Current Topics in Theoretical Chemisty,
Cumbaya, Ecuador, July5, 2019

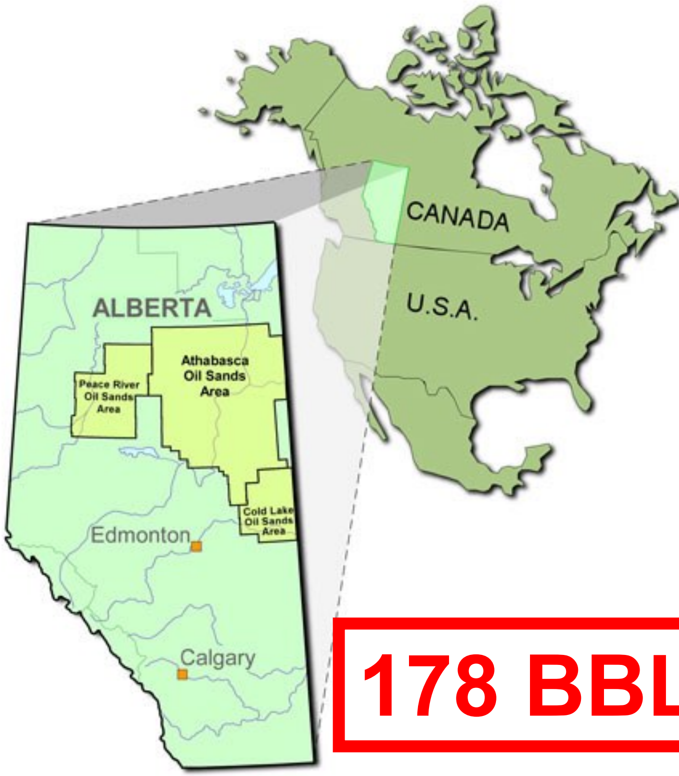
Towards free-energy profiles for nano-catalyzed reactions in complex environments

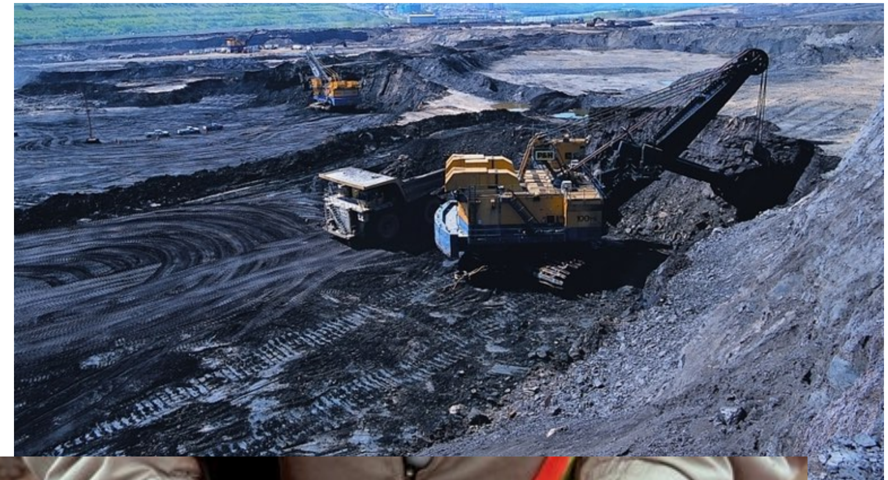
Dennis Salahub, Xingchen Liu, Baojing Zhou, Alex Tkalych, Andreas Köster, Farouq Ahmed, Thomas Heine, Augusto Oliveira, Muhammad Wahiduzzaman, Mauricio Chagas da Silva, Jiří Hostaš, Shideh Ahmadi, Morteza Chehelamirani, Lizandra Barrios, Patrizia Calaminici, Domingo Cruz-Olvera

University of Calgary (Chemistry, CMS, IQST, Quantum Alberta), Alberta, Canada

NSERC (CIAM) /SENER/CONACYT
Andreas Köster (CINVESTAV, Mexico City), Helio Duarte (Belo Horizonte, Brazil), Pedro Pereira (Chem and Petroleum Eng, Calgary)

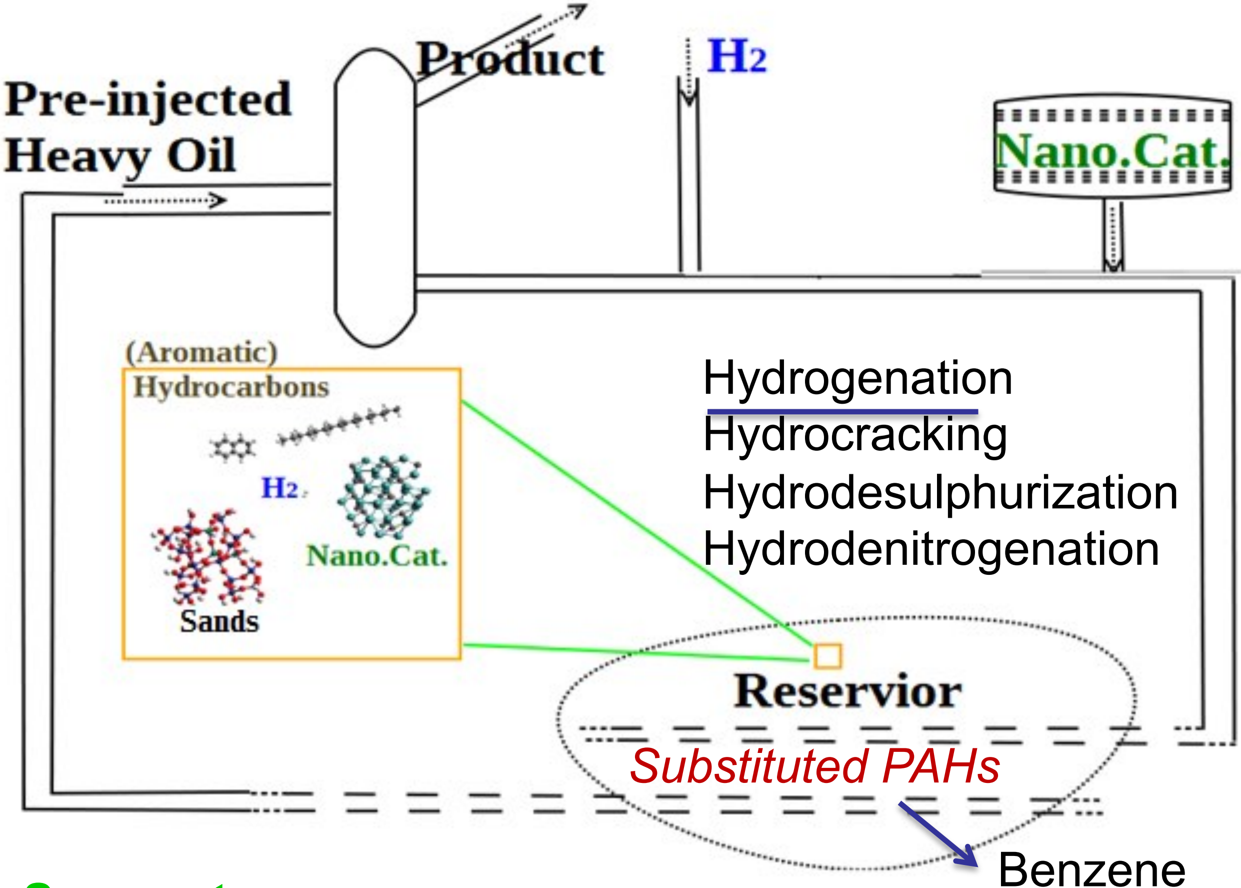
EU – IRSES TEMM1P – Computer simulations of thermally excited molecules and materials by first principles
Thomas Heine (Jacob's University, Bremen) + 6 others


CANADA
U.S.A.
ALBERTA
Athabasca
Oil Sands
Area
Peace River
Oil Sands
Area
Cold Lake
Oil Sands
Area
Edmonton
Calgary

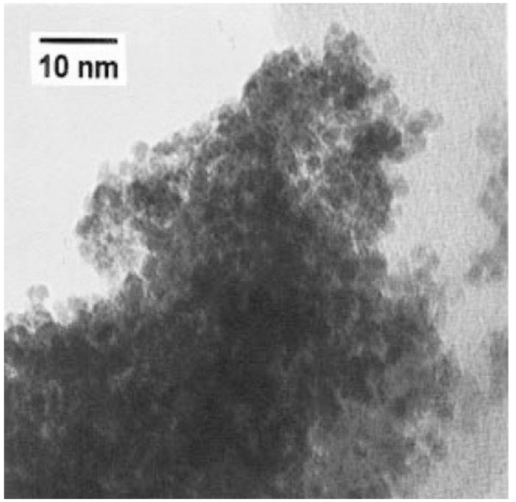


178 BBL

Motivation: *In-situ* Upgrading of Oil Sands via MCNPs



2-3 nm MCNPs



Hyeon et al. J Am Chem Soc 1996, 118 (23), 5492

- ✓ **Save water**
- ✓ **Save energy**
- ✓ **Save the environment**

Goal: understand the chemistry in the oil reservoir

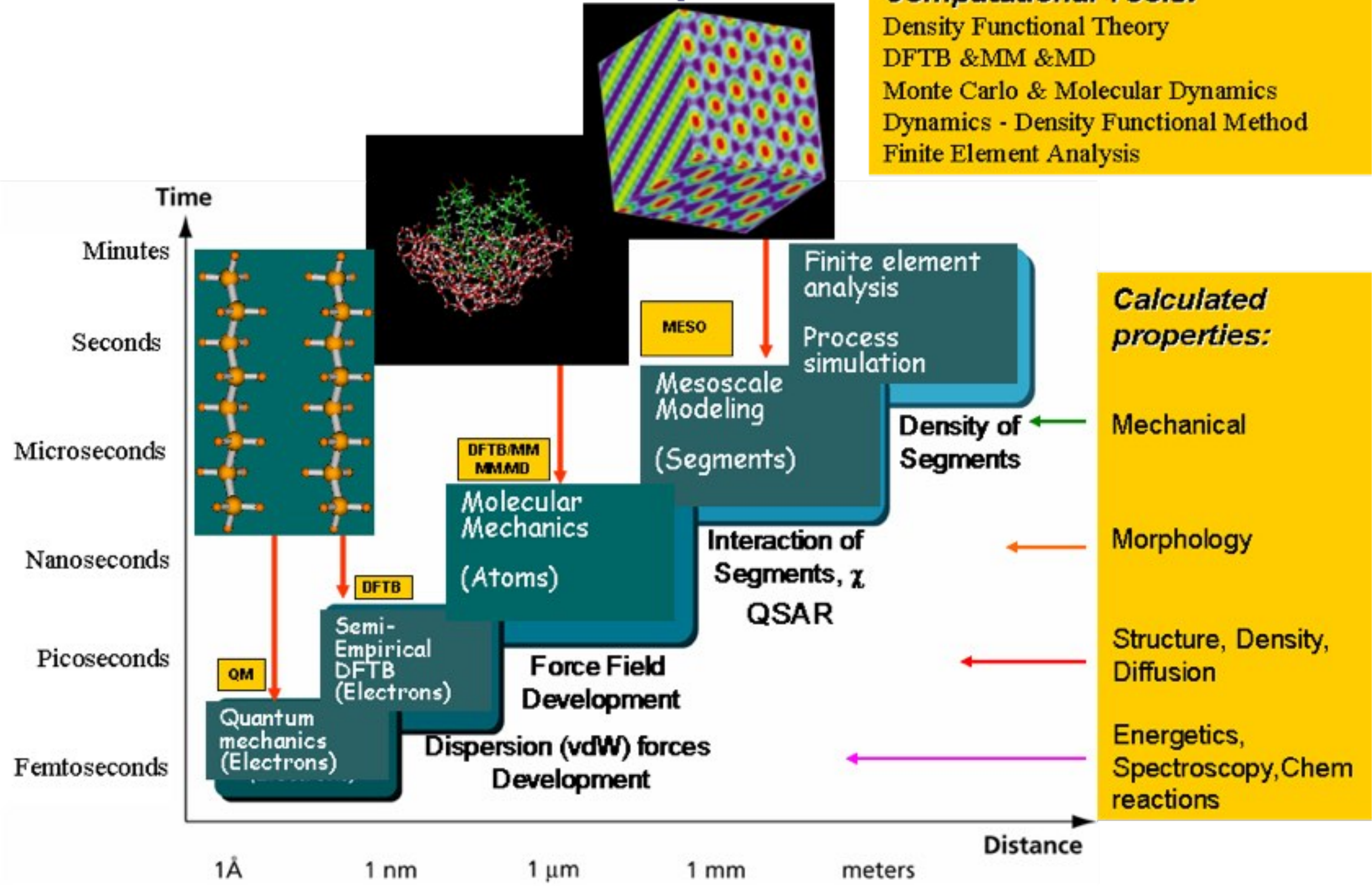
P. Pereira, *Can. J Chem. Eng.* **2012,** *90* (2), 320

NSERC CRSNG

Inter-American Collaboration in Materials Research (CIAM)

Multiscale modeling of self-assembled and functionalized nanoobjects

Computational Tools:
Density Functional Theory
DFTB &MM &MD
Monte Carlo & Molecular Dynamics
Dynamics - Density Functional Method
Finite Element Analysis



Outline

- **Benzene hydrogenation mechanism:** [Adsorption of Cyclic C_6H_6, C_6H_8, C_6H_{10} and C_6H_{12} on the (0001) surface of α-Mo_2C]
- **Parameterizing a faster quantum mechanical method:** [DFTB Parameterization of Mo, C, H, O and Si]
- **The role of entropy and the environment:** [Molybdenum Carbide Nanocatalysts at Work in the *in-situ* Environment: a DFTB and QM(DFTB)/MM Study]

X. Liu, D. R. Salahub et al, *J Phys Chem C* **2013,** *117* (14), 7069
X. Liu, D. R. Salahub et al, Theor.Chem.Acc., **2016,** 135:168, 1-14
X. Liu, D. R. Salahub et al, *J Am Chem Soc* **2015**, 137, 4249

Outline

- **Benzene hydrogenation mechanism:** [Adsorption of Cyclic C_6H_6, C_6H_8, C_6H_{10} and C_6H_{12} on the (0001) surface of α-Mo_2C]
- **Parameterizing a faster quantum mechanical method:** [DFTB Parameterization of Mo, C, H, O and Si]
- **The role of entropy and the environment:** [Molybdenum Carbide Nanocatalysts at Work in the *in-situ* Environment: a DFTB and QM(DFTB)/MM Study]

X. Liu, D. R. Salahub et al, *J Phys Chem C* **2013,** *117* (14), 7069

X. Liu, D. R. Salahub et al, Theor. Chem. Acc, in press

X. Liu, D. R. Salahub et al, *J Am Chem Soc 2015, <u>137</u>, 4249*

$$\hat{H}\Psi = E\Psi \qquad E = E[\rho(\mathbf{r})]$$

KS-DFT Choices - deMon2k

$$E_{xc}\,, \quad v_{xc} = \delta E_{xc} / \delta\rho$$

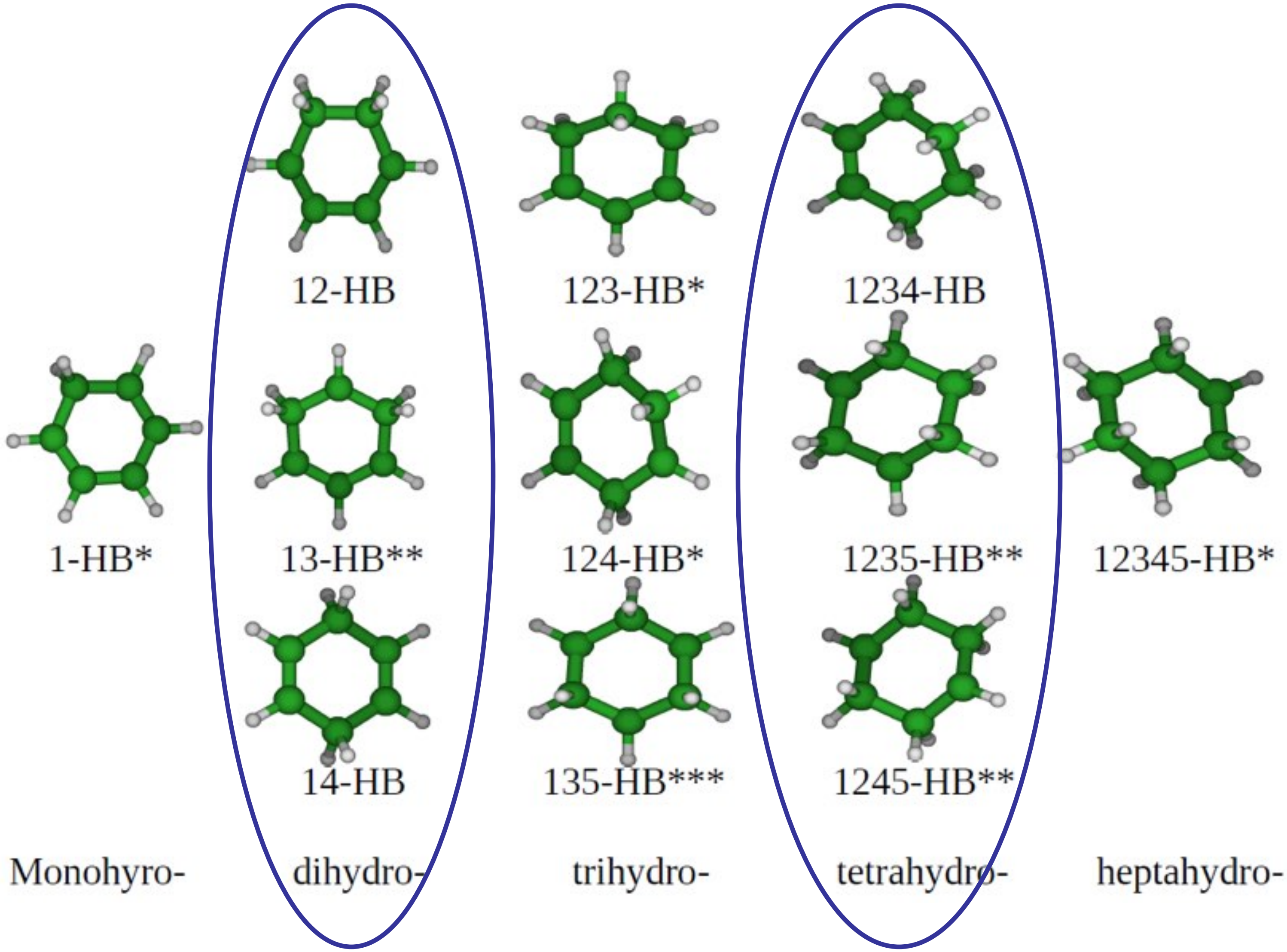
- Exact
- Fitting
- Multipoles

- Gaussians
- Slaters
- Numerical Functions
- Plane waves
- APW etc.

$$(-1/2\ \nabla^2 + V_{ne} + V_C + V_{XC})\ \phi_i^{KS} = \varepsilon_i\ \phi_i^{KS}$$

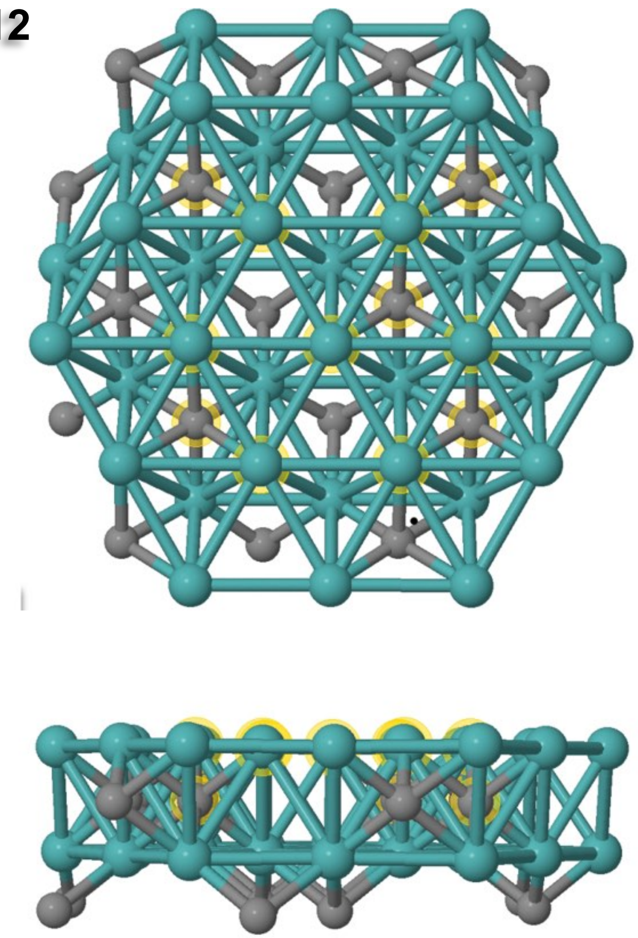
- L(S)DA
- GGA
- Hybrid
- AC
- LAP,TAU (meta-GGA)
- OEP

Benzene Hydrogenation Reaction Intermediates



C_6H_7 C_6H_8 C_6H_9 C_6H_{10} C_6H_{11}

Adsorption of C_6H_6, C_6H_8, C_6H_{10}, C_6H_{12}

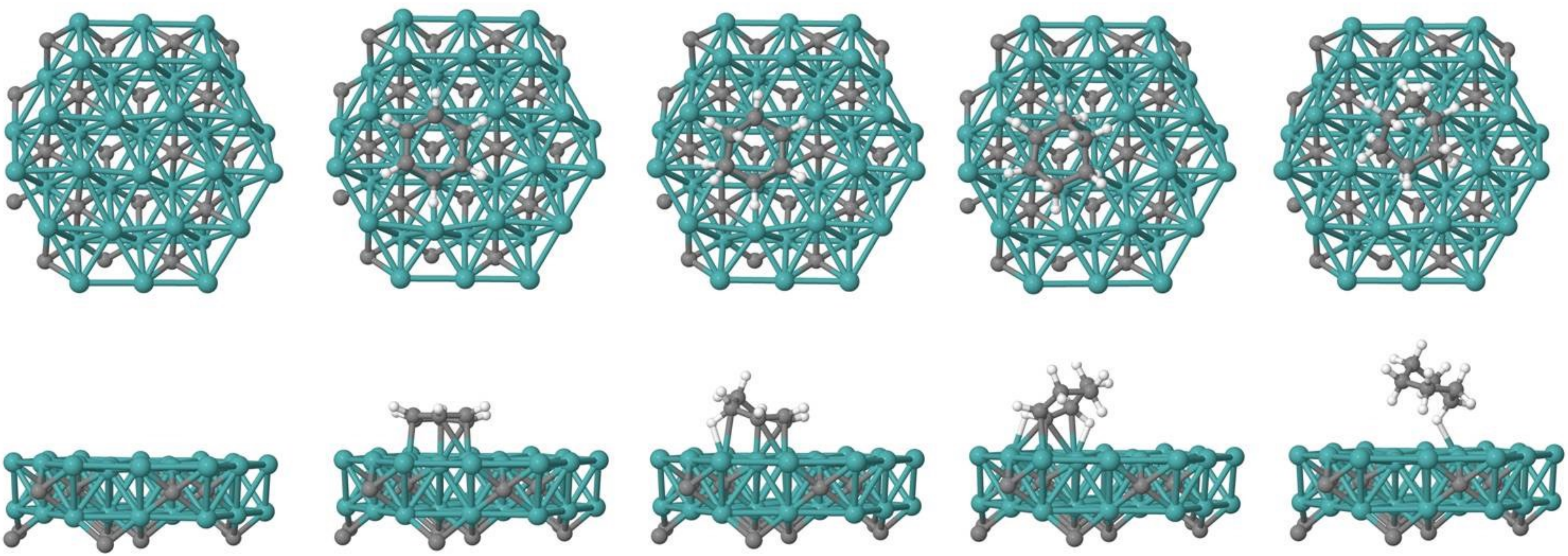


cluster model $Mo_{38}C_{19}$

PBE/RMCP(DZVP)/GEN-A2/Dispersion

X. Liu, D. R. Salahub et al J Phys Chem C **2013,** *117* (14), 7069

Thermodynamically Stable Adsorption Configuration



C_6H_6 | $12\text{-}C_6H_8$ | $1234\text{-}C_6H_{10}$ | C_6H_{12}

What are missing from the DFT studies?

- The topology (shape) of the real active sites on MCNPs;
- The electron delocalization over the real nanoparticles;
- The reaction barriers

$\Downarrow$

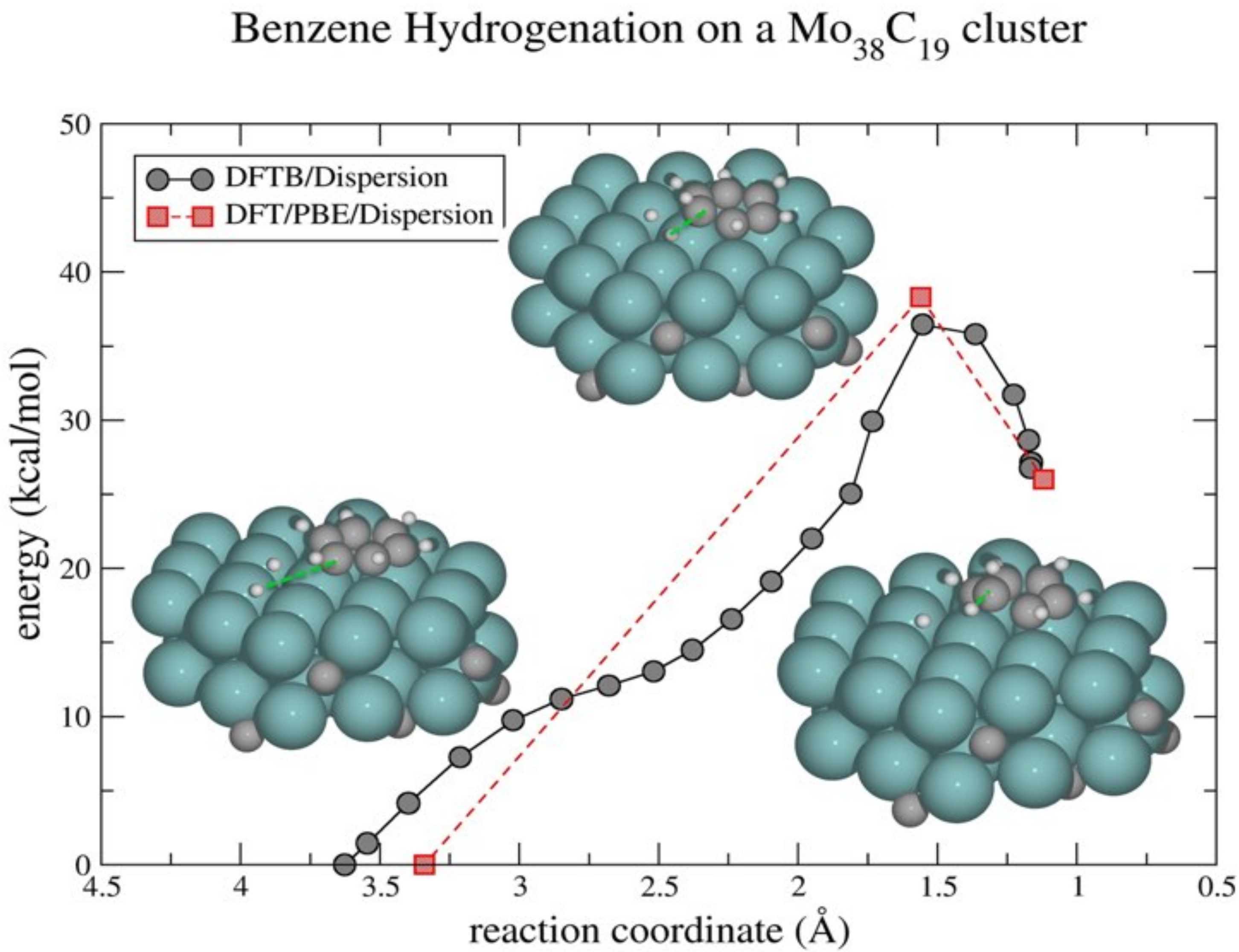
An even faster method is needed!

Outline

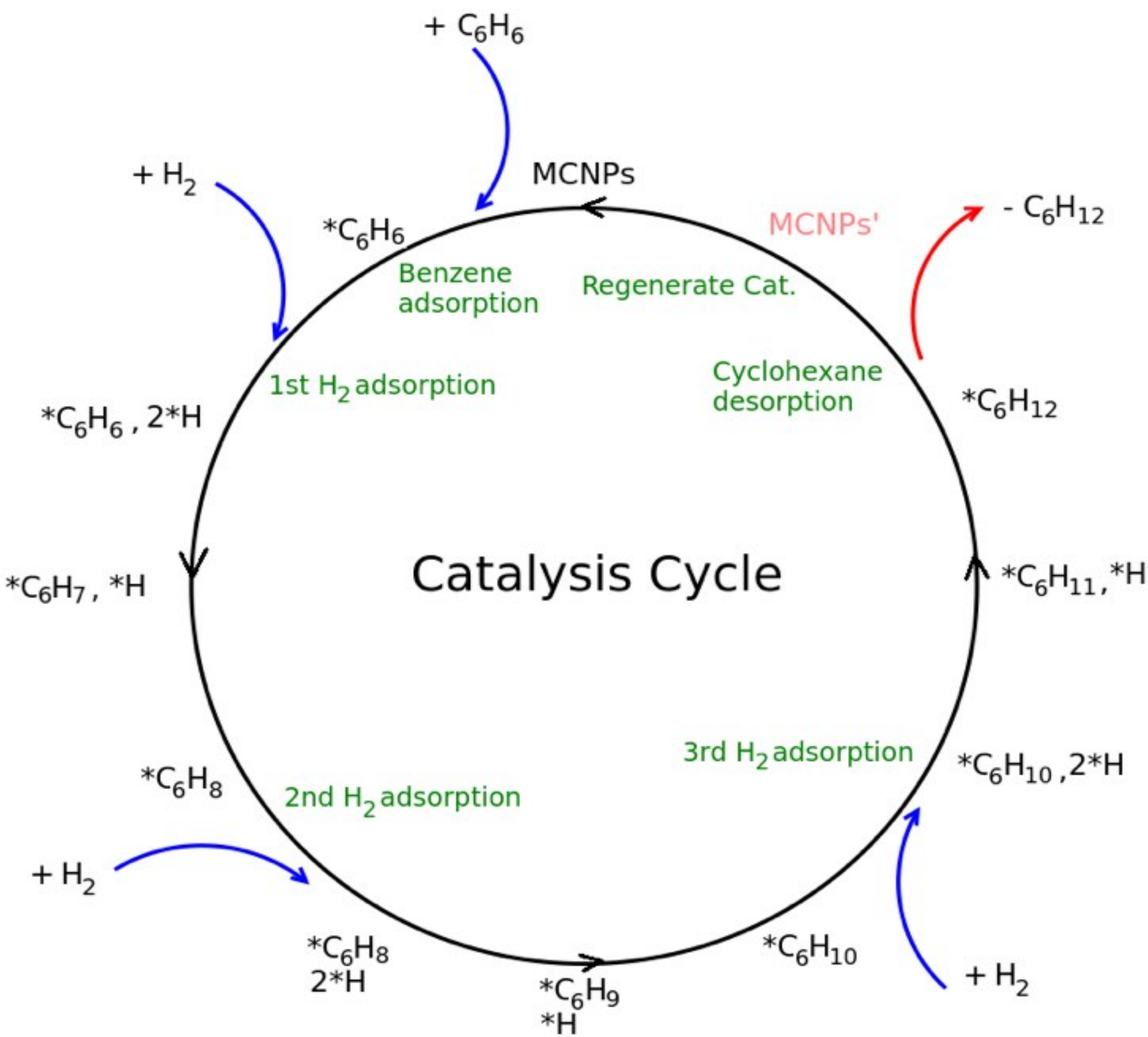
- **Benzene hydrogenation mechanism:** [Adsorption of Cyclic C_6H_6, C_6H_8, C_6H_{10} and C_6H_{12} on the (0001) surface of α-Mo_2C]
- **Parameterizing a faster quantum mechanical method:** [DFTB Parameterization of Mo, C, H, O and Si]
- **The role of entropy and the environment:** [Molybdenum Carbide Nanocatalysts at Work in the *in-situ* Environment: a DFTB and QM(DFTB)/MM Study]

X. Liu, D. R. Salahub et al, *J Phys Chem C* **2013,** *117* (14), 7069
X. Liu, D. R. Salahub et al, Theor.Chem.Acc., **2016,** 135:168, 1-14
X. Liu, D. R. Salahub et al, *J Am Chem Soc* 2015,137, 4249

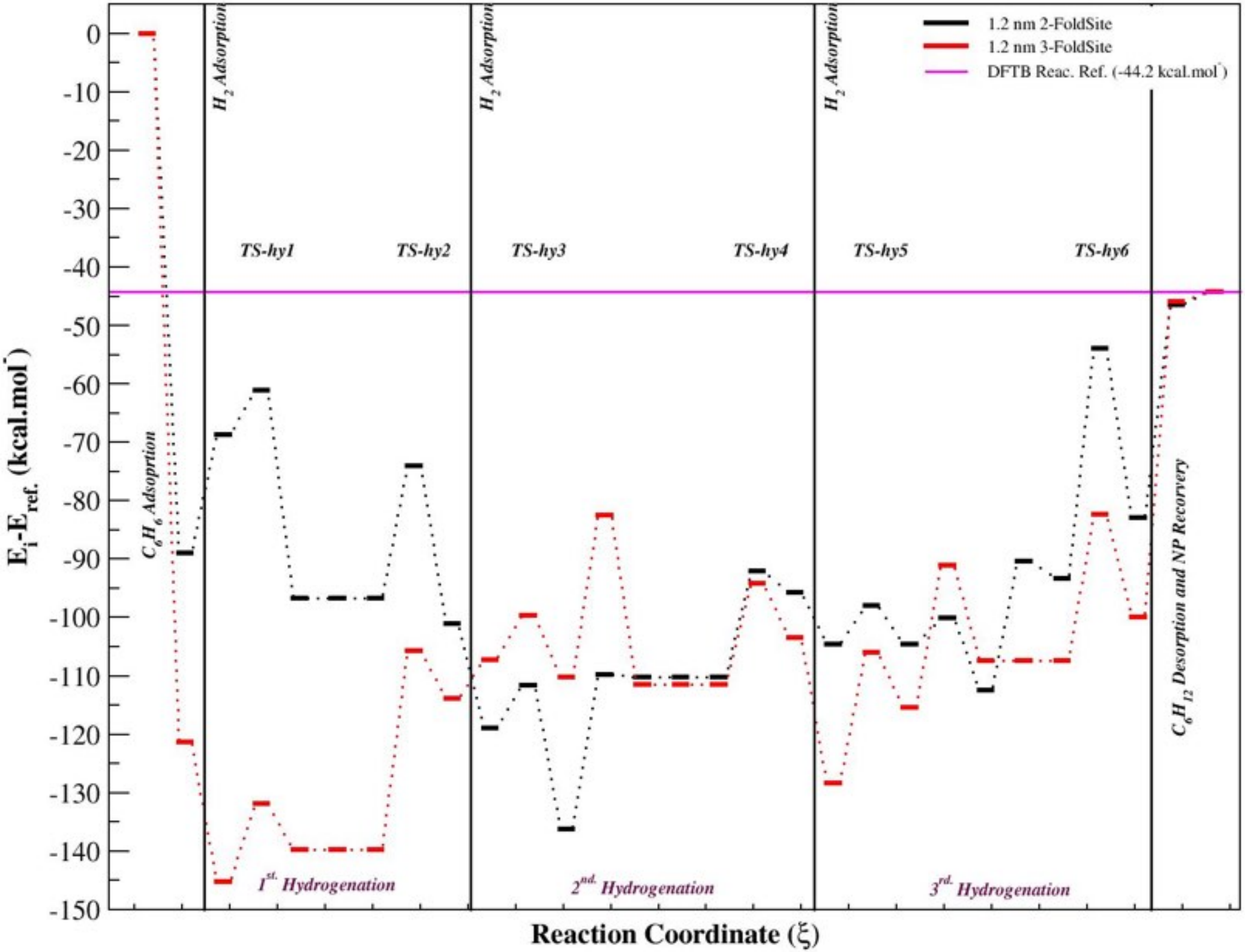
DFTB Benchmarking: Reaction Path Energies



X. Liu, T. Heine, and D. R. Salahub In preparation



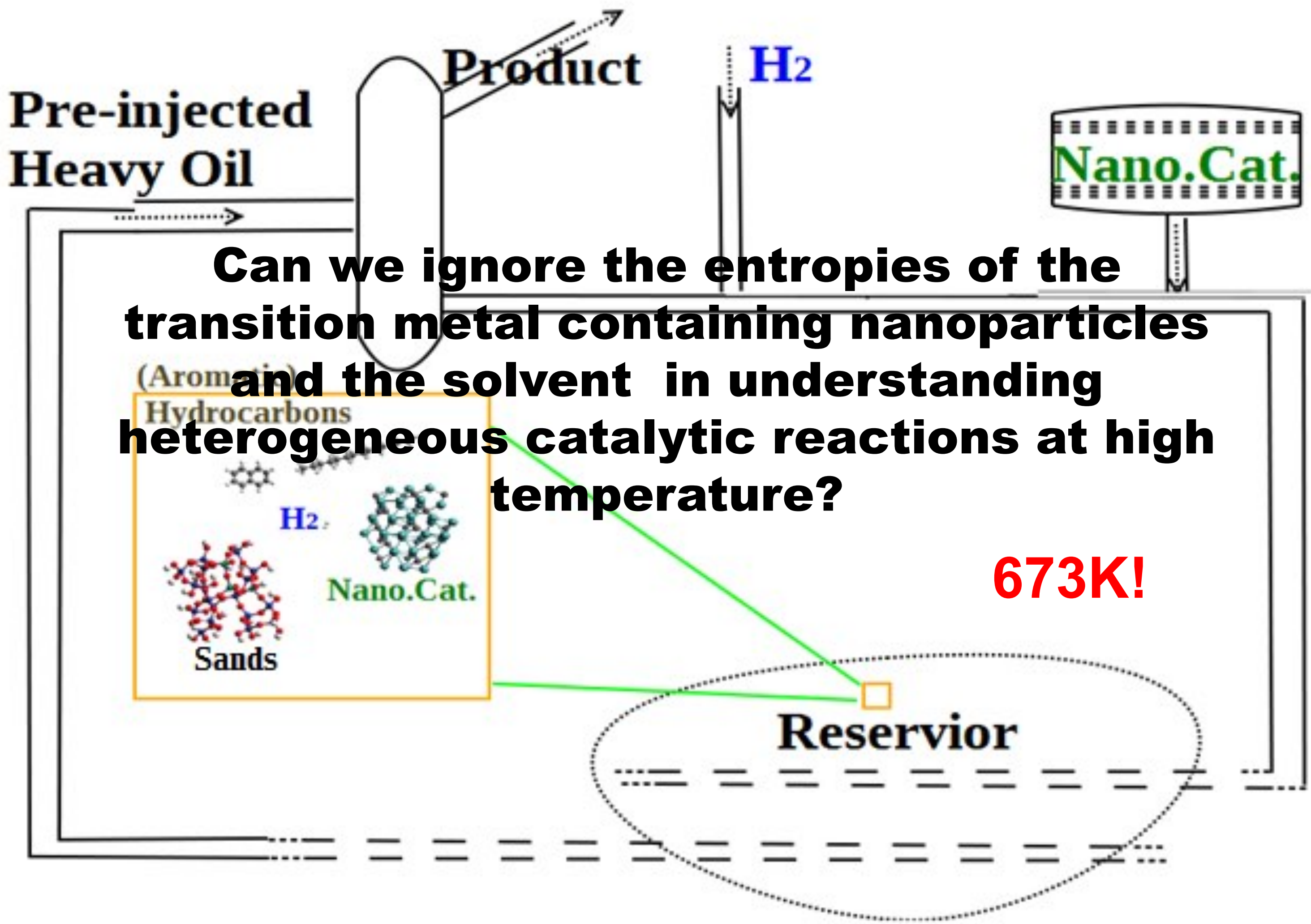
+ C_6H_6
MCNPs
MCNPs'
- C_6H_{12}
+ H_2
*C_6H_6
Benzene adsorption
Regenerate Cat.
1st H_2 adsorption
Cyclohexane desorption
*C_6H_{12}
*C_6H_6 , 2*H
*C_6H_7 , *H
Catalysis Cycle
*C_6H_{11} ,*H
3rd H_2 adsorption
*C_6H_{10} ,2*H
*C_6H_8
2nd H_2 adsorption
+ H_2
*C_6H_8
2*H
*C_6H_9
*H
*C_6H_{10}
+ H_2



1.2 nm 2-FoldSite
1.2 nm 3-FoldSite
DFTB Reac. Ref. (-44.2 kcal.mol-)
H2 Adsorption
H2 Adsorption
H2 Adsorption
C6H6 Adsoprtion
TS-hy1
TS-hy2
TS-hy3
TS-hy4
TS-hy5
TS-hy6
C6H12 Desorption and NP Recorvery
1st. Hydrogenation
2nd. Hydrogenation
3rd. Hydrogenation
Ei-Eref. (kcal.mol-)
Reaction Coordinate (ξ)
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
-120
-130
-140
-150

Entropy? The *in-situ* Environment?

Something is still missing!



Can we ignore the entropies of the transition metal containing nanoparticles and the solvent in understanding heterogeneous catalytic reactions at high temperature?

673K!

Outline

- **Benzene hydrogenation mechanism:** [Adsorption of Cyclic C_6H_6, C_6H_8, C_6H_{10} and C_6H_{12} on the (0001) surface of α-Mo_2C]
- **Parameterizing a faster quantum mechanical method:** [DFTB Parameterization of Mo, C, H, O and Si]
- **The role of entropy and the environment:** [Molybdenum Carbide Nanocatalysts at Work in the *in-situ* Environment: a DFTB and QM(DFTB)/MM Study]

X. Liu, D. R. Salahub et al, *J Phys Chem C* **2013,** *117* (14), 7069
X. Liu, D. R. Salahub et al, Theor.Chem.Acc., **2016,** 135:168, 1-14 .
X. Liu, D. R. Salahub et al, *J Am Chem Soc* **2015**, 137, 4249

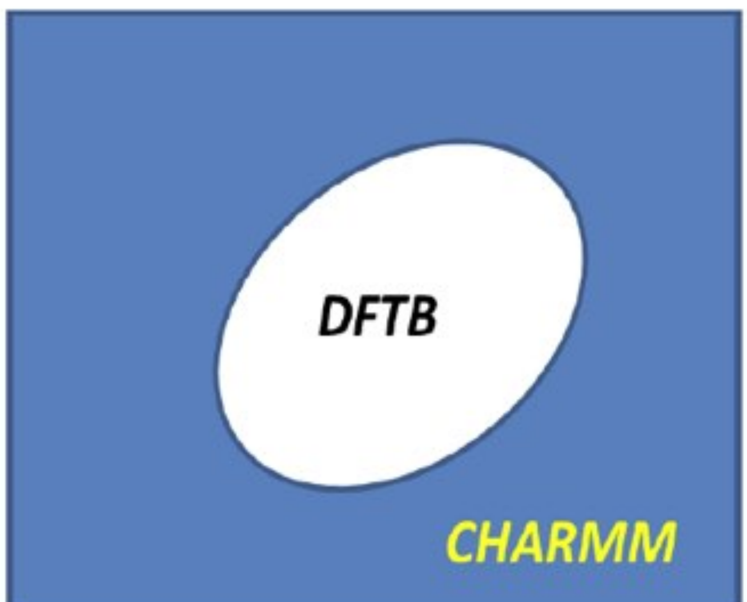
QM(DFTB)/MM Scheme

$$E^{tot} = E^{QM} + E^{MM} + E^{QM/MM}$$

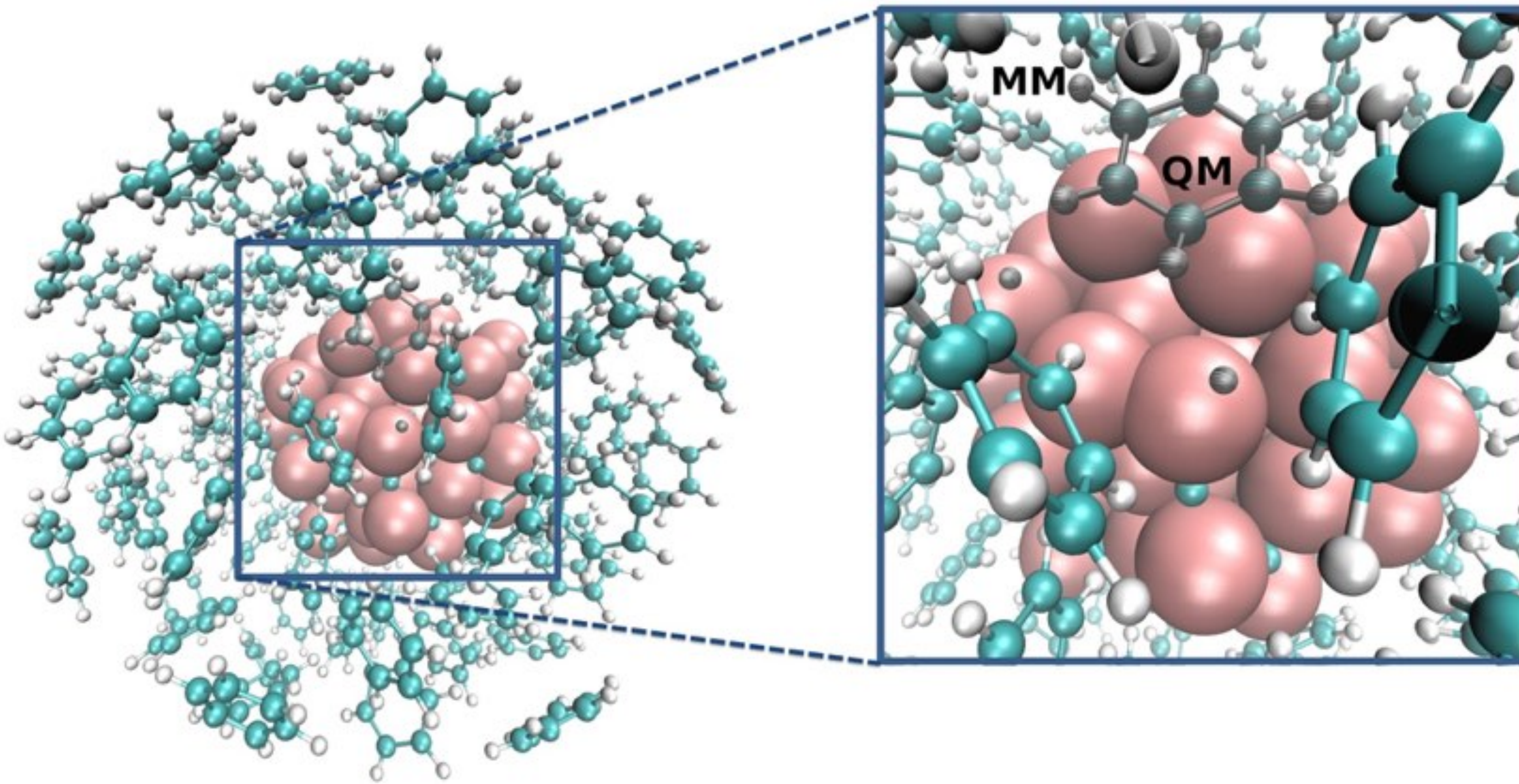
Warshel and Levitt, *J. Mol. Biol.* 1976

$$E^{tot} = \left\langle \Psi \middle| \widehat{H}^{QM} + \widehat{H}_{esd}^{QM/MM} \middle| \Psi \right\rangle + E_{vdw}^{QM/MM} + E^{MM}$$

Cui et al. *J Phys. Chem. B* **2000,** *105* (2), 569

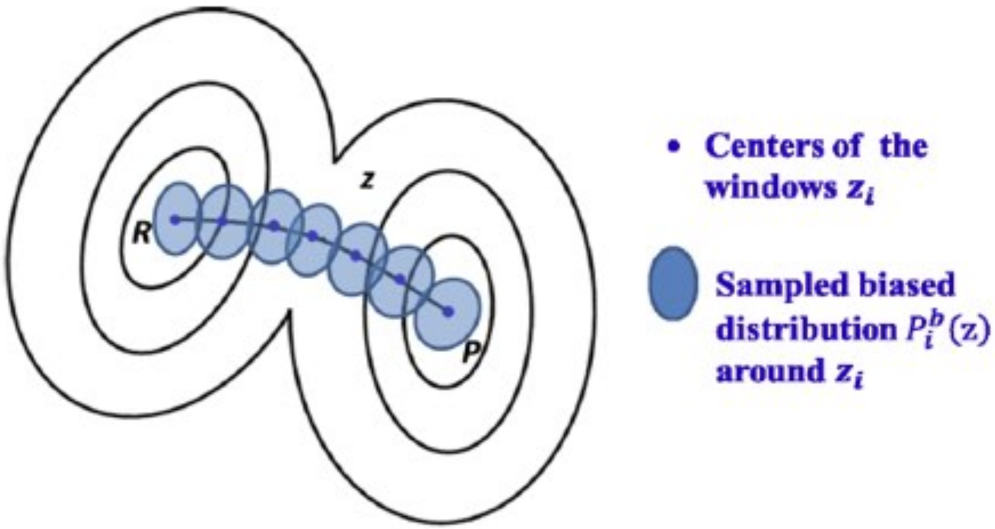


The QM/MM Model



A 1.2 nm MCNP embedded in 100 benzene molecules

Free Energy and Umbrella Sampling



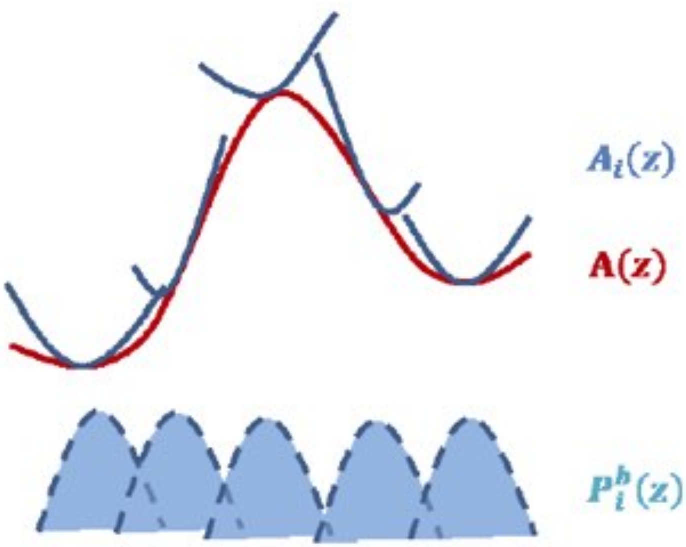
$$\mathrm{A} = -\frac{1}{\beta} \ln Q_{NVT}$$

$$P(\mathrm{z}) = \frac{\int e^{-\beta V(\boldsymbol{R})} \delta[z'(\boldsymbol{R}) - z] d\boldsymbol{R}}{\int e^{-\beta V(\boldsymbol{R})} d\boldsymbol{R}}$$

$$V_b(\boldsymbol{R}) = V_u(\boldsymbol{R}) + \omega_i(z)$$

$$P_i^b(\mathrm{z}) = \frac{\int e^{-\beta[V_u(\boldsymbol{R}) + \omega_i(z)]} \delta[z'(\boldsymbol{R}) - z] d\boldsymbol{R}}{\int e^{-\beta[V_u(\boldsymbol{R}) + \omega_i(z)]} d\boldsymbol{R}}$$

$$\mathrm{A}_i(\mathrm{z}) = -\frac{1}{\beta} \ln P_i^u(\mathrm{z}) = -\frac{1}{\beta} \ln P_i^b(\mathrm{z}) + \omega_i(z) + F_i$$



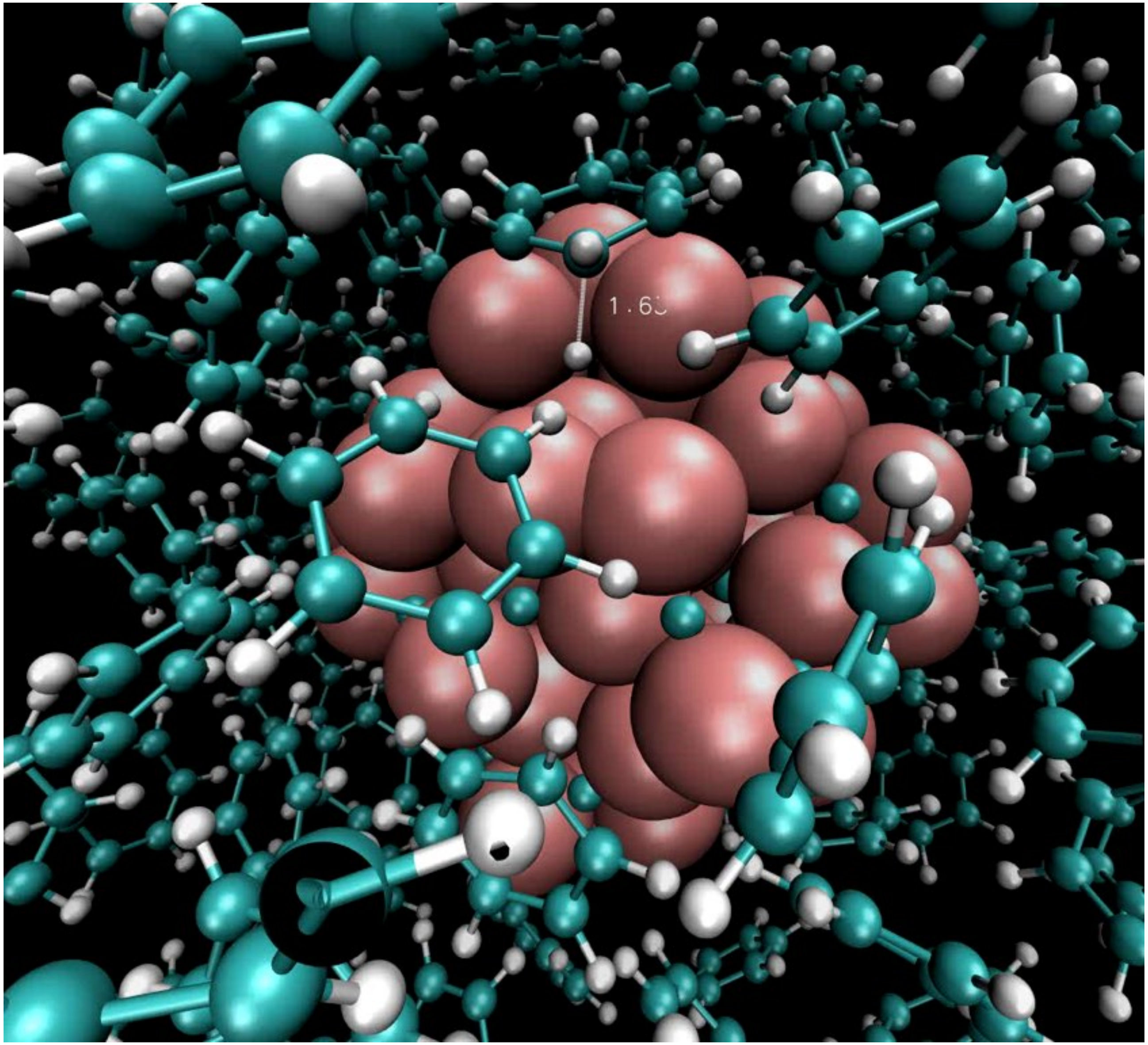
$$P^u(\mathrm{z}) = \frac{\sum_i^W N_i e^{-\beta\omega_i(z)}}{\sum_j^W N_j e^{[-\beta\omega_i(z) + \beta F_i]}}$$

$$e^{-\beta F_i} = \int P^u(\mathrm{z})\, e^{-\beta\omega_i(z)} dz$$

WHAM

QM/MM Umbrella Sampling (a typical window)

Addition of the 1st hydrogen



The Physical Quantities

Electronic (potential) energy **QM-NEB**

$$\mathrm{U} = E_e + E_t(T) + E_r(T) + E_v(T)$$ **Internal Energy**

$$\mathrm{S} = S_e + S_t + S_r + S_v$$ **Entropy (harmonic)**

$$\mathrm{A} = \mathrm{U} - \mathrm{TS}$$ **Free Energy (harmonic)**

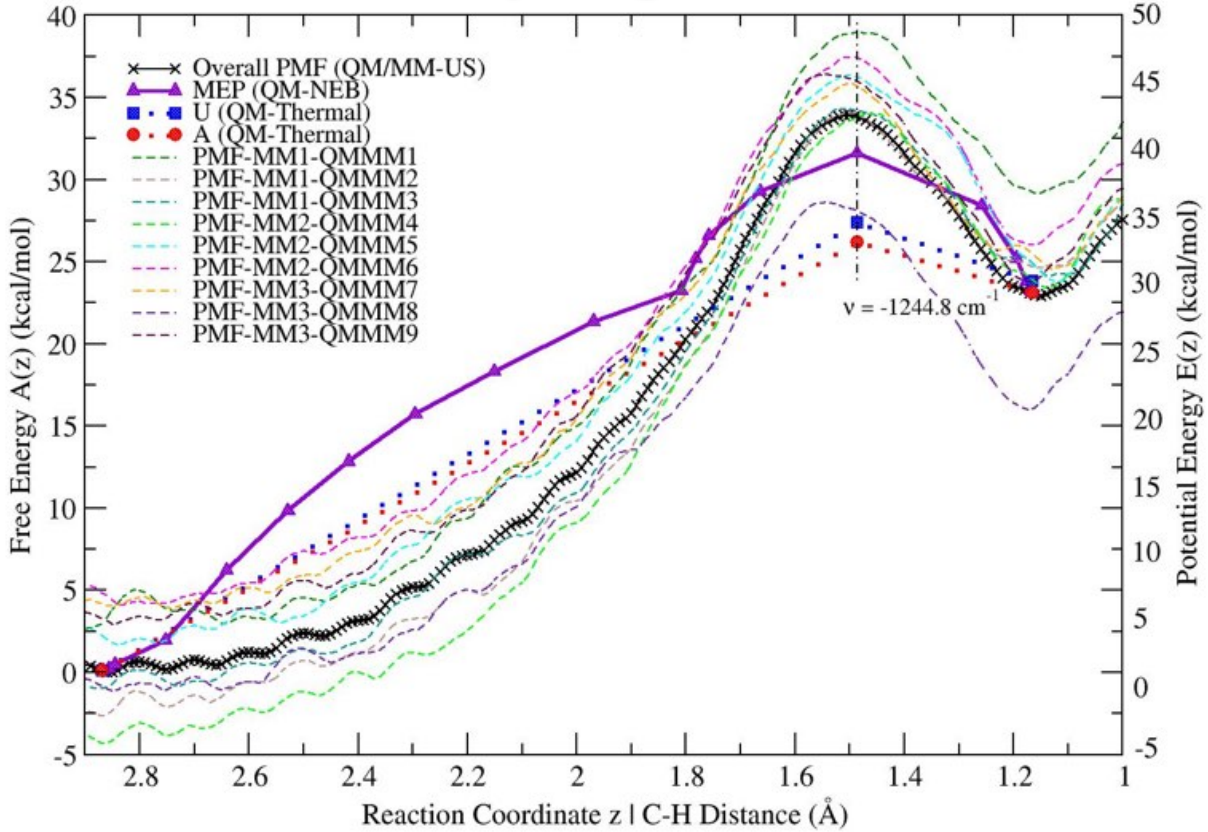
Traditional Approach

$$A' = -\frac{1}{\beta} lnQ_{NVT}$$ **Free Energy (anharmonic)**

Multiscale modelling Approach

Average PMF

The Second Hydrogenation Reaction



States	*E(QM-NEB)*	*U(QM-Thermal)*	*A(QM-Thermal)*	*A'(QM/MM-US)*	*A'- A*
R(QM-NEB)	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
TS	**31.5**	**27.3**	**26.1**	**33.8**	**7.7**
P	**23.0**	**23.7**	**23.0**	**10.9**	**12.1**

Final Remarks - important concepts

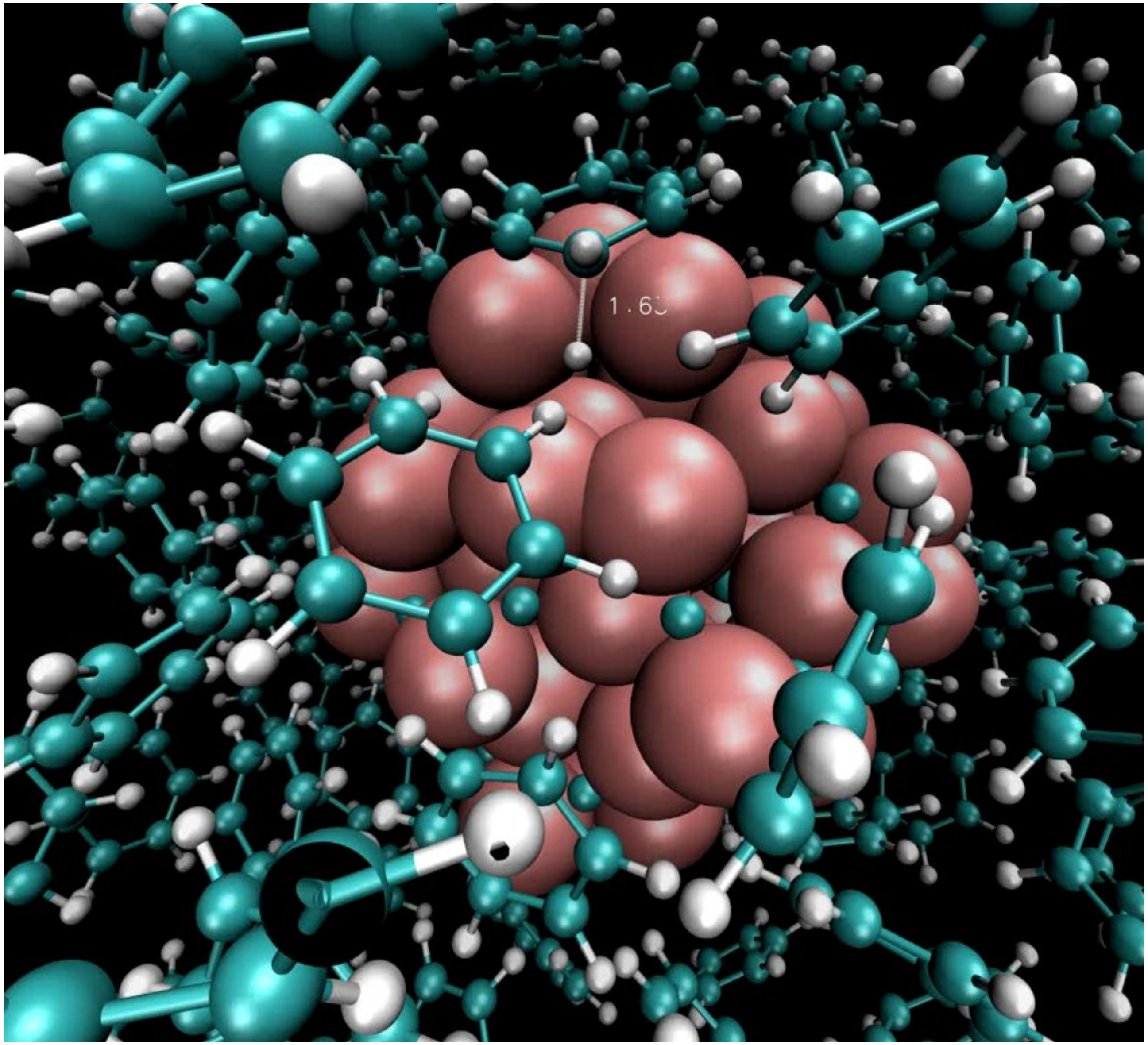
- ✓ **Benzene hydrogenation on molybdenum carbide: Langmuir-Hinshelwood mechanism.**
- ✓ **MCNPs: metallic nanoparticles.**
- ✓ **The key to improve the catalytic activity: controlling the morphology of the MCNPs.**
- ✓ **The MCNPs are flexible under working conditions.**
- ✓ **The entropic (including anharmonic entropy) effect and the solvent environment are crucial.**
- ✓ **Nanoparticles are neither clusters nor bulk**
- ✓ **Need a new paradigm for nanocatalysis that includes flexibility, anharmonicity and entropy**

How to choose/find reaction coordinates (collective variables)?

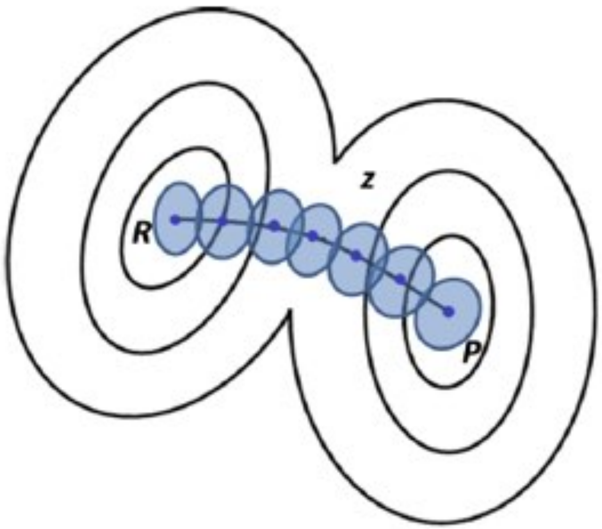
?

QM/MM Umbrella Sampling (a typical window)

Addition of the 1st hydrogen



Free Energy and Umbrella Sampling



- **Centers of the windows z_i**

Sampled biased distribution $P_i^b(\mathrm{z})$ around z_i

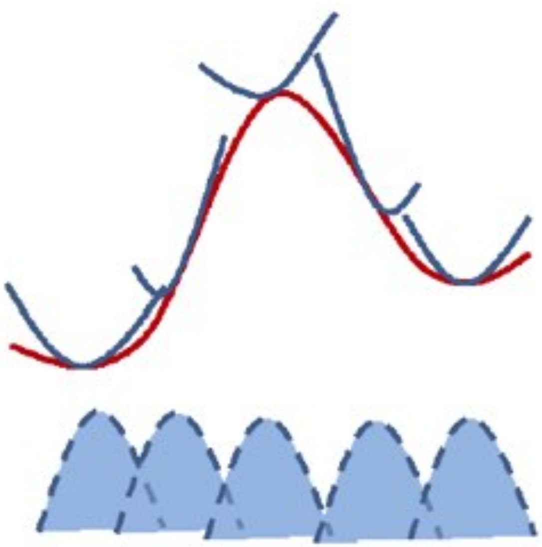
$$\mathrm{A} = -\frac{1}{\beta} ln Q_{NVT}$$

$$P(\mathrm{z}) = \frac{\int e^{-\beta V(\boldsymbol{R})} \delta[\mathrm{z}'(\boldsymbol{R}) - \mathrm{z}] d\boldsymbol{R}}{\int e^{-\beta V(\boldsymbol{R})} d\boldsymbol{R}}$$

$$V_b(\boldsymbol{R}) = V_u(\boldsymbol{R}) + \omega_i(\mathrm{z})$$

$$P_i^b(\mathrm{z}) = \frac{\int e^{-\beta[V_u(\boldsymbol{R}) + \omega_i(\mathrm{z})]} \delta[\mathrm{z}'(\boldsymbol{R}) - \mathrm{z}] d\boldsymbol{R}}{\int e^{-\beta[V_u(\boldsymbol{R}) + \omega_i(\mathrm{z})]} d\boldsymbol{R}}$$

$$\mathrm{A}_i(\mathrm{z}) = -\frac{1}{\beta} ln P_i^u(\mathrm{z}) = -\frac{1}{\beta} ln P_i^b(\mathrm{z}) + \omega_i(\mathrm{z}) + F_i$$



$\boldsymbol{A_i(z)}$

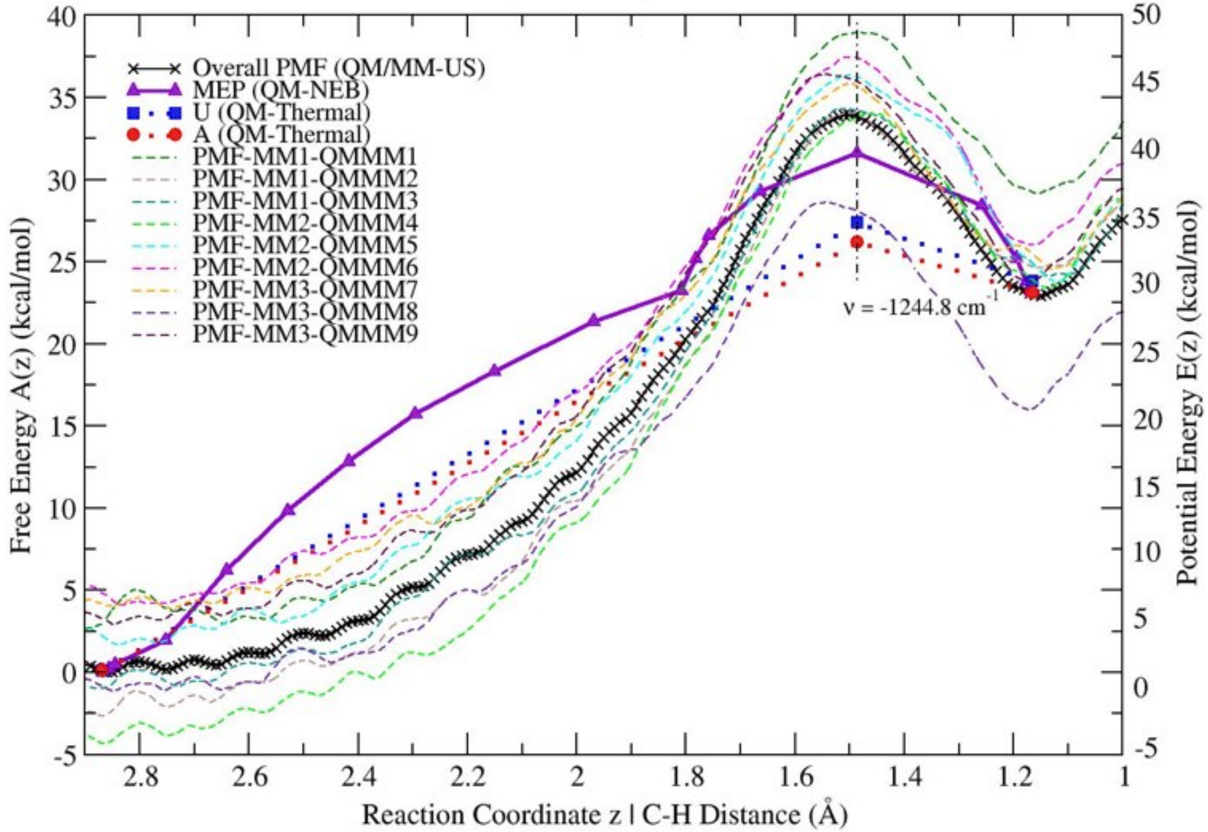
$\mathbf{A(z)}$

$\boldsymbol{P_i^b(z)}$

$$\boxed{P^u(\mathrm{z}) = \frac{\sum_i^W N_i e^{-\beta\omega_i(\mathrm{z})}}{\sum_j^W N_j e^{[-\beta\omega_i(\mathrm{z}) + \beta F_i]}} \qquad e^{-\beta F_i} = \int P^u(\mathrm{z})\, e^{-\beta\omega_i(\mathrm{z})} dz}$$

WHAM

The Second Hydrogenation Reaction



States	*E(QM-NEB)*	*U(QM-Thermal)*	*A(QM-Thermal)*	*A'(QM/MM-US)*	*A'- A*
R(QM-NEB)	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
TS	**31.5**	**27.3**	**26.1**	**33.8**	**7.7**
P	**23.0**	**23.7**	**23.0**	**10.9**	**12.1**

A sampling of enhanced sampling methods

1. Umbrella sampling (Torrie, Valleau, 1977)
2. Transition path sampling (Bolhuis, Chandler, 1998)
3. Replica exchange (Parallel tempering) (Sugita, 1999)
4. Meta-dynamics (Laio, Parrinello, 2002)
5. String method (Vanden Eijnden, 2002)
6. Path collective variables (Parrinello, 2007)
7. Adaptive Biasing Force (Pohorille, 2008, Klein, 2010)
8. Free-energy without CVs (Laio, 2018)
9. Time-lagged autoencoders/deep learning (Noé, 2018)
10. CV discovery with deep Bayesian models (Schöberl, 2018)

Predictive Collective Variable Discovery with Deep Bayesian Models

Markus Schöberl,[1,2,a)] Nicholas Zabaras,[1,b)] and Phaedon-Stelios Koutsourelakis[2,c)]

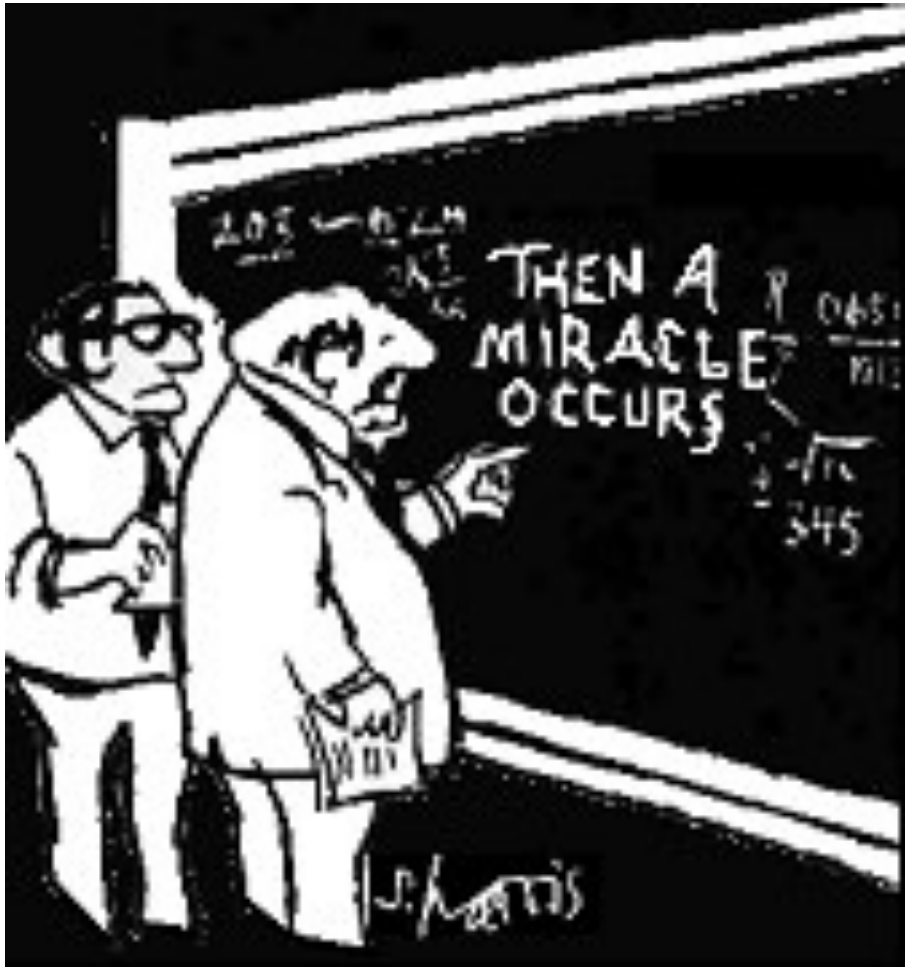
[1)] *Center for Informatics and Computational Science, University of Notre Dame, 311 Cushing Hall, Notre Dame, IN 46556, USA.*

[2)] *Continuum Mechanics Group, Technical University of Munich, Boltzmannstraße 15, 85748 Garching, Germany.*

(Dated: 20 September 2018)

Extending spatio-temporal scale limitations of models for complex atomistic systems considered in biochemistry and materials science necessitates the development of enhanced sampling methods. The potential acceleration in exploring the configurational space by enhanced sampling methods depends on the choice of collective variables (CVs). In this work, we formulate the discovery of CVs as a Bayesian inference problem and consider the CVs as hidden generators of the full-atomistic trajectory. The ability to generate samples of the fine-scale atomistic configurations using limited training data allows us to compute estimates of observables as well as our probabilistic confidence on them. The methodology is based on emerging methodological advances in machine learning and variational inference. The discovered CVs are related to physicochemical properties which are essential for understanding mechanisms especially in unexplored complex systems. We provide a quantitative assessment of the CVs in terms of their predictive ability for alanine dipeptide (ALA-2) and ALA-15 peptide.

J. Chem. Phys. **150**, 024109 (2019)



THEN A
MIRACLE
OCCURS

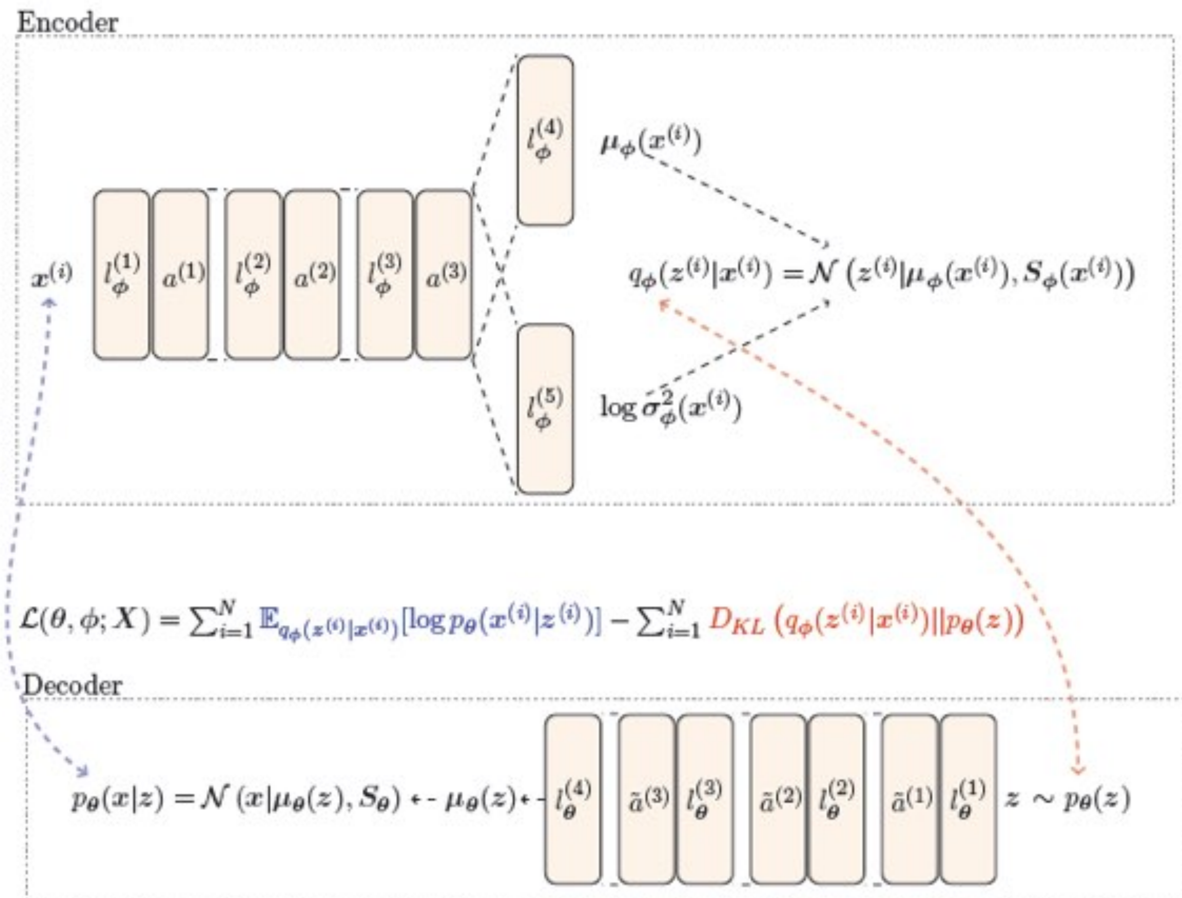
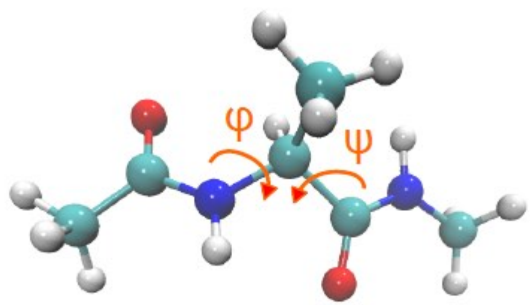


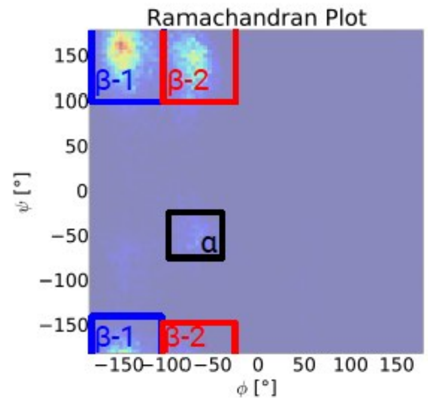
FIG. 2. Schematic of the AEVB depicting the employed network architecture. Fully connected linear layers are denoted with $l^{(i)}$ and non-linear activation functions with $a^{(i)}$. The indices ϕ and θ indicate encoding and decoding networks, respectively. The maximization of the lower-bound on the log-likelihood $\mathcal{L}(\theta, \phi; X)$ in Eq. (11) simultaneously optimizes the parametrization of the encoder and decoder. The first term in $\mathcal{L}(\theta, \phi; X)$ accounts for the reconstruction of the training data $x^{(i)}$ with $z^{(i)}$ distributed according $q_\phi(z^{(i)}|x^{(i)})$. The second term, in aggregation of all data $x^{(i)}$, ensures that $q_\phi(z^{(i)}|x^{(i)})$ is close to $p(z)$.

1. Simulation of ALA-2

Alanine dipeptide consists of 22 atoms leading to $\dim(\boldsymbol{x}) = 66$ in a Cartesian representation comprising the coordinates of *all* atoms which we will use later on as the model input. It is well-known that ALA-2 exhibits distinct conformations which are categorized depending on the dihedral angles (ϕ, ψ) (as indicated in Fig. 1(a)) of the atomistic configuration. We label the three characteristic modes as α, β-1, and β-2 in accordance with [104] (see Fig. 1(b)).



(a) ALA-2 peptide with indicated dihedral angels.



(b) Characteristic conformations and their labelling as used in the sequel.

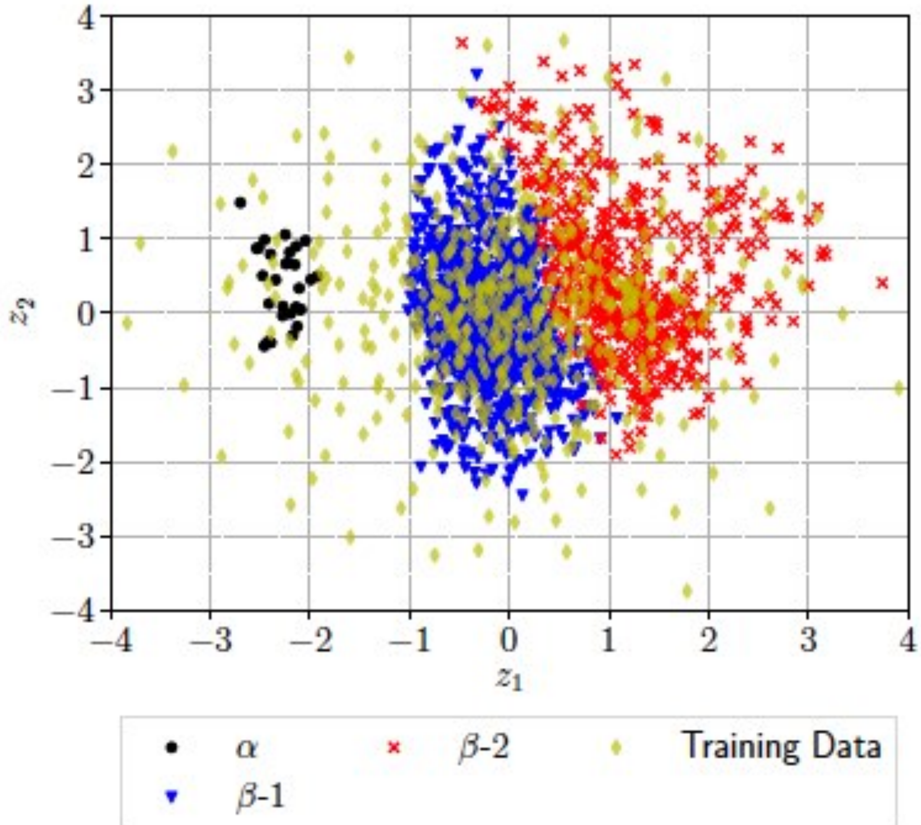
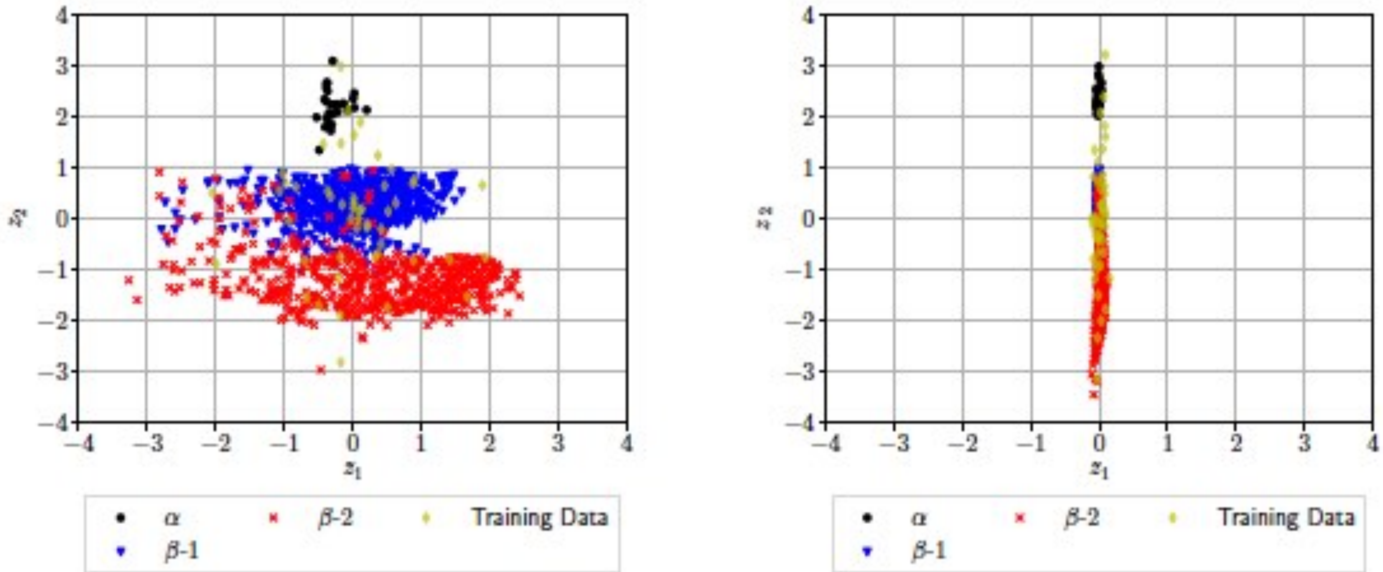


FIG. 5. Representation of the $\boldsymbol{z}$-coordinates of the training data $\boldsymbol{X}$ with $N = 500$ in the CV space (yellow diamonds). Using the trained model and the mean of $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{z})$ we computed the $\boldsymbol{z}$-coordinates of 1527 test samples corresponding to different conformations of the alanine dipeptide to α (black), β-1 (blue), and β-2 (red). Without any prior physical information, the encoder yields three distinct clusters in the CV space.



(a) Active ARD prior.

(b) Without ARD prior.

FIG. 6. Representation of the $\boldsymbol{z}$-coordinates of the training data $\boldsymbol{X}$ with $N = 50$ in the CV space (yellow diamonds). Using the trained model and the mean of $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{z})$ we computed the $\boldsymbol{z}$-coordinates of 1527 test samples corresponding to different conformations of the alanine dipeptide to α (black), β-1 (blue), and β-2 (red). In the case of limited training data, the ARD prior facilitates the identification of physically meaningful CVs (left) compared to the representation on the right obtained without the ARD prior. Note that the changed positioning of the conformations in the CV space compared to Fig. 5 is due to symmetries in $p_{\boldsymbol{\theta}}(\boldsymbol{z})$.

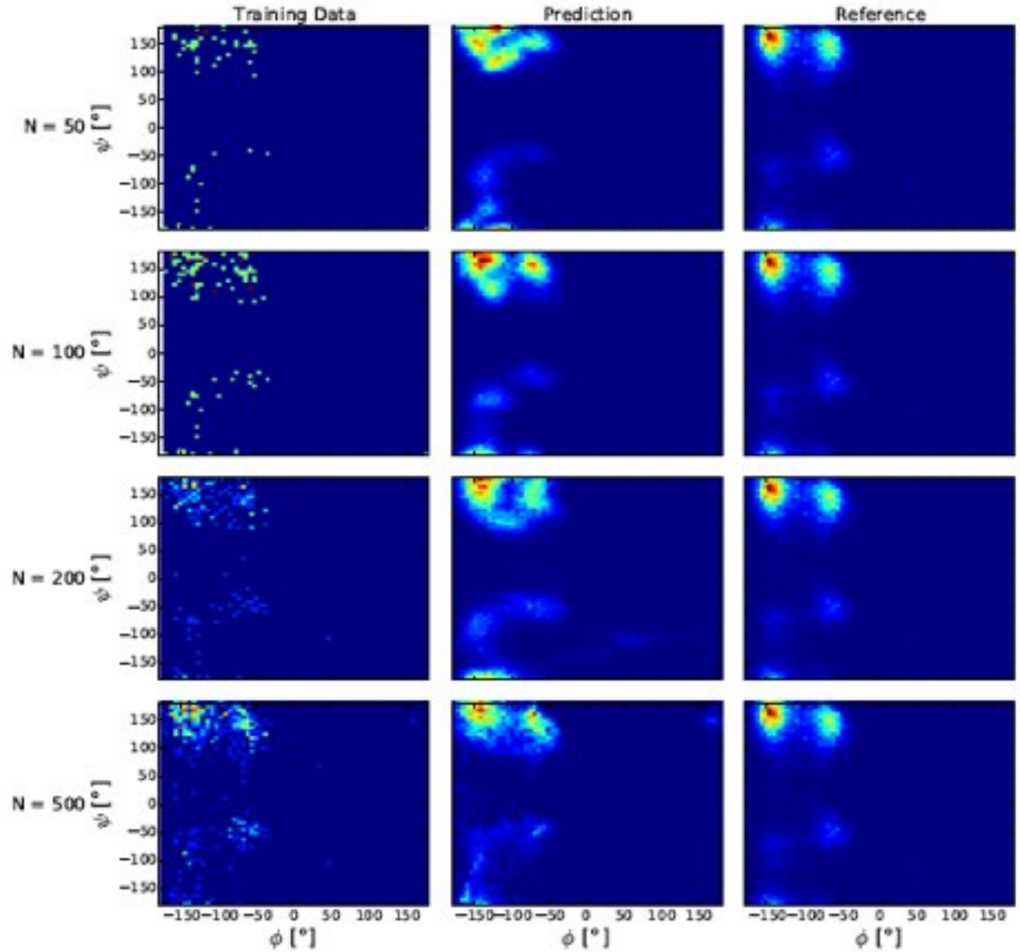


FIG. 7. Ramachandran plot estimated with the training data $\boldsymbol{X}$ (left column), using predictions of the trained model (middle column), and the reference (right column, estimated with $N = 10,000$). Each row refers to different size N of training datasets (the figure on the right column is repeated to allow easy comparison with the results on the first two columns). The represented predictions are obtained by applying Algorithm [2] with $T = 10,000$ samples. The generative nature of the model allows more accurate estimates than when using the training data alone. In addition, the Bayesian approach allows for predictions with their associated uncertainties as discussed subsequently.

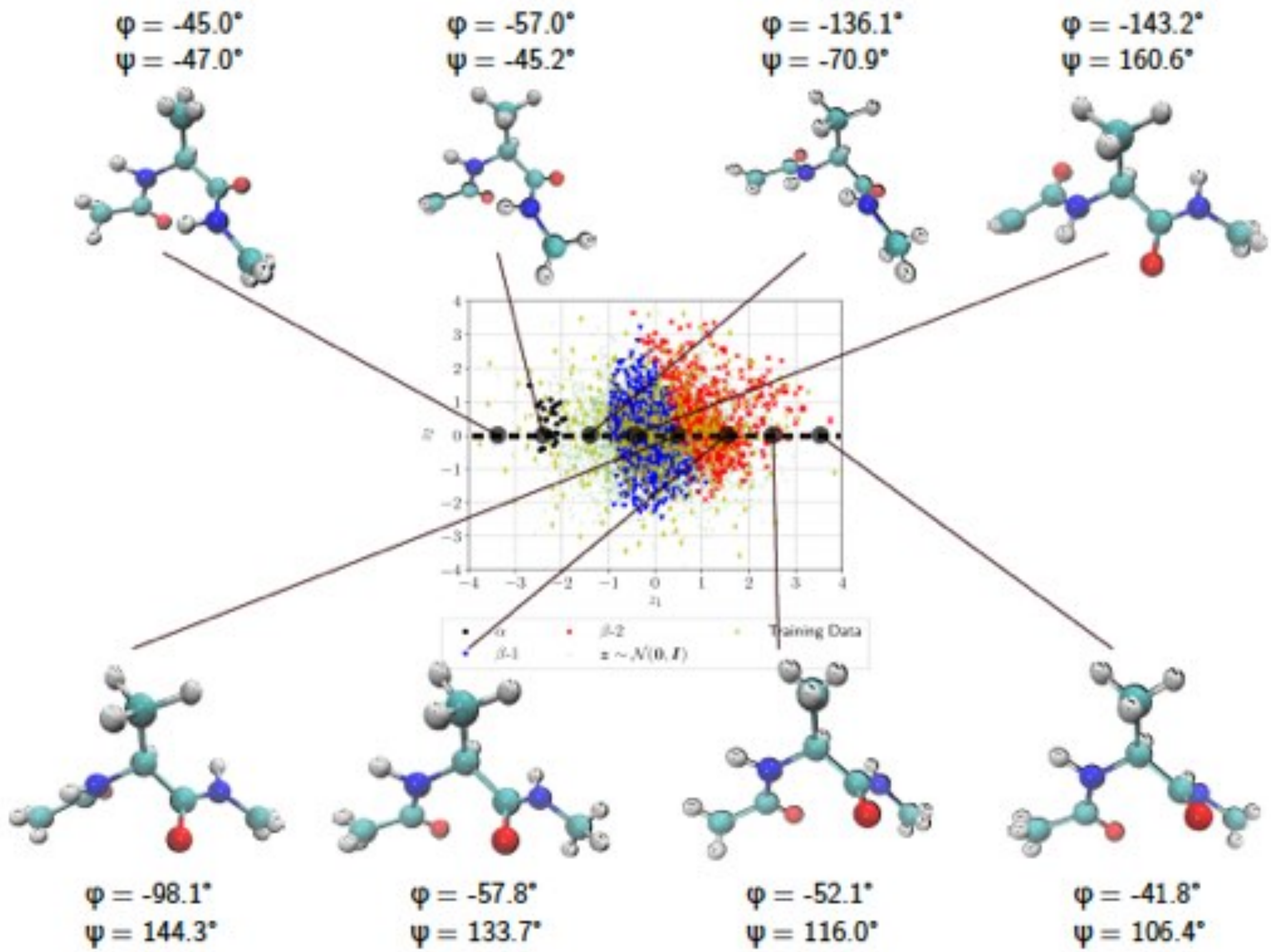
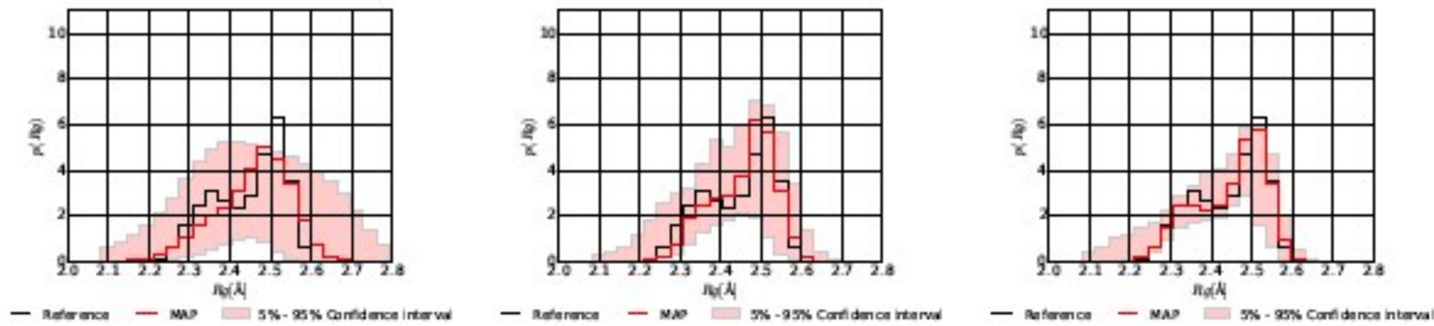


FIG. 8. *Predicted* configurations $\boldsymbol{x}$ (including dihedral angle values) for $\{\boldsymbol{z}|z_1 = \{-3.5, -2.5, \ldots, 3.5\}, z_2 = 0\}$ with $\boldsymbol{\mu_\theta}(\boldsymbol{z})$ of $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})$. As one moves along the z_1 axis, we obtain for the given CVs atomistic configurations $\boldsymbol{x}$ reflecting the conformations α, β-1, and β-2. Rendered atomistic representations are created by VMD[124].



(a) $N = 50$. (b) $N = 200$. (c) $N = 500$.

FIG. 12. Predicted radius of gyration with $\dim(\boldsymbol{z}) = 2$ for various sizes N of the training dataset. The MAP estimate indicated in red is compared to the reference (black) solution. The latter is estimated by $N = 10,000$. The shaded area represents the 5%-95% confidence interval, reflecting the induced epistemic uncertainty from the limited amount of training data.

B. ALA-15

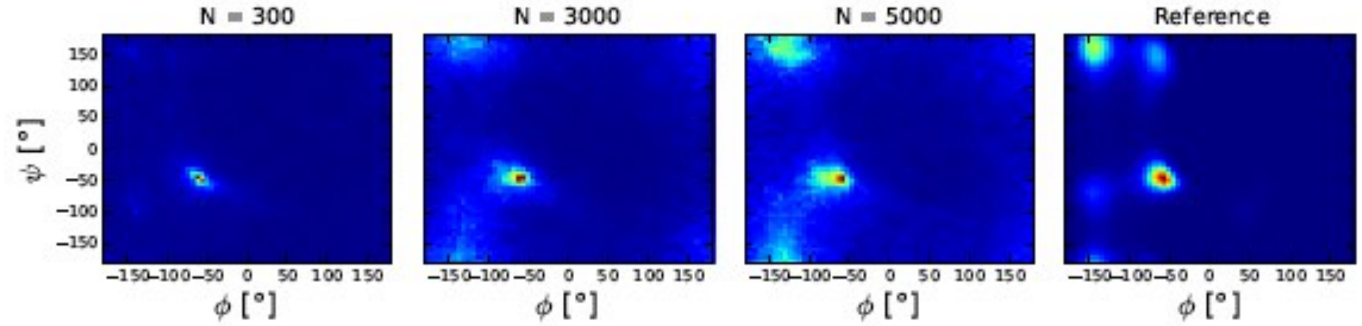
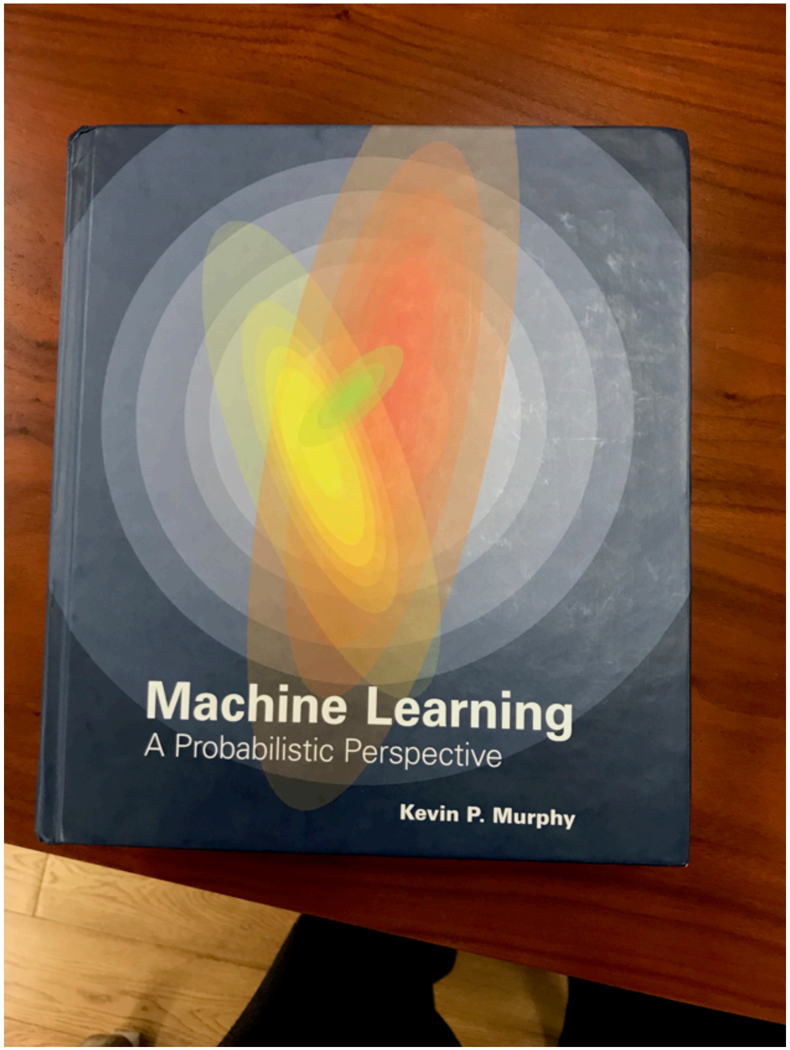
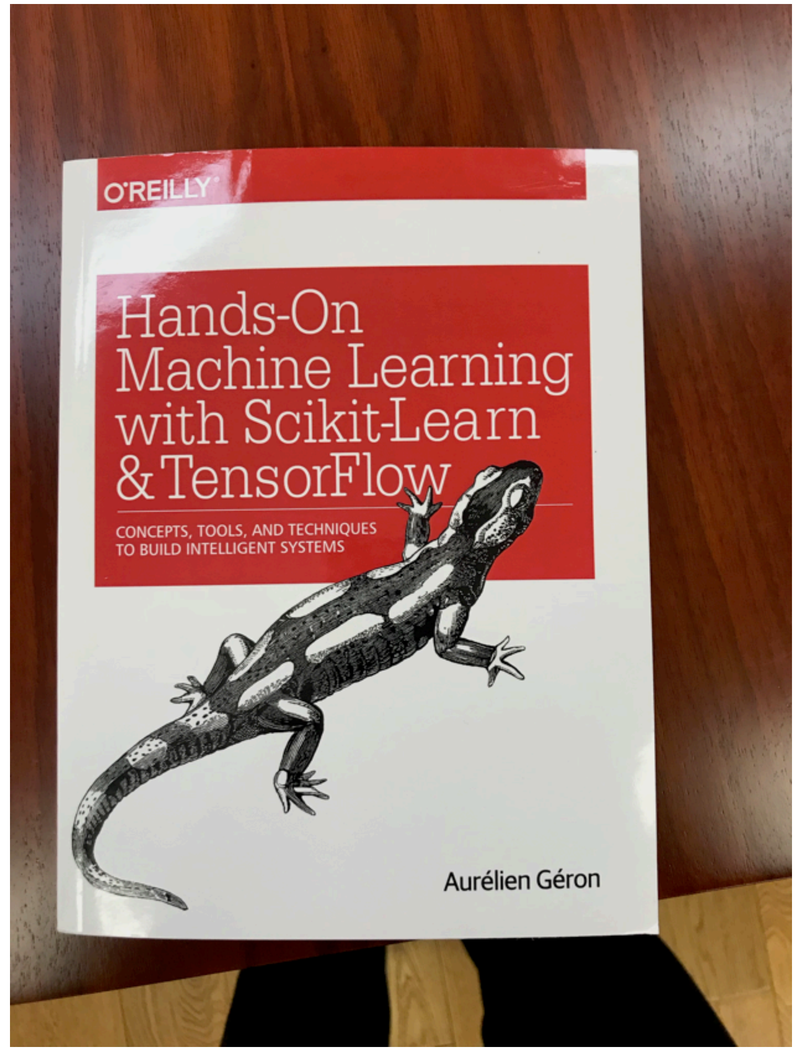
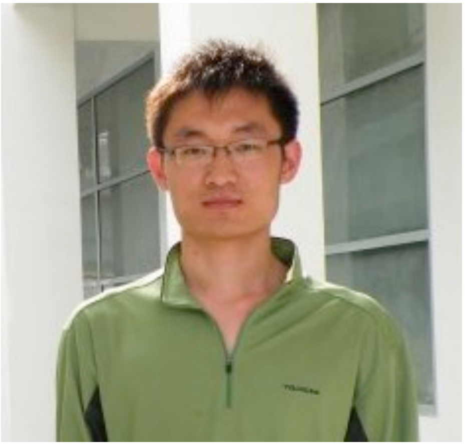


FIG. 15. Predicted Ramachandran plot with $\dim(\boldsymbol{z}) = 2$ for various sizes N of the training dataset (first three plots from the left). Depicted predictions are MAP estimates based on $T = 10,000$ samples. The plot on the right is the reference MD prediction with $N = 10,000$ configurations.

findingcollectivevariables-l@mailman.ucalgary.ca







Xingchen Liu

Baojing Zhou

Alex Tkalych

Farouq Ahmed

Pedro Pereira

Andreas Köster

Helio Duarte

Acknowledgement

Cinvestav

IPICYT
INSTITUTO POTOSINO DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA, A.C.

PEMEX

UNIVERSITY OF CALGARY