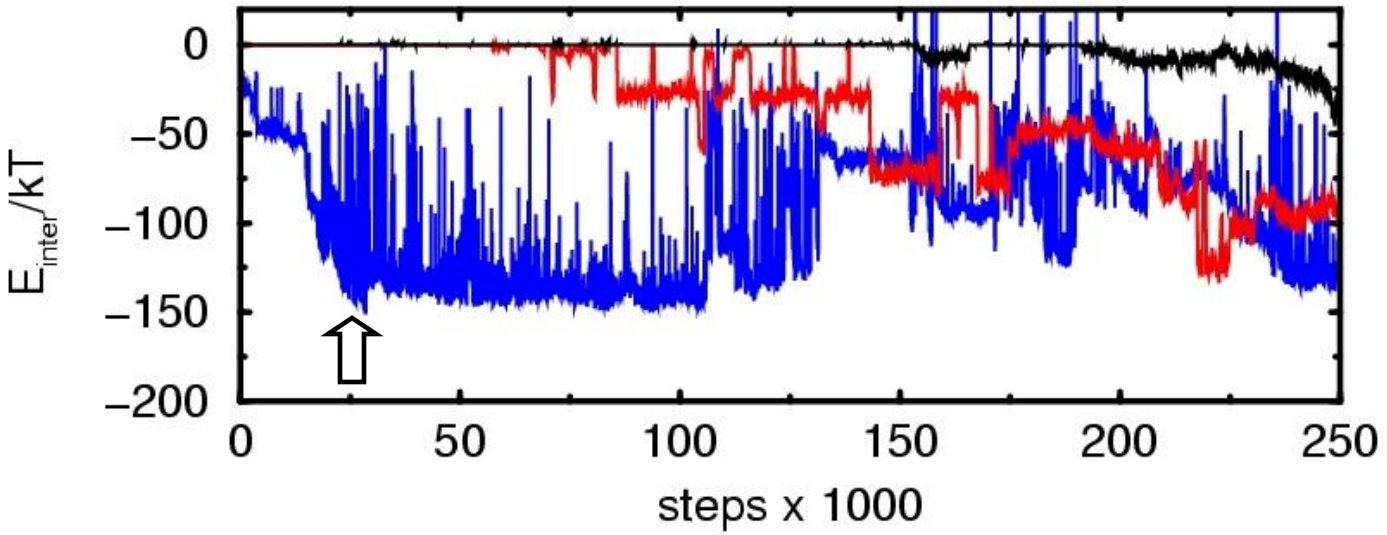


J. B. Abrams and MET
*JPCB***112** 15742 (2008)

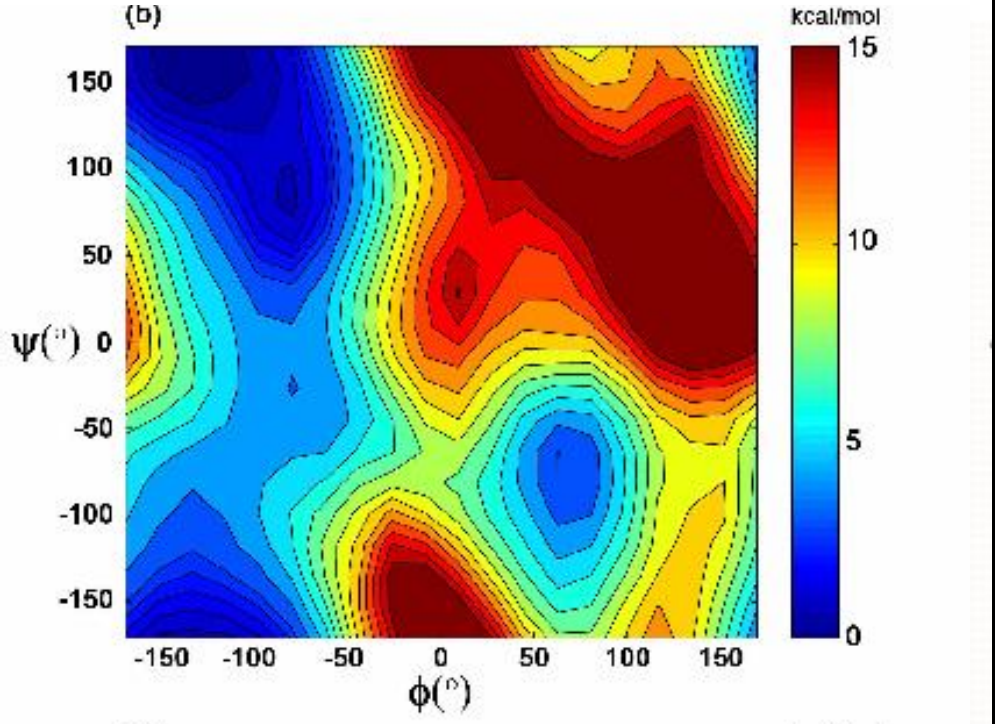
Conformational equilibrium

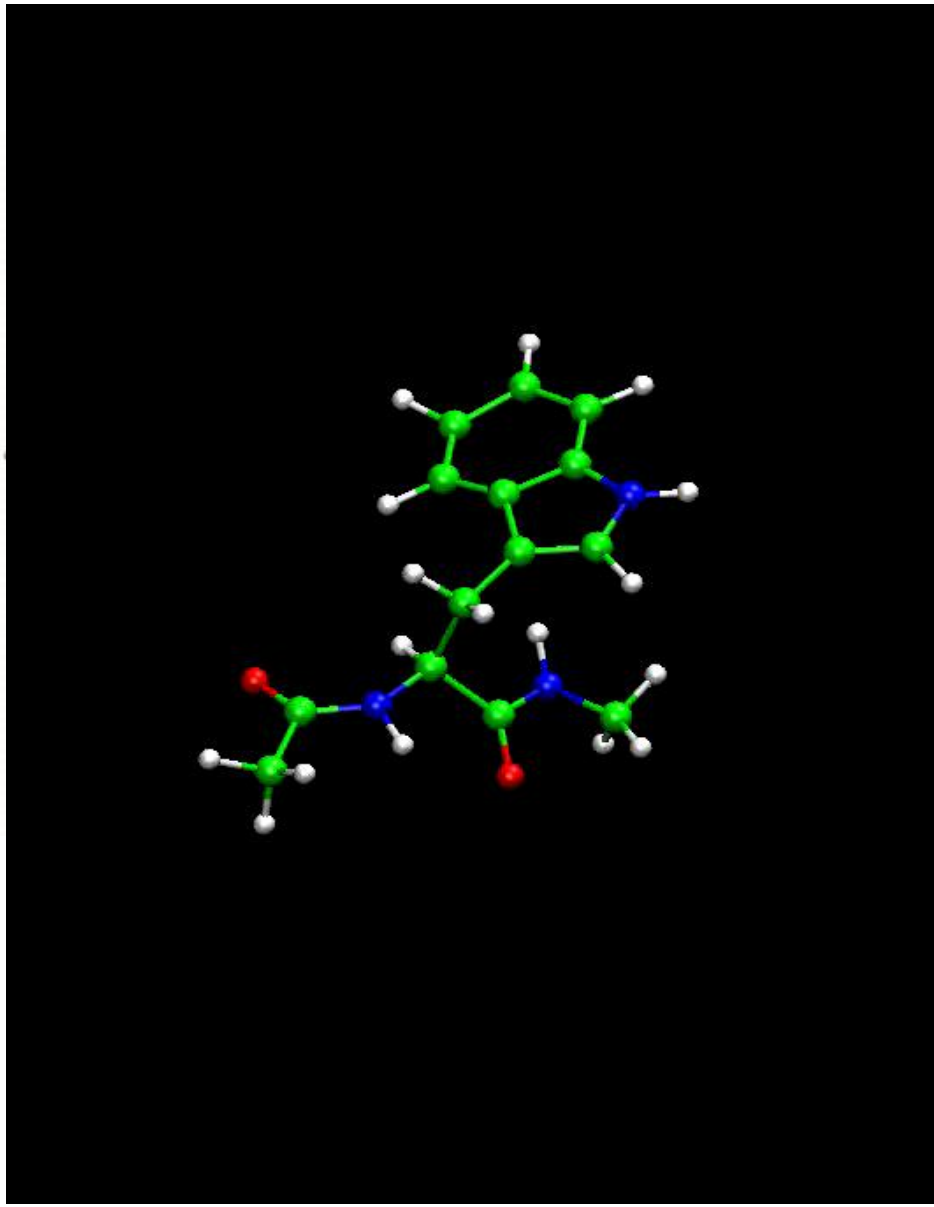


P. Minary, MET, G. J. Martyna *SIAM J. Sci. Comput*. **30**, 2055 (2008)

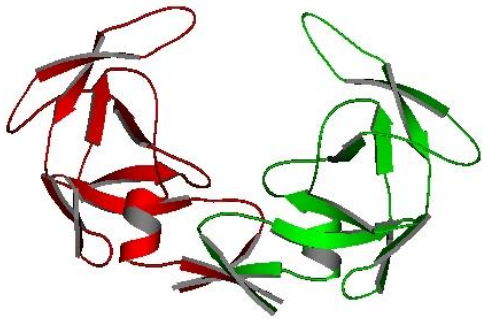


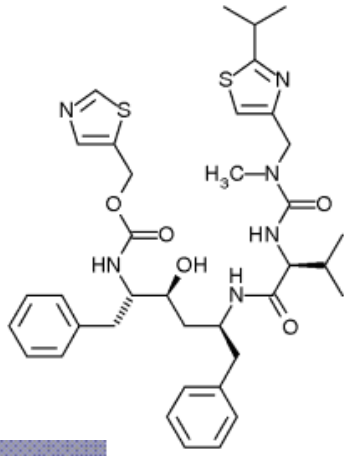
Ramachandran surface of NATMA





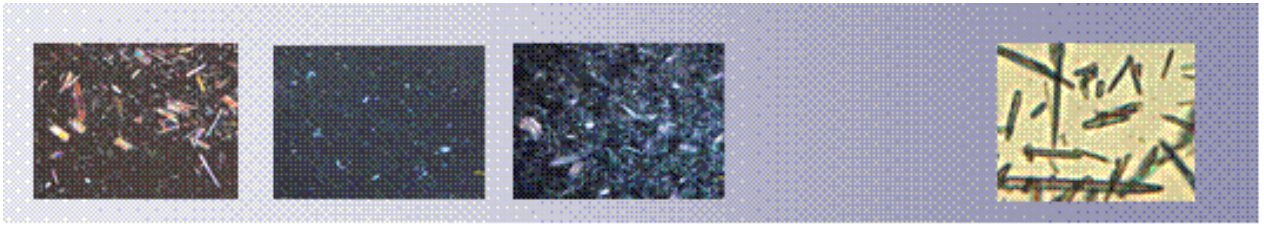
Polymorphic compounds

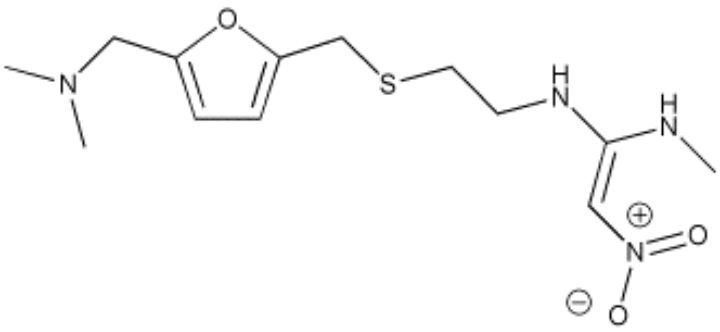




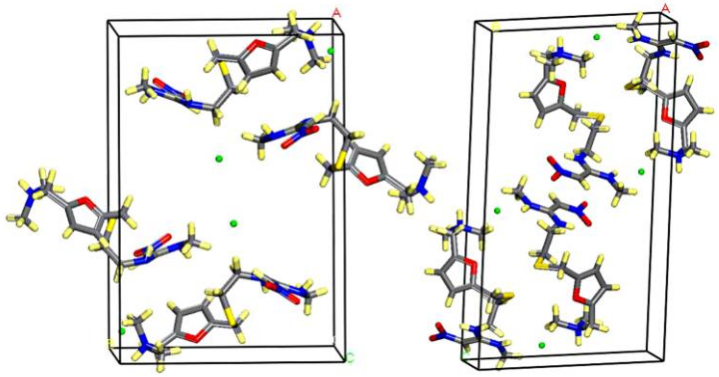
Norvir

1996 (launch) 1998 (recall) 2002 (Trans form)





Chemical Formula: $C_{13}H_{22}N_4O_3S$



Ranitidine
(Zantac)

Adiabatic free-energy dynamics

L. Rosso and MET *Mol. Simulat.* **28**, 91 (2002); L. Rosso, P. Minary, Z. Zhu, MET *J. Chem. Phys.* **116**, 4389 (2002)

In a transformation to generalized coordinates:

$$q_\alpha = q_\alpha(\mathbf{r}_1, \ldots, \mathbf{r}_N) \qquad \alpha = 1, \ldots, 3N \quad ; \qquad \mathbf{r}_i = \mathbf{r}_i(q_1, \ldots, q_{3N})$$

Suppose first *n* are of particular interest.

$$P(s_1, \ldots, s_n) = \int d^N\mathbf{p}\, d^N\mathbf{r}\; e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{\alpha=1}^{n} \delta(q_\alpha(\mathbf{r}) - s_\alpha)$$

$$= \int d^N\mathbf{p}\, d^{3N}q\; e^{-\beta \tilde{H}(\mathbf{p},q)} \prod_{\alpha=1}^{n} \delta(q_\alpha - s_\alpha)$$

$$H = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} + U(\mathbf{r}) \qquad \tilde{H} = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} + U(\mathbf{r}(q)) - kT \ln J(q)$$

Adiabatic and temperature conditions: $m_{1,\ldots,n}$ □ $m_{n+1,\ldots,3N}$ T_s □ T

Free-energy surface:

$$A(s_1, \ldots, s_n, T) = -kT_s \ln P_{\text{adb}}(s_1, \ldots, s_n)$$

Driven Adiabatic Free Energy Dynamics (d-AFED)

L. Rosso, P. Minary, Z. Zhu, MET *J. Chem. Phys*. **116**, 4389 (2002)
Margliano and Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006)
J. B. Abrams and MET, *J. Phys. Chem. B* **112**, 14752 (2008)

Suppose *n* collective variables characterize a free energy landscape of interest

$$q_\alpha = q_\alpha(\mathbf{r}_1, \ldots, \mathbf{r}_N) \qquad \alpha = 1, \ldots, n$$

Canonical probability distribution:

$$P(s_1, \ldots, s_n) = \int d^N\mathbf{p}\, d^N\mathbf{r}\; e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{\alpha=1}^{n} \delta(q_\alpha(\mathbf{r}) - s_\alpha)$$

Write δ-functions as product of Gaussians:

$$P(s_1, \ldots, s_n) = \lim_{\{\kappa_\alpha \to \infty\}} \int d^N\mathbf{p}\, d^N\mathbf{r}\; e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{\alpha=1}^{n} \left(\frac{\beta\kappa_\alpha}{2\pi}\right)^{1/2} \exp\left[-\frac{\beta\kappa_\alpha}{2}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]$$

Introduce uncoupled Gaussian integrations:

$$P_{\{\kappa\}}(s_1, \ldots, s_n) = C_{\{\kappa\}} \int d^N\mathbf{p}\, d^N\mathbf{r}\; d^n p_s \exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r}) + \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \sum_{\alpha=1}^{n} \frac{\kappa_\alpha}{2}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]\right\}$$

Effective Hamiltonian:

$$\mathcal{H}(\mathbf{p}, \mathbf{r}, s, p_s) = H(\mathbf{p}, \mathbf{r}) + \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \sum_{\alpha=1}^{n} \frac{\kappa_\alpha}{2}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2$$

Driven Adiabatic Free Energy Dynamics (d-AFED)

Introduce high temperature $T_s \gg T$ for extended variables and high masses $m_\alpha \gg m_i$

Adiabatically decoupled equations of motion:

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial H}{\partial \mathbf{r}_i} + \sum_\alpha \kappa_\alpha \left(s_\alpha - q_\alpha(\mathbf{r}) \right) \frac{\partial q_\alpha}{\partial \mathbf{r}_i} + \text{heat bath}(T)$$

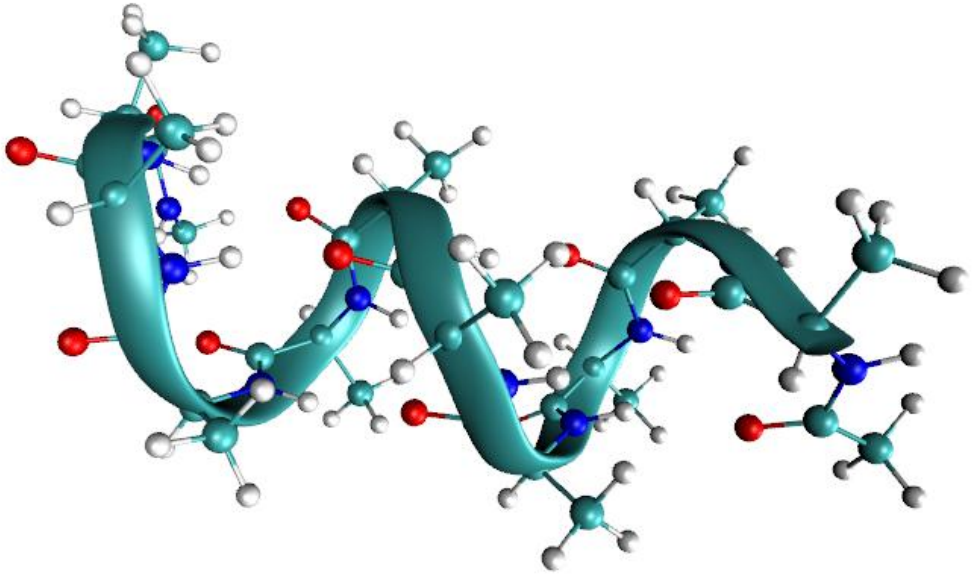
$$m_\alpha \ddot{s}_\alpha = -\kappa_\alpha \left(s_\alpha - q_\alpha(\mathbf{r}) \right) + \text{heat bath}(T_s)$$

Under adiabatic conditions, we generate a distribution $P_{\text{adb}}^{(\{\kappa\})}(s_1, \ldots, s_n, T_s, T)$

$$P_{\text{adb}}^{(\{\kappa\})}(s_1, \ldots, s_n, T_s, T) = \tilde{C}_{\{\kappa\}} \int d^n p e^{-\beta_s \sum_\alpha \frac{p_\alpha^2}{2 m_\alpha}} \left[P_{\{\kappa\}}(s_1, \ldots, s_n, T \right]^{T/T_s}$$

$$\lim_{\{\kappa \to \infty\}} \left[-k T_s \ln P_{\text{adb}}^{(\{\kappa\})}(s_1, \ldots, s_n, T_s) \right] = A(s_1, \ldots, s_n, T)$$

Alanine Decamer (gas phase)



Force field: CHARMM22

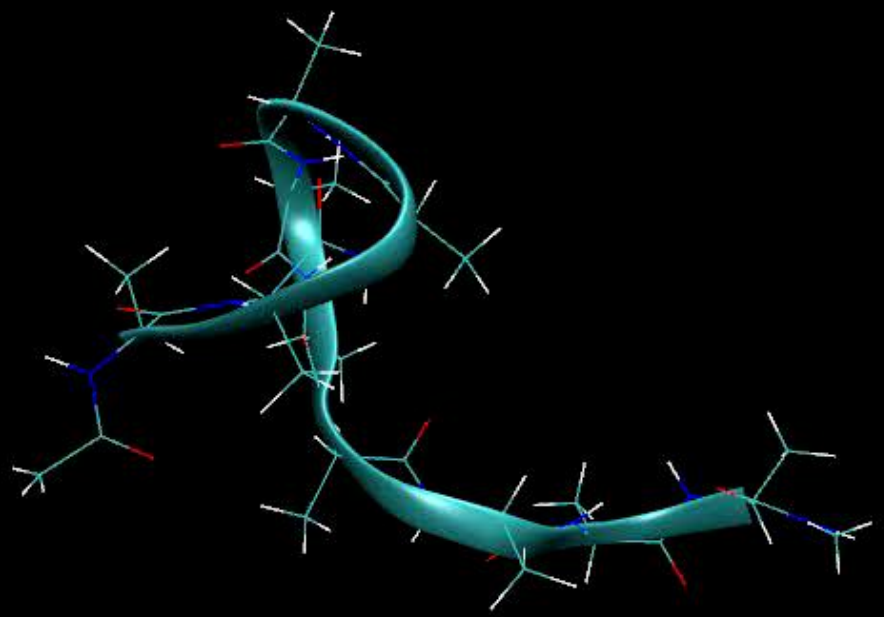
20 CVs: All (φ,ψ) pairs

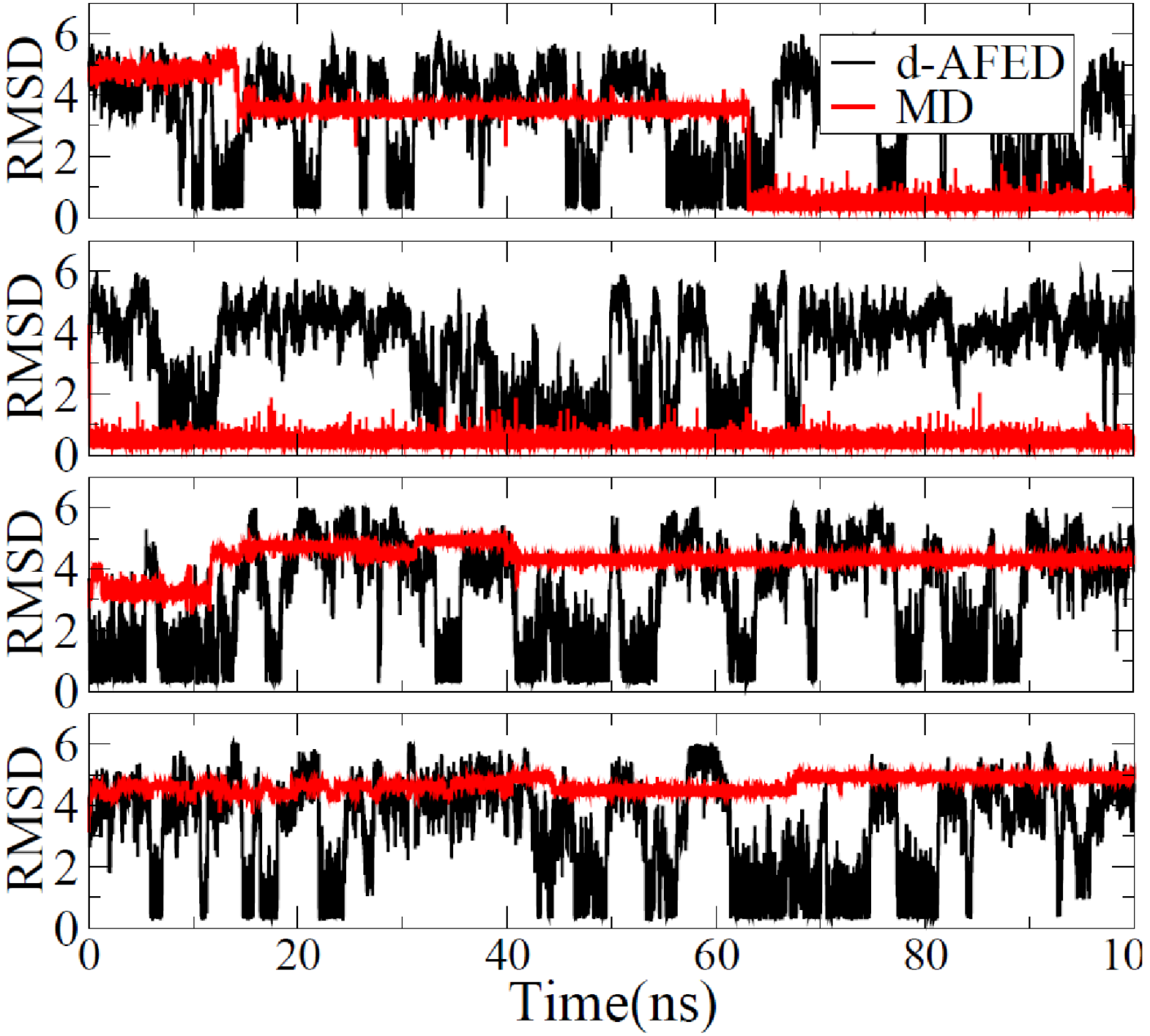
CV Temperature: T_s = 900 K

Physical Temp: T = 300 K

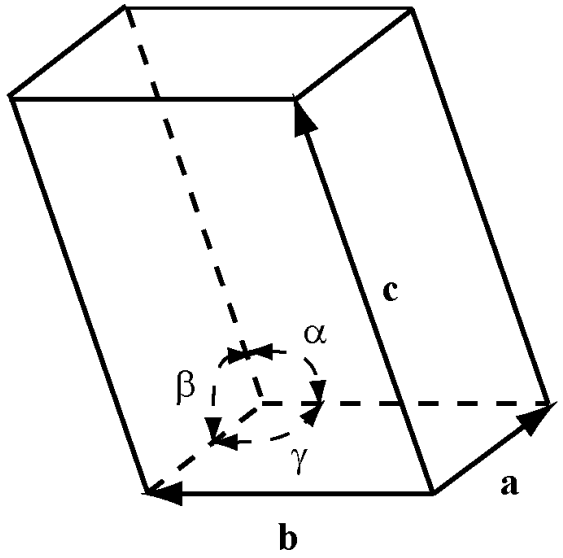
CV mass: $m_{(\psi,\varphi)} = 600 m_H$

Harmonic coupling: 300.5 kcal/mol/rad^2




d-AFED
MD
RMSD
RMSD
RMSD
RMSD
0
2
4
6
0
20
40
60
80
10
Time(ns)

Crystal-AFED [T. Q. Yu and MET *PRL* (2011)]



$$\mathbf{h} = \begin{pmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{pmatrix}$$

$$V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \det(\mathbf{h})$$

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{\mathbf{p}_\mathbf{g}}{W}\mathbf{r}_i, \qquad \dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{\mathbf{p}_\mathbf{g}}{W}\mathbf{p}_i - \frac{1}{N_f}\frac{\mathrm{Tr}\left[\mathbf{p}_\mathbf{g}\right]}{W}\mathbf{p}_i + \text{heat bath}(T)$$

$$\dot{\mathbf{h}} = \frac{\mathbf{p}_\mathbf{g}\mathbf{h}}{W}, \qquad \dot{\mathbf{p}}_\mathbf{g} = \det(\mathbf{h})\left(\mathbf{P}^{(\text{int})} - P\mathbf{I}\right) + \frac{1}{N_f}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\mathbf{I} + \text{heat bath}(T_\mathbf{h})$$

Large W, $\quad T_\mathbf{h} \to T$

Predicting crystalline polymorphs using adiabatic dynamics

T. Q. Yu and MET *Phys Rev. Lett.* (2011)

Using the AFED framework, we run NPT dynamics using the **h**-matrix as collective variables.

Generates the Gibbs free energy surface from the adiabatic probability distribution:

$$G(\mathbf{h}, T) = -kT_{\mathbf{h}} \ln P_{\text{adb}}(\mathbf{h}, T, T_{\mathbf{h}})$$

Test case: Benzene crystal using the Gromos 96 force field

$$T = 300 \text{ K}, \qquad T_{\mathbf{h}} = 32,000 - 40,000 \text{ K}$$

$$N = 216 \ (3 \times 3 \times 3), \qquad P = 2 \text{ GPa}$$

Nine different trajectories were initiated from randomly generated initial conditions. Total run time over all nine trajectories is 500 ps.

“Phase diagram” of benzene

Raiteri, Martonak, Parrinello *Angew. Chem. Intl. Ed.* **44**, 3769 (2005)

I

II

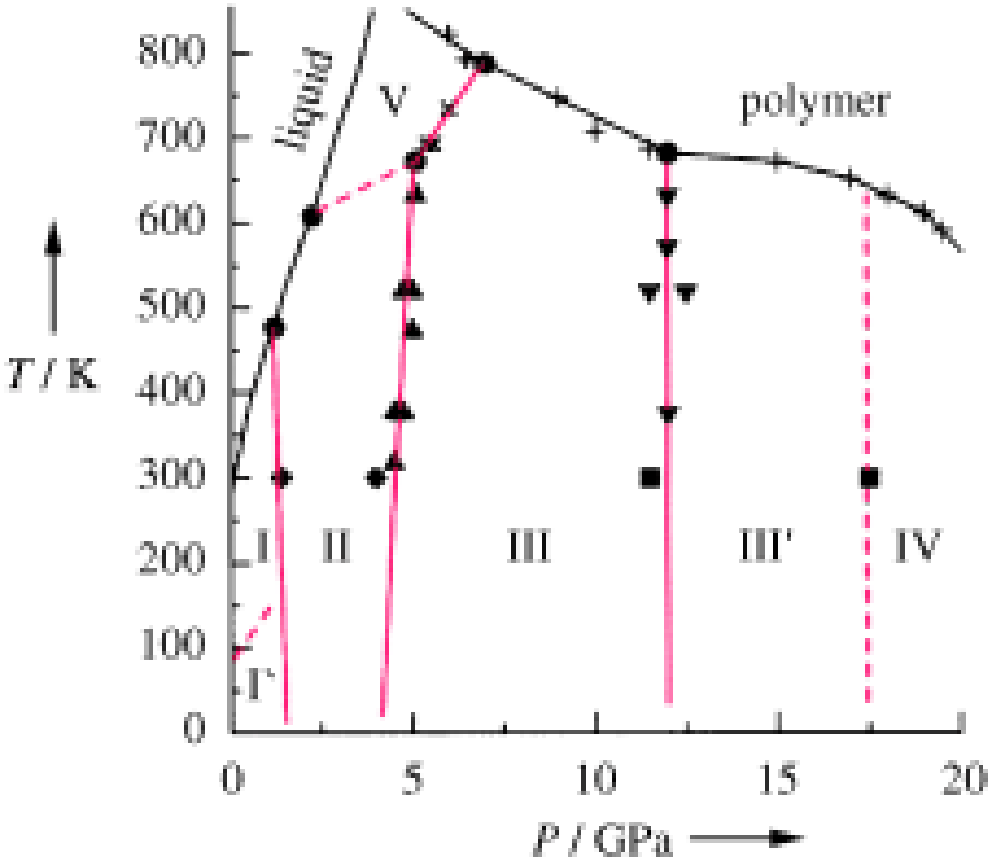
III

III′

IV

V

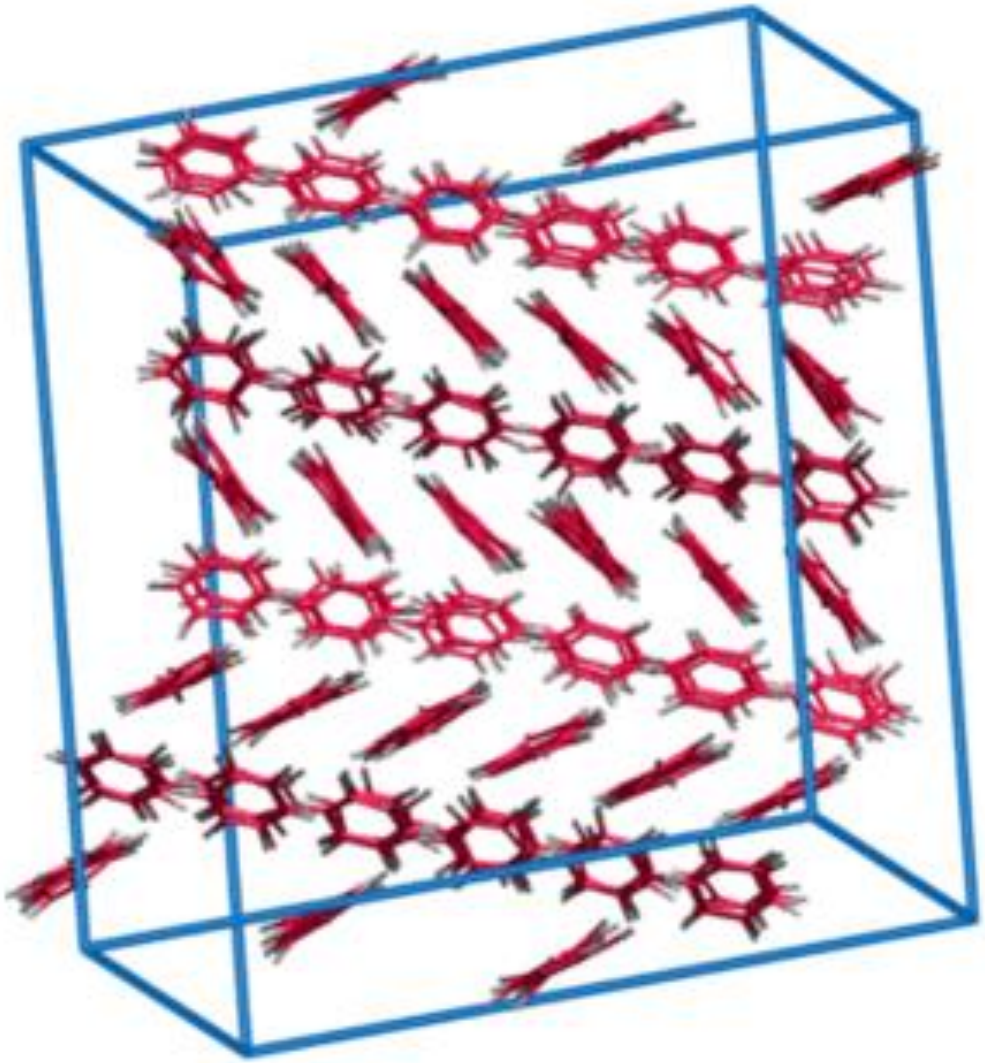
I′



$P2_1/c$ (II96) $P4_32_12$(II98) II01

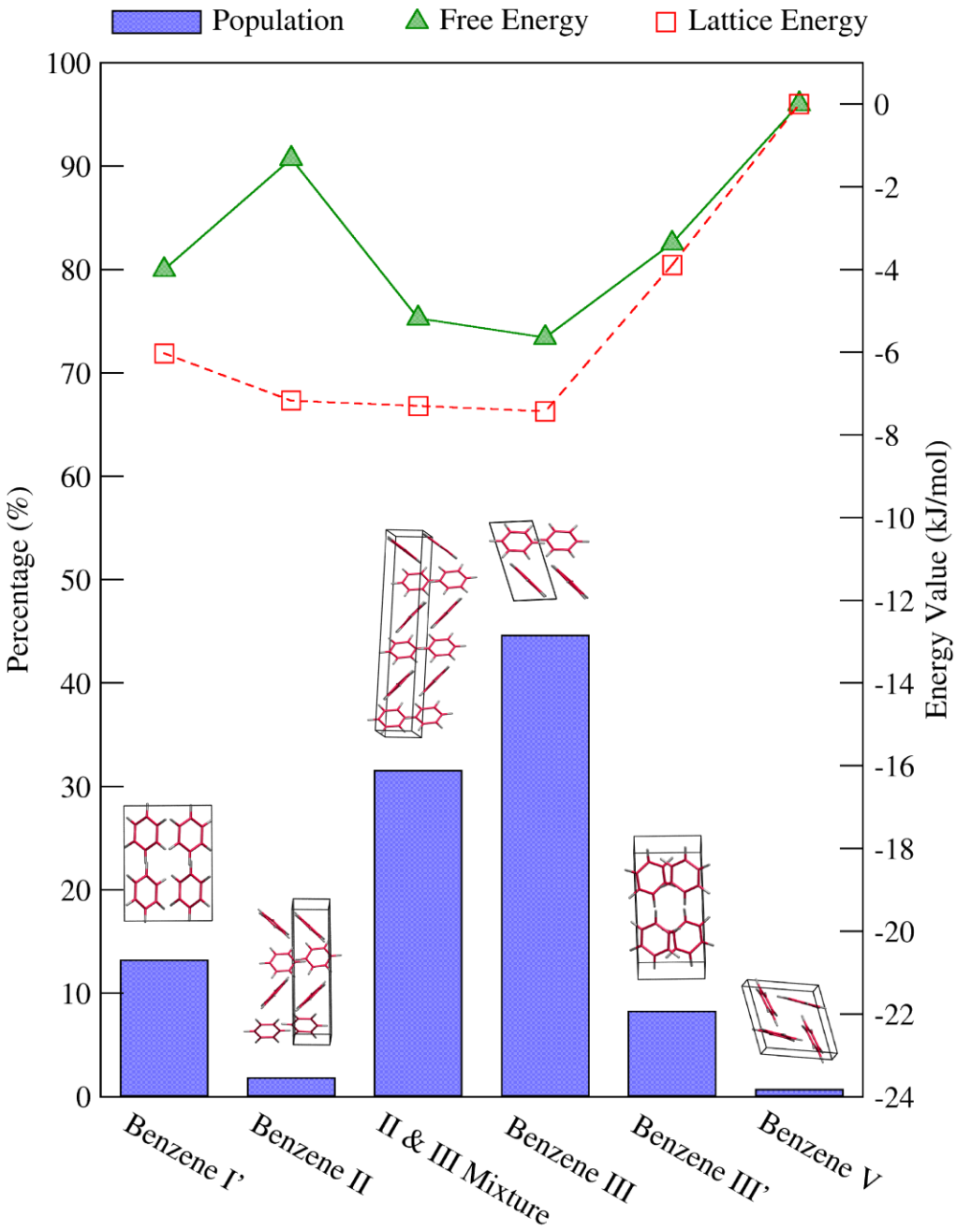
I, II, III, III′, IV known from X-ray and Raman scattering

I′ and V more speculative



GROMOS FF, $T = 300$ K, $T_h = 32,000 - 40,000$ K

$N = 216$ $(3 \times 3 \times 3)$, $P = 2$ GPa, Total run time = 500 ps, 5 ns


Population
Free Energy
Lattice Energy
Percentage (%)
Energy Value (kJ/mol)
100
90
80
70
60
50
40
30
20
10
0
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
Benzene I'
Benzene II
II & III Mixture
Benzene III
Benzene III'
Benzene V

Adiabatic conditions on $\{s,p_s\}$: $T_s \approx T$, $m_\alpha >> m$

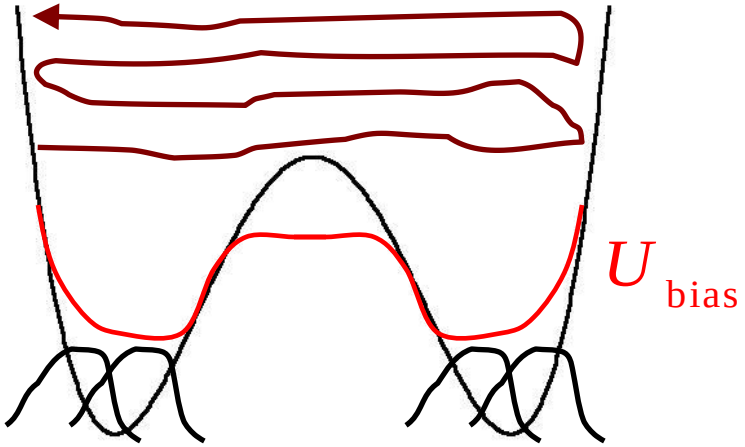
d-AFED with a bias potential

M. Chen, M. Cuendet, and MET *PNAS* (submitted)

Suppose a bias potential $U_{\text{bias}}(s_1,\ldots,s_n)$ is applied in the extended phase space.

Apply metadynamics-like bias in the extended-variable space:

$$U_{\text{bias}}(s) = \sum_i A e^{-\sum_\alpha \left\| s_\alpha - s_\alpha(t_i) \right\|^2 / 2\sigma^2}$$



Effective Hamiltonian:

$$H(\mathbf{p},\mathbf{r},s,p_s) = H(\mathbf{p},\mathbf{r}) + \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \frac{1}{2}\sum_{\alpha=1}^{n} \kappa_\alpha \left(q_\alpha(\mathbf{r}) - s_\alpha \right)^2 + U_{\text{bias}}(s)$$

Free energy surface:

$$A_{\{\kappa\}}(s_1,\ldots,s_n) = -kT_s \ln P_{\text{adb}}^{(\{\kappa\})}(s_1,\ldots,s_n) - U_{\text{bias}}(s_1,\ldots,s_n)$$

Using the free energy gradient

If the space of CVs is not too large, then we can employ free energy gradients:

$$F_\alpha(s_1,\ldots,s_n) = -\frac{\partial A}{\partial s_\alpha} = \left\langle \kappa_\alpha \left(q_\alpha(\mathbf{r}) - s_\alpha \right) \right\rangle$$

Manifestly independent of T_s and can be shown to be independent of bias.

Free energy reconstruction:

1. Expand free energy in a basis set: $A(\{s\}) = \sum_i C_i \psi_i(\{s\})$
2. Minimize gradient expansion on a grid:

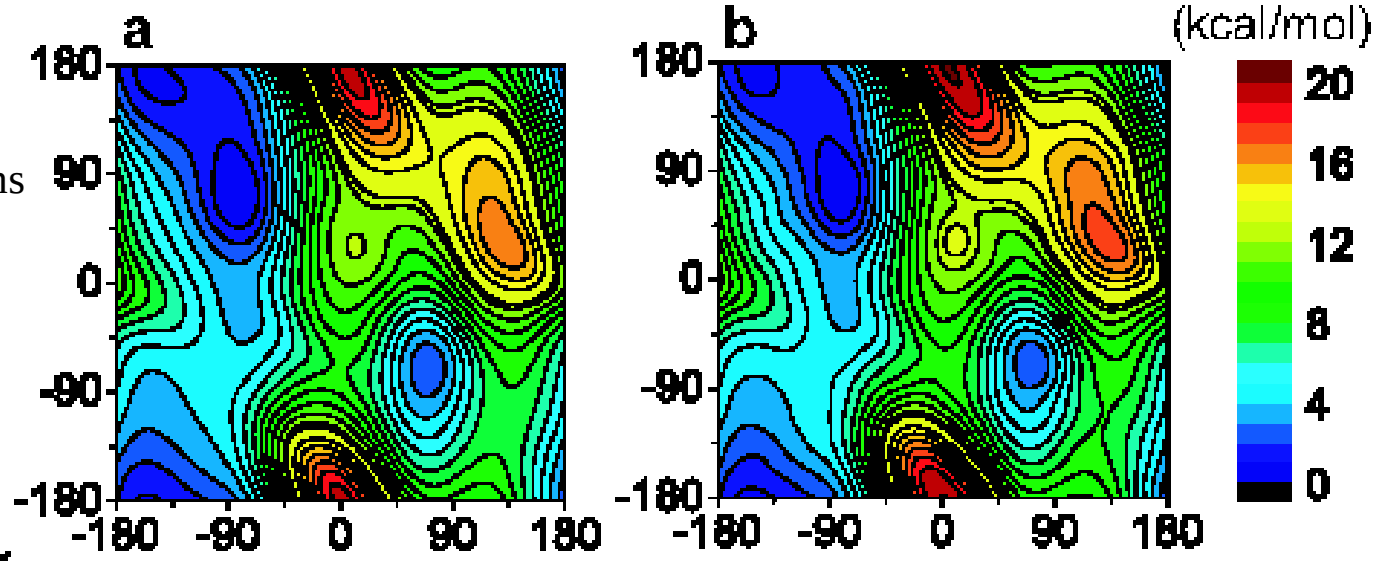
$$f(\{C\}) = \sum_k \sum_\alpha \left| \partial_\alpha A(\{s^{(k)}\}) + F_\alpha(\{s^{(k)}\}) \right|^2$$

3. Gives a set of linear equations

Overall scheme: Adiabatic dynamics, high T_s, gradients, bias – termed U-FED (unified FED).
M. Chen, M. Cuendet, MET PNAS (submitted).

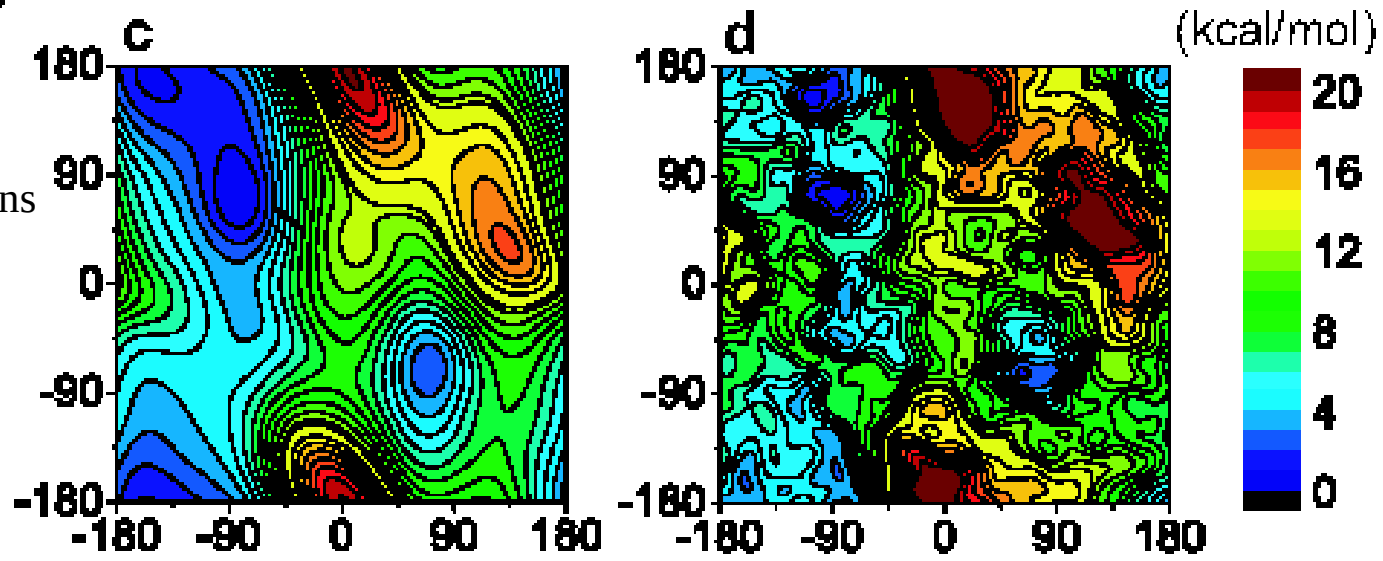
Ramachandran Surfaces of alanine-dipeptide in gas phase

U-FED
Small Gaussians
0.02 kcal/mol



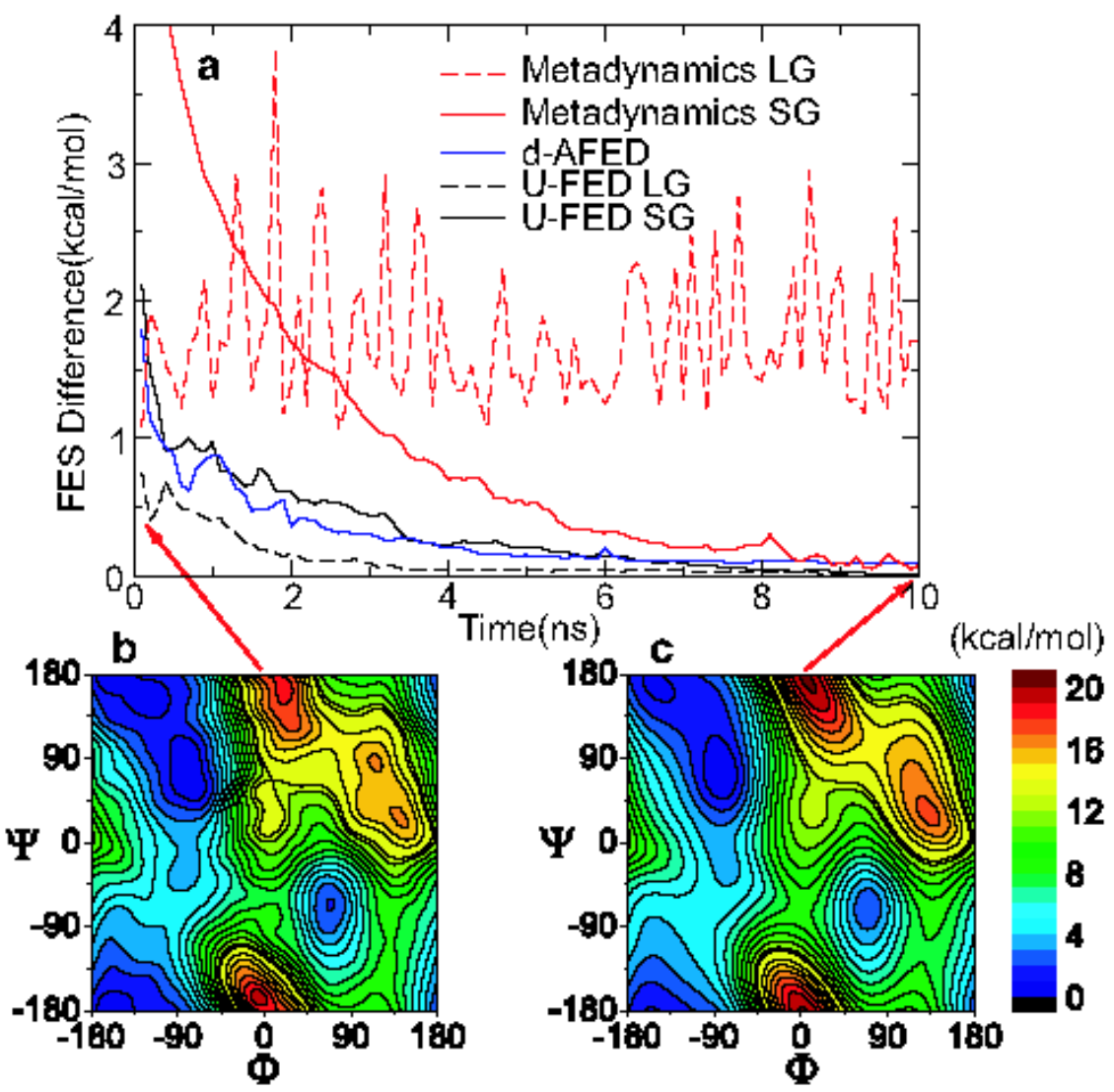
Metadyn.
Small Gauss.

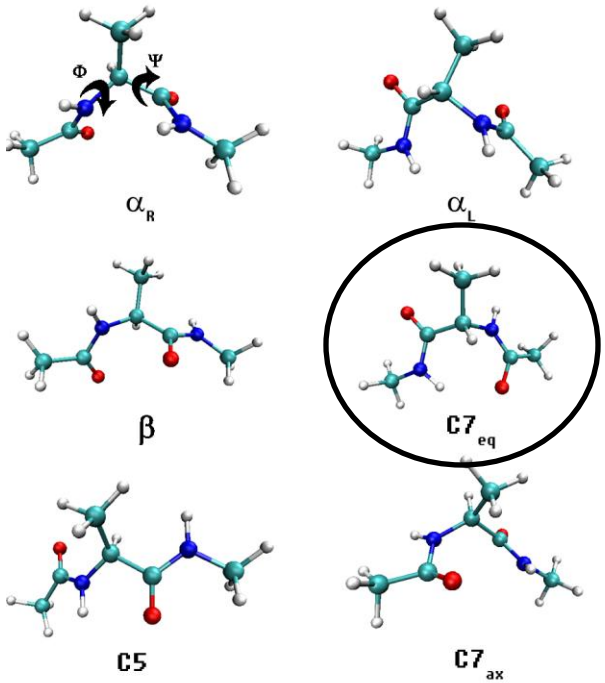
U-FED
Large Gaussians
0.2 kcal/mol



Metadyn.
Large Gauss.

Convergence of the FES for gas-phase alanine dipeptide



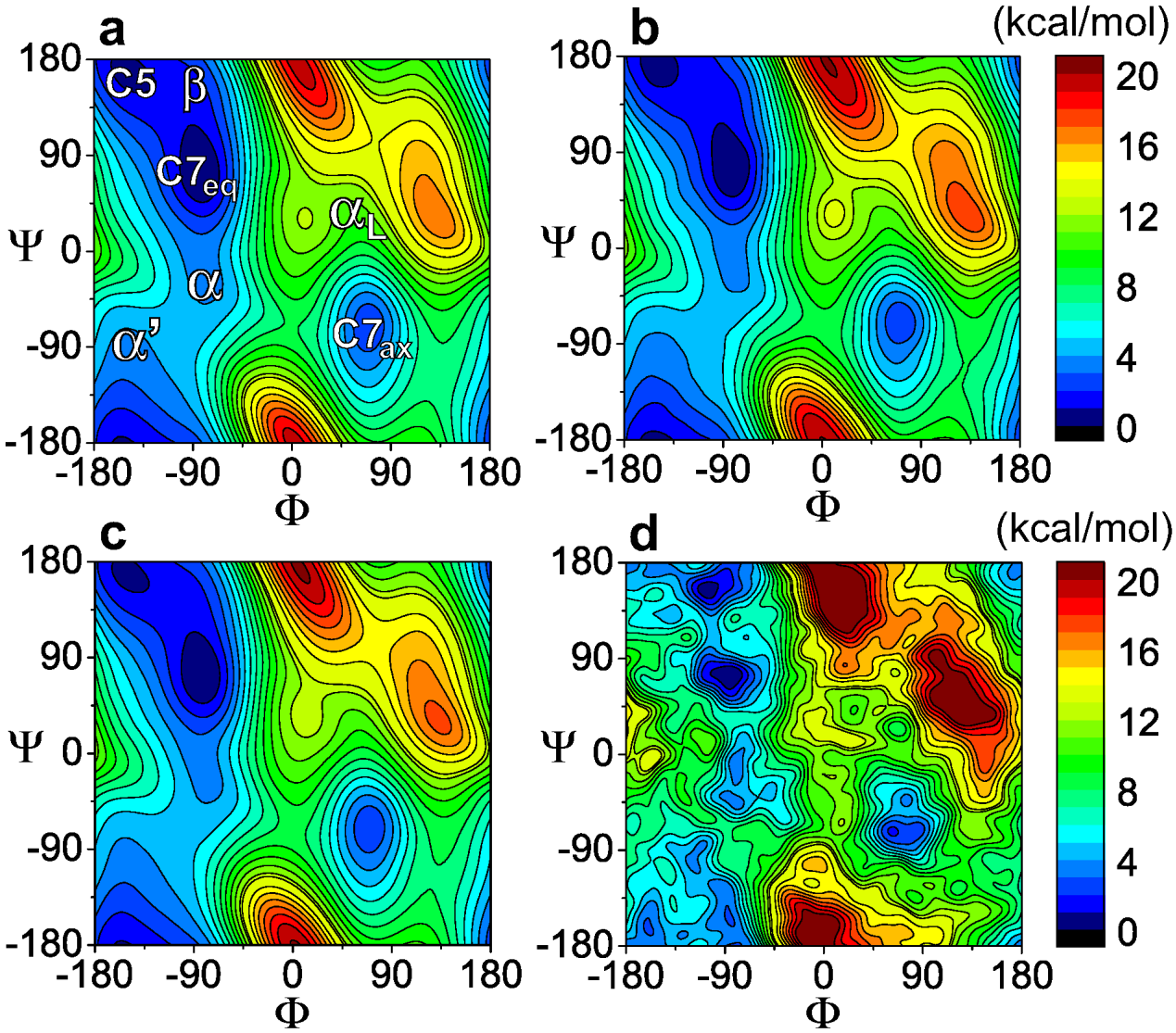


Biased d-AFED 500 ps Large Gauss.

Biased d-AFED 10 ns Large Gauss.

Convergence of FESs for alanine dipeptide in solution

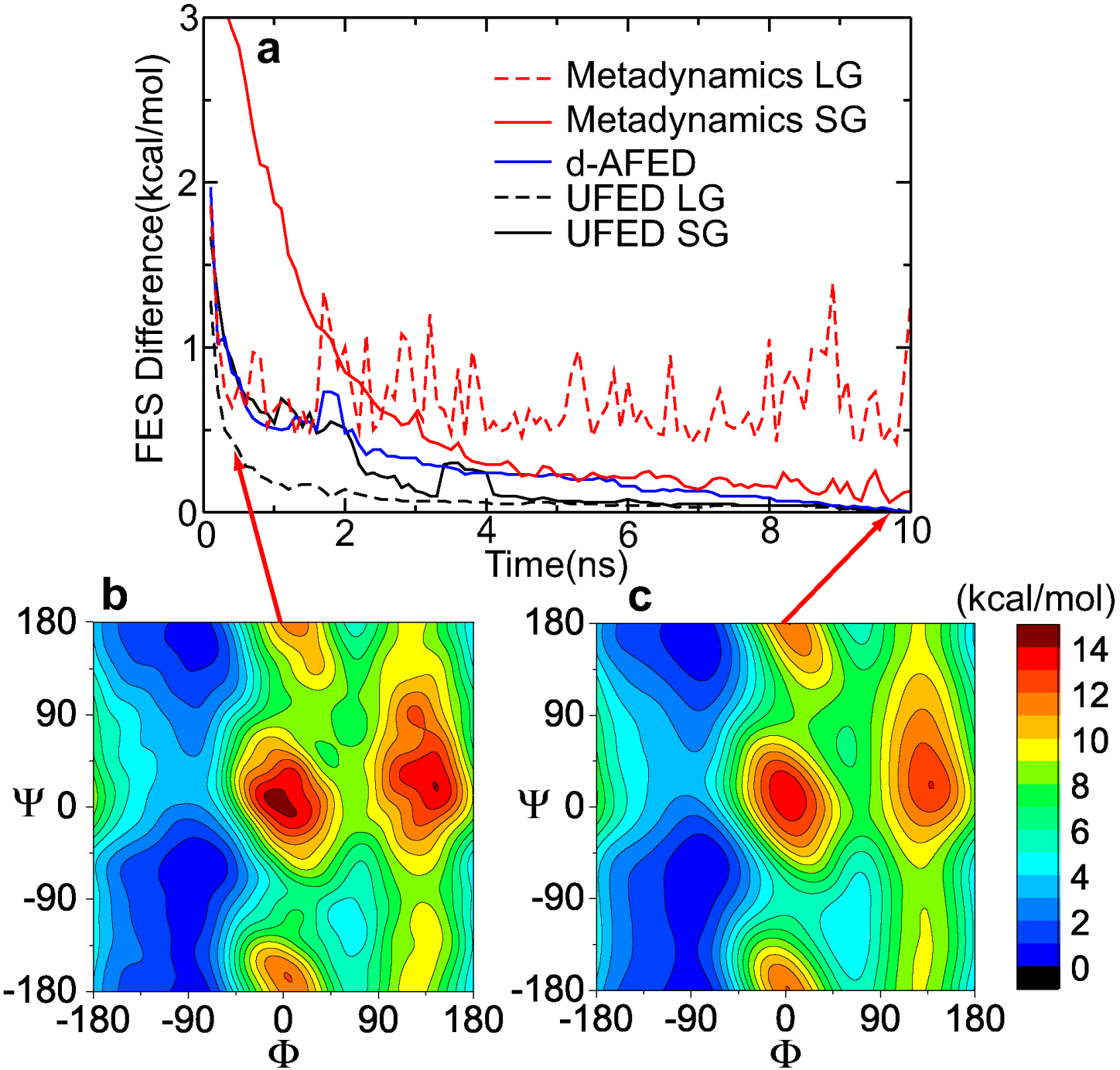
U-FED
Small Gaussians
0.02 kcal/mol



Metadyn.
Small Gauss.

U-FED
Large Gaussians
0.2 kcal/mol

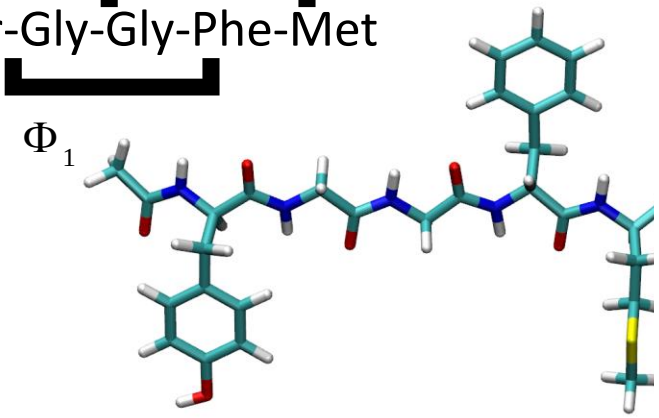
Metadyn.
Large Gauss.

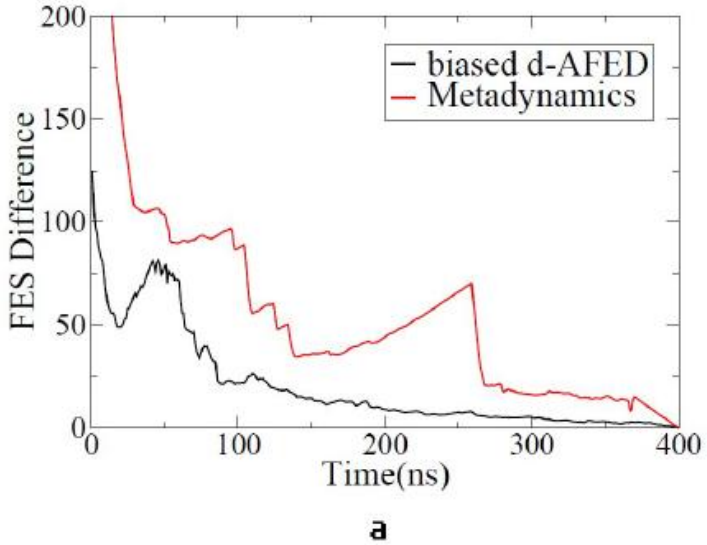

a
Metadynamics LG
Metadynamics SG
d-AFED
UFED LG
UFED SG
FES Difference(kcal/mol)
Time(ns)
0
1
2
3
0
2
4
6
8
10
b
c
(kcal/mol)
14
12
10
8
6
4
2
0
180
90
0
-90
-180
Ψ
-180
-90
0
90
180
Φ

Met-enkephalin

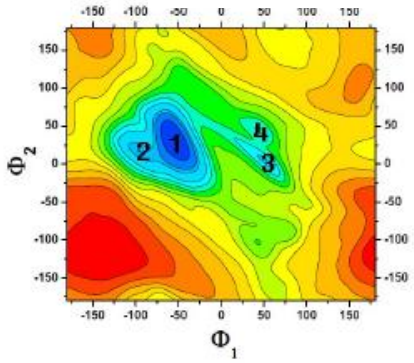
Tyr-Gly-Gly-Phe-Met

Φ_1 Φ_2

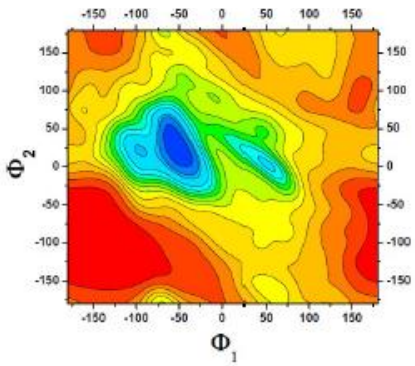




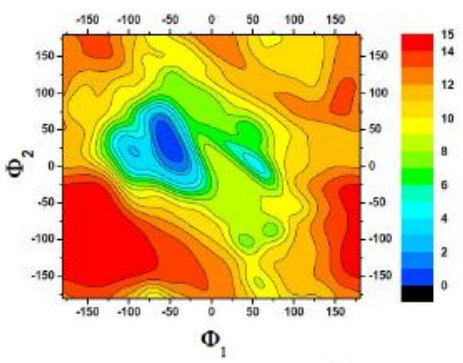
a



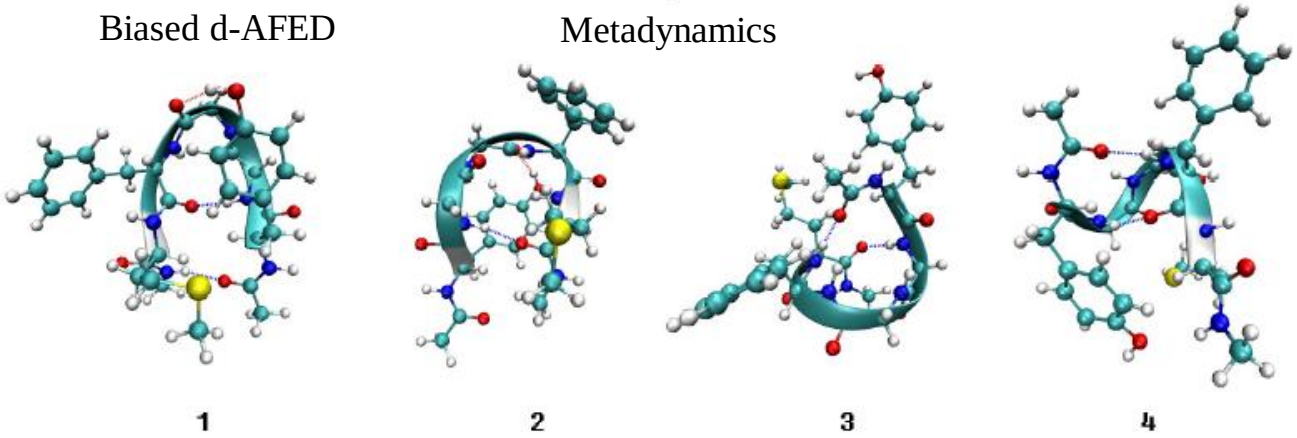
Biased d-AFED



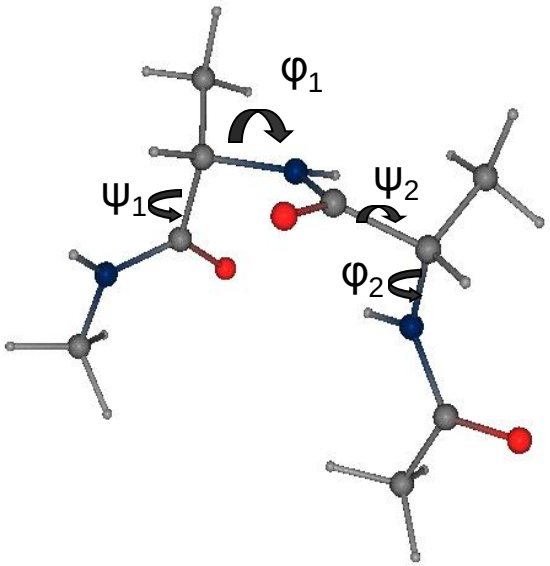
Metadynamics

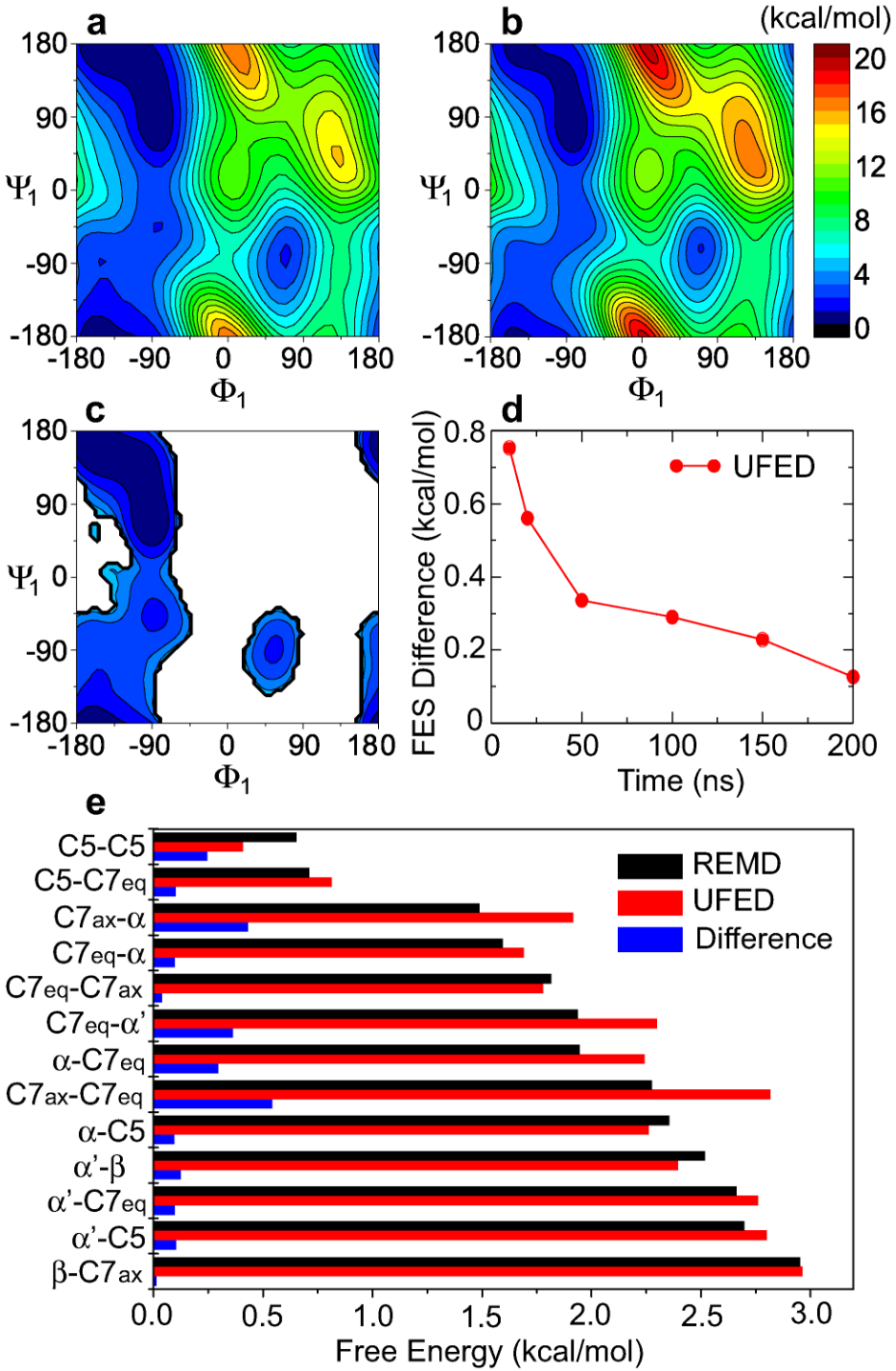


Metadynamics (GROMACS)



Conformations of the alanine tripeptide





Conclusions and perspectives

1. Adiabatic dynamics allows free energy surfaces to be explored and generated in a relatively large number of collective variables.

2. Applications include exploration of crystal polymorphs and small oligopeptides.

3. Previously also applied to solvation and binding free energies [J. B. Abrams and MET, *J. Chem. Phys*. (2006)]

4. Combining adiabatic dynamics with bias potentials generated in the spirit of metadynamics further improves the performance. The key is to add the Gaussians only the extended phase space and not to use them to construct the FES.

5. Applications to larger biomolecules and to drug docking currently under way.

6. Combine with Crystal-AFED for studying crystals of strongly interacting molecules.

Acknowledgments

Students past and present

- Zhongwei Zhu
- Peter Minary
- Tang-Qing Yu
- Ming Chen
- Jerry Abrams
- Lula Rosso

Postdocs

- I-Chun Lin
- Michel Cuendet

Funding

- NSF Chemistry
- NYU Whitehead Award
- Camille and Henry Dreyfus Foundation
- NSF MRSEC