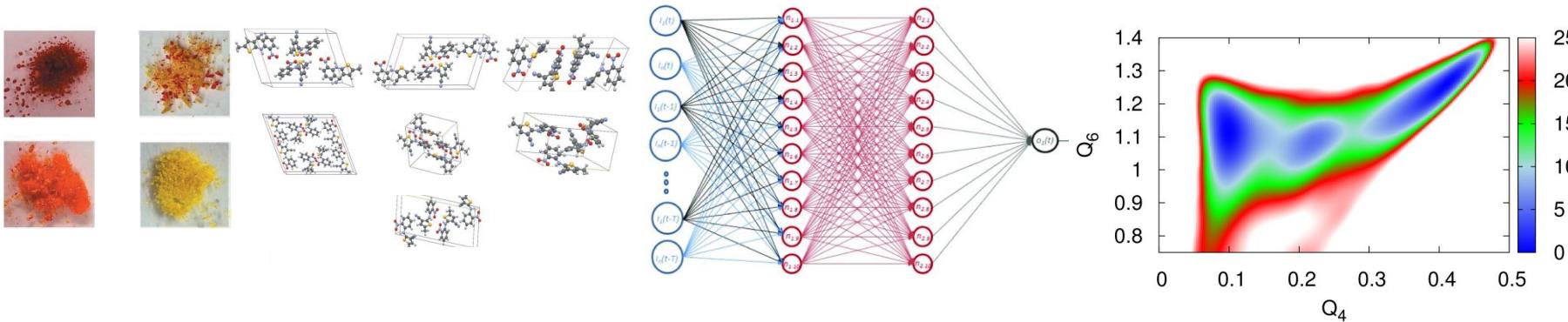


# *Molecular simulation and Machine Learning as Routes to Exploring Structure and Phase Behavior in Atomic and Molecular Crystals*



**CECAM Summer School Seminar  
SISSA/Trieste  
June 12, 2019**

**Mark E. Tuckerman**

*Dept. of Chemistry and Courant Institute of Mathematical Sciences*

*New York University, 100 Washington Square East, NY 10003*

*NYU-ECNU Center for Computational Chemistry at NYU Shanghai 200062, China*

**华东师范大学-纽约大学 计算化学联合研究中心**



**NEW YORK UNIVERSITY MRSEC**





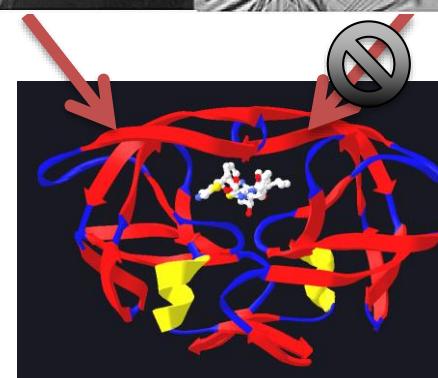
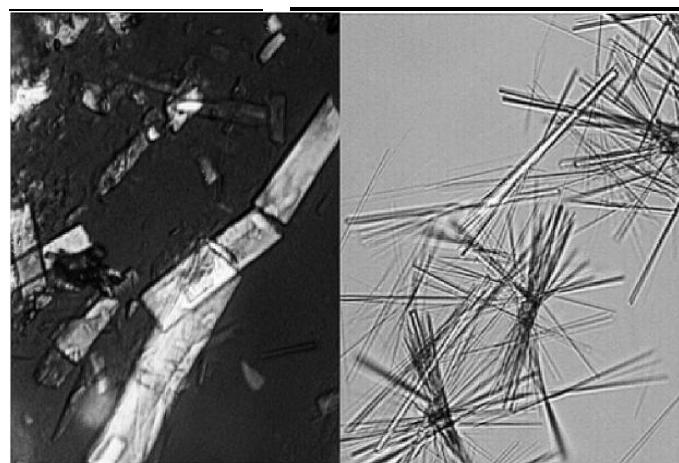
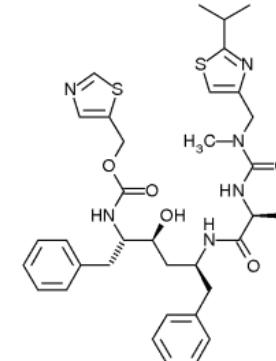
# Polymorphism and Pharmaceuticals: Ritonavir



***Polymorphism*** refers to the ability of a compound to form multiple crystal structures.

## Ritonavir: HIV protease inhibitor

- NIH: \$3,500,000 investment, Abbott Labs (now Abbvie): \$200,000,000 investment.
  - Originally dispensed in 1996 as ordinary capsules, no refrigeration required
  - Converted to lower energy (hitherto unknown) polymorph (form I to form II) on the shelf
  - Form II: Low free energy,  
poor solubility, lower bioavailability
  - Required recall (1998) and reformulation as  
gel caps (2002).
  - Estimates of 15% to 45% of pharmaceutical  
compounds could have late-appearing  
polymorphs [Neumann, *Faraday Disc.* (2018)].



## HIV-1 protease



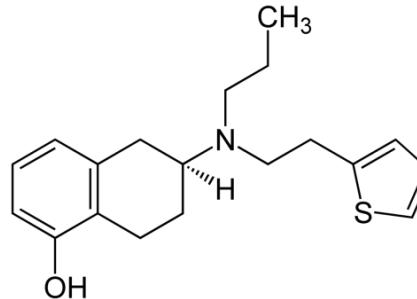
# Polymorphism and Pharmaceuticals: Rotigotine



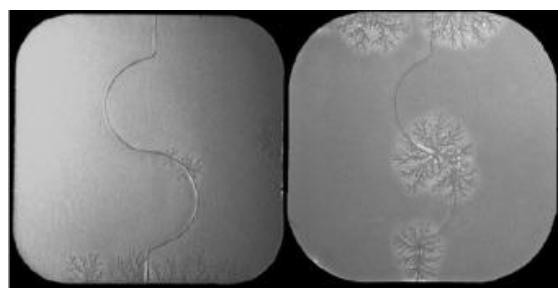
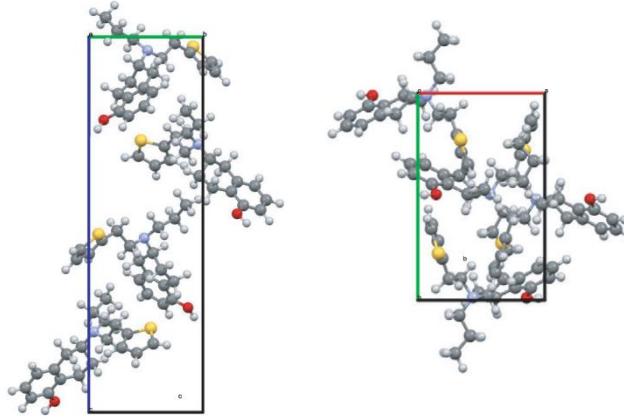
Not “Rigatoni”



Rotigotine



Used to relieve symptoms from Parkinson's

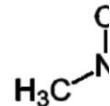


Rietveld & Ceolin *J. Pharm. Sci.* **105**, 4117 (2015)

- Developed in 1985 at U. Groningen.
- Licensed to Schwarz Pharma in 1998.
- Sold in Europe in 2006 and approved for use in the US in 2007.
- Until 2007, only form I was known, but soon, snowflake like patches appeared on patches, reducing efficacy.
- Identified as a late-appearing polymorph, now called form II.
- Recalled in 2008 and was unavailable until 2012, amorphous matrix form was reintroduced in the US.



# Pharmaceutical Polymorphism: Zantac



United States Court of Appeals, Federal Circuit.

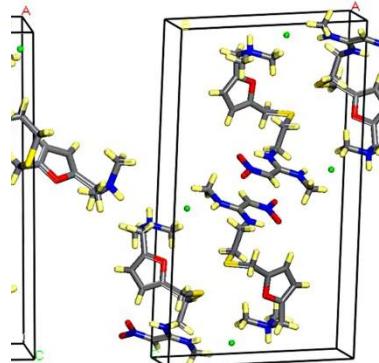
GLAXO INC. and Glaxo Group Limited, Plaintiffs-Appellees, v. NOVOPHARM LTD., Defendant-Appellant.

No. 94-1026.

Decided: April 21, 1995

Before ARCHER, Chief Judge, RICH, and MAYER, Circuit Judges. Stephen B. Judlowe, Hopgood, Calimafde, Kalil, Blaustein & Judlowe, New York City, argued for plaintiffs-appellees. With him on the brief were William G. Todd and Janet B. Linn. Also on the brief were Steven P. Lockman, Arnold & Porter, Washington, DC, and Joseph W. Eason, Moore & Van Allen, Raleigh, NC, of counsel. Robert F. Green, Leydig, Voit & Mayer, LTD., Chicago, IL, argued for defendant-appellant. With him on the brief were John E. Rosenquist and Jeffrey S. Ward.

Novopharm Ltd. (Novopharm) appeals the judgment of the United States District Court for the Eastern District of North Carolina, Glaxo, Inc. v. Novopharm Ltd., 830 F.Supp. 871, 29 USPQ2d 1126 (E.D.N.C.1993), that United States Patent No. 4,521,431 was not invalid and was infringed, and enjoining Novopharm from the commercial manufacture or sale of the patented crystalline form of ranitidine hydrochloride. We affirm.



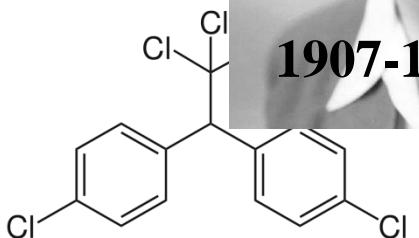
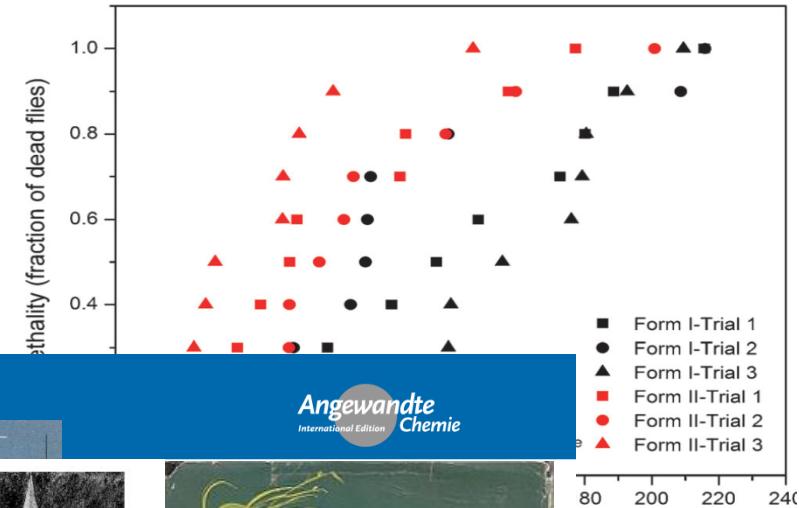
Form II

## Ranitidine hydrochloride (Zantac)

- Form I patented ('658) by GSK in 1978.
- In 1980, a new polymorph (form II) unexpectedly obtained that was also bioactive.
- GSK filed for a 2<sup>nd</sup> patent ('431) for form II.
- In 1991, sales of Zantac reached \$3.5 billion.
- A generic company, Novopharm sought to market form I at the expiration of the '658 patent.
- Procedure in '658 patent only yielded form II; Novopharm sought to market form II, claiming '658 patent only covered form II and had expired.
- GSK sues Novopharm for patent infringement, showed that they could obtain form I in a small fraction of experiments. Courts ruled in favor of GSK.
- Novopharm then sought to market form I, produced their own samples.
- GSK tried to claim that Novopharm actually produced mixtures of forms I and II. Novopharm showed their samples have negligible amounts of form II. Courts permit them to market their product.
- GSK lost monopoly on "billion \$ polymorph."



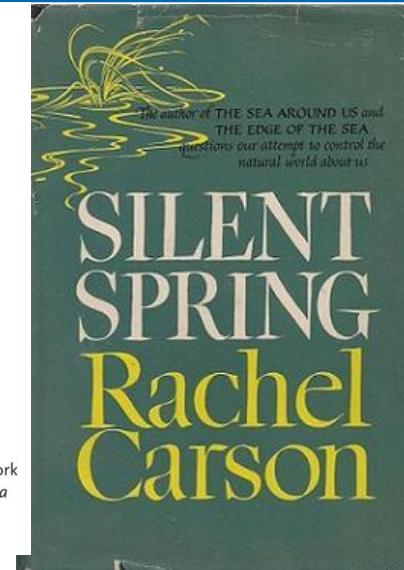
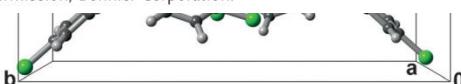
# Crystal Walkers and DDT polymorphs



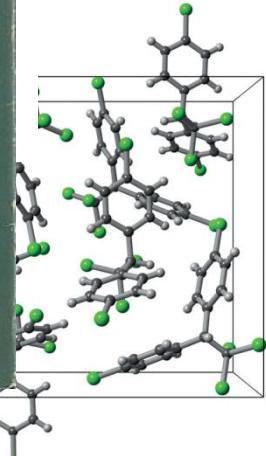
ESSAYS  
RACHEL



Figure 2. Illustration by Alexander Leydenfrost of a warmer New York City which accompanied excerpts of Rachel Carson's article *The Sea Around Us*<sup>[4]</sup> in *Popular Science*, November 1951.<sup>[5]</sup> Reprinted with permission, Bonnier Corporation.



phila faster  
]



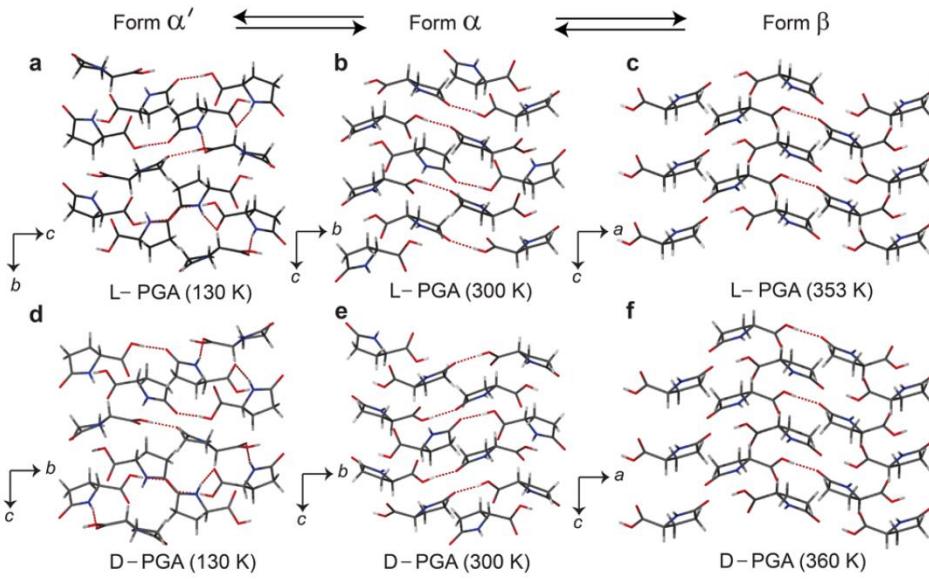
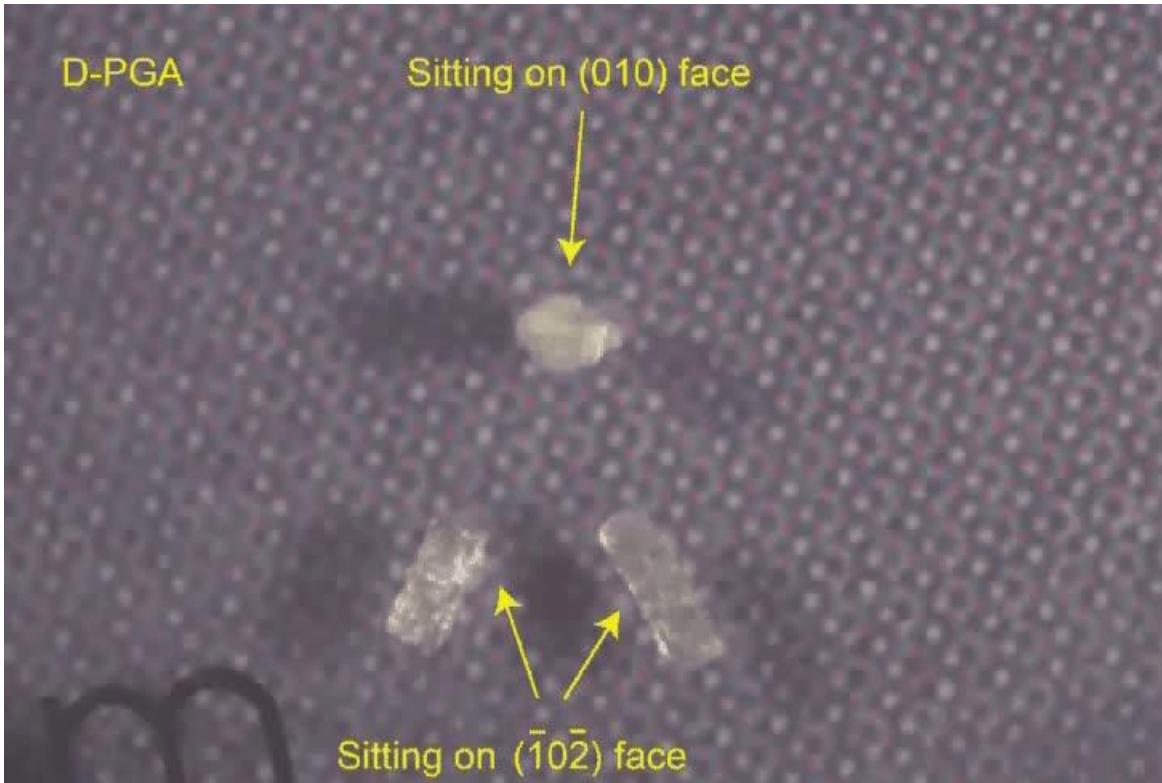
Ostwald's Rule of Stages: Least thermodynamically stable phase forms first,  
Our corollary on lethality: Least thermodynamically stable phase kills fastest.

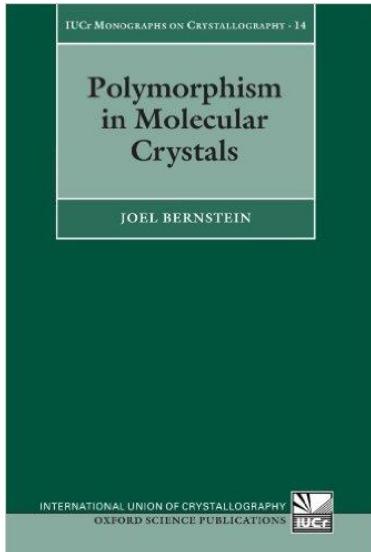


Pance Naumov

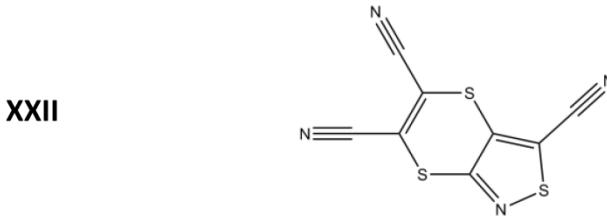
## L- and D- PGA crystals

M. K. Panda *et al.*  
*JACS* **137**, 1895 (2015)

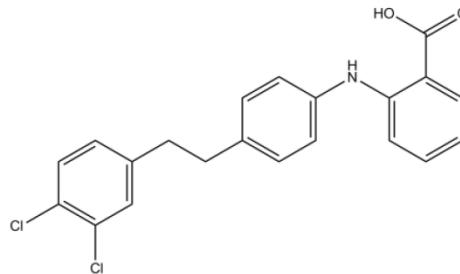




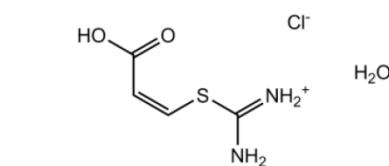
## Targets of the sixth blind structure prediction test



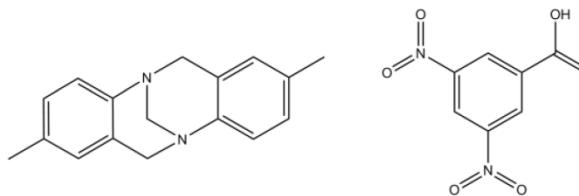
XXII



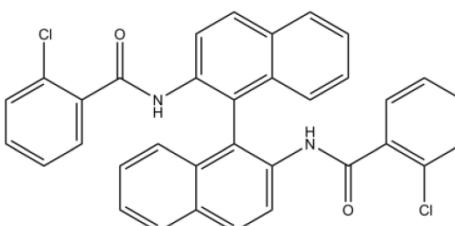
XXIII



XXIV



xxv

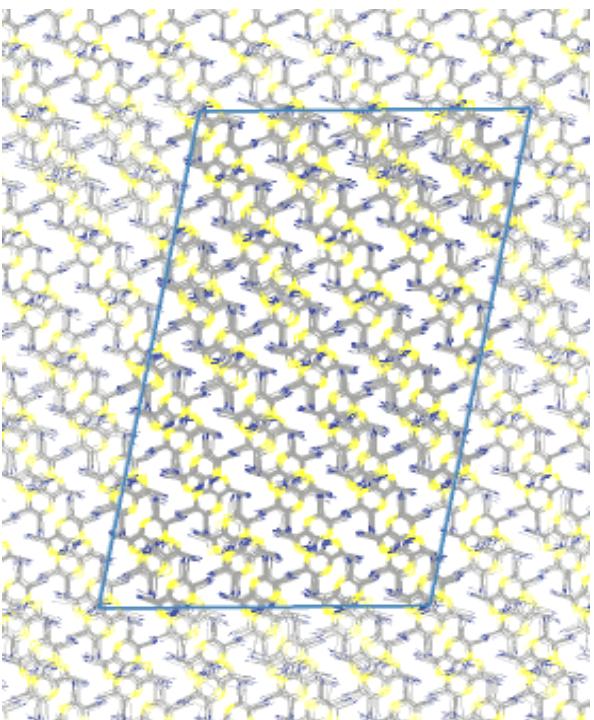
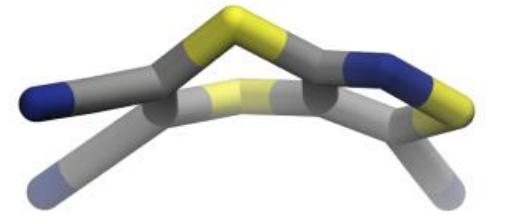


xxvi

Joel Bernstein (1941-2019)  
NYU Abu Dhabi  
NYU Shanghai



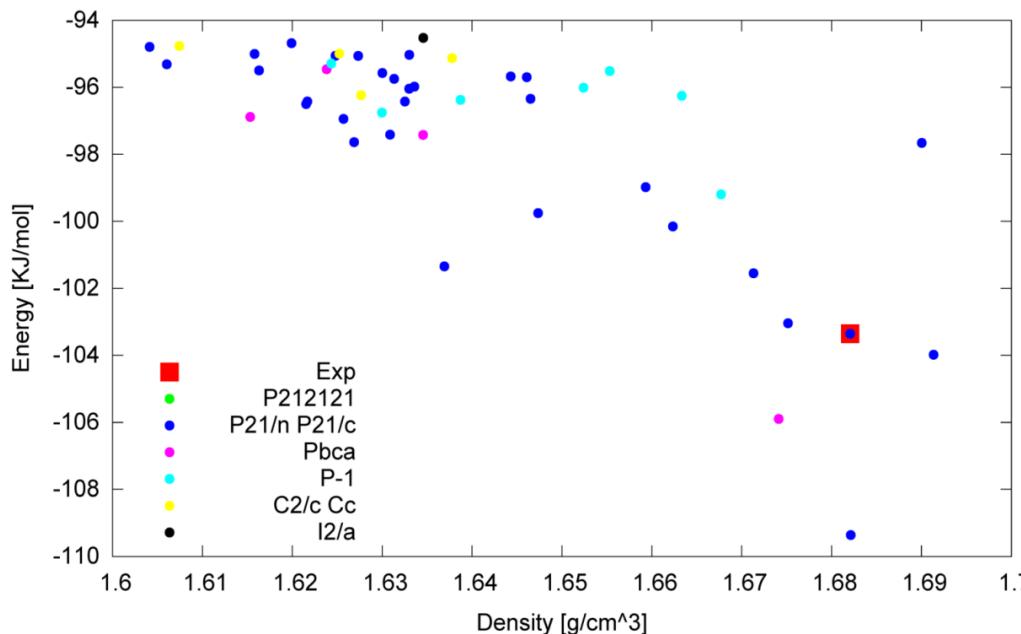
# Summary of ~ 50 predicted structures for the blind test



## Protocol:

1. Optimize the monomer using quantum chemistry.
2. Fit pairwise forcefield from quantum chemistry calculations on database of dimer configurations.
3. Pack molecules into random structures according to symmetry rules of top 20 space groups  
=> **48,000 structures** 😱
4. Cluster to find unique structures (**11,000** 😰) and save all structures below 10 kJ/mol (**131 structures** 🤔)
5. Optimize the structures using NPT MD.
6. Average over trajectories to predict final structures.

$$\hat{\mathcal{H}}\Psi = \mathcal{E}\Psi$$

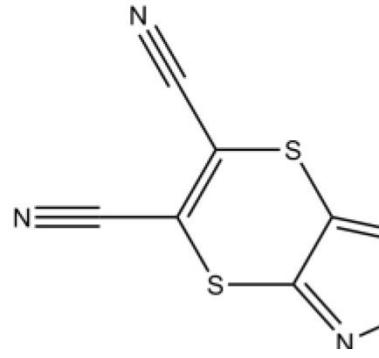


## Prediction of target XXII molecule from 6<sup>th</sup> CCDC blind structure prediction test

Reilly *et al.* *Acta Cryst. B* (2016);

E. Schneider, L. Vogt, MET *Acta Cryst. B* (2016)

Given information:



Final structure ge  
experimental *P* at

	<i>Spc g</i>				
<b>Expt.</b>	$P2_1/\bar{c}$				
<b>Calc.</b>	$P2_1/n$	12.02	6.72	12.48	108.7

At 300 K:  $\text{RMSD}_{20}$  from experimental structure = 0.187 Å

At 150 K:  $\text{RMSD}_{20}$  from experimental structure = 0.140 Å.

Attempted predictions = 21, Times predicted = 12, Times ranked 5 or lower = 7

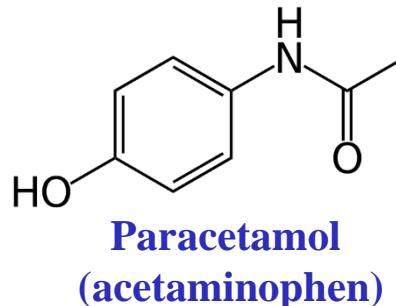


ranking step. However, it is worth remembering that experimental structures feature thermal-expansion effects, whereas the majority of the CSP methods are predicting 0 K ‘equilibrium’ geometries. MD simulations, which have been used by two submissions (Podeszwa *et al.* and Tuckerman, Szalewicz *et al.*), should capture these effects and provide better comparison with experiment. Such simulations require the temperature of the diffraction experiment as input though, which was not disclosed to participants. For (XXII), MD simulations at 300 K gave an  $\text{RMSD}_{20}$  of 0.187 Å (Tuckerman, Szalewicz *et al.*), but a post-test MD simulation at the experimental temperature of 150 K, gives a value of 0.140 Å, which is smaller than any of the RMSD values for the submitted structures. This demonstrates the significant contribution of

$$+\frac{1}{2}, \bar{z} + \frac{1}{2}$$

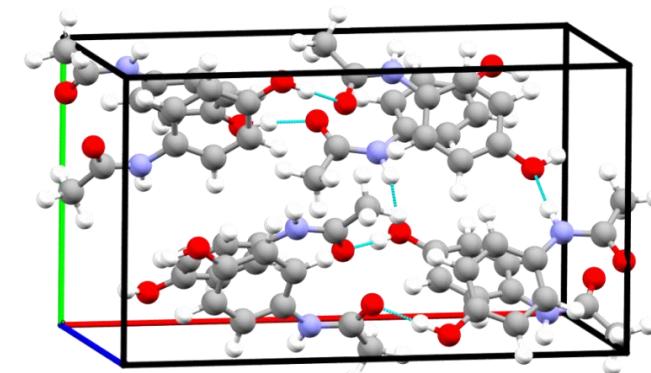
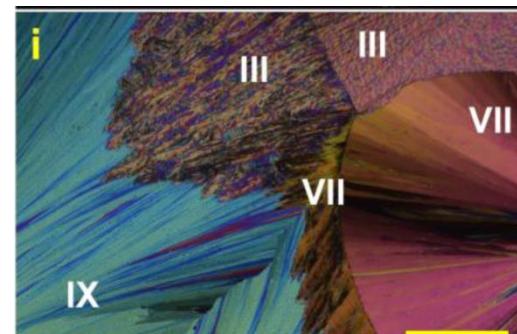
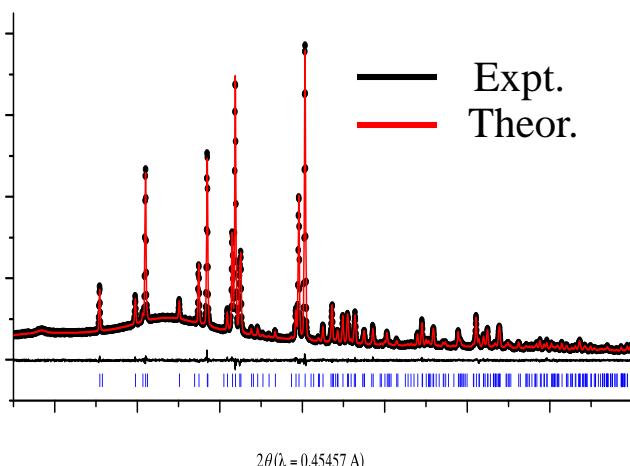
$$x + \frac{1}{2}, \bar{y} + \frac{1}{2}, z + \frac{1}{2}$$

# Crystal structure prediction of a new polymorph of acetaminophen

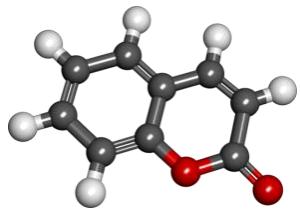


- New form called Form VII (metastable) crystallized in lab of B. Kahr long with new minor phases VIII and IX.
- Attempts to index the PXRD resulted in space groups  $P2_12_12_1$ ,  $P2_12_12$ ,  $P22_12_1$ ,  $P2_122_1$  and  $P222$  with  $Z' = 2$ .
- Energy evaluations of these all resulted in high energy values  $> 10 \text{ kJ/mol}$

- Our CSP protocol with  $Z' = 2$ , including  $P1$  in the search, yielded  $Pna2_1$  structure.
- Optimizing with DFT (PBE/TZVP) preserved the space group and brought the energy to  $< 10 \text{ kJ/mol}$ .
- Lattice parameters:  $a = 16.85 \text{ \AA}$ ,  $b = 9.49 \text{ \AA}$ ,  $c = 9.14 \text{ \AA}$ , good match to PXRD

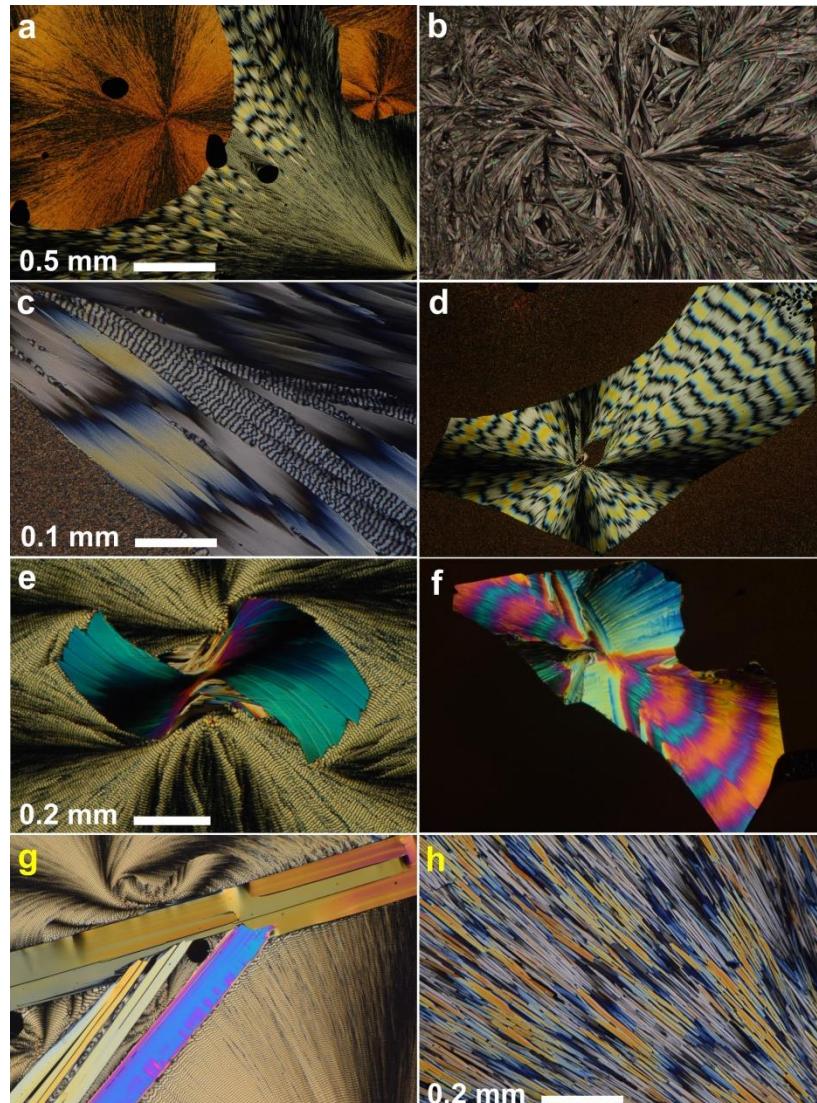


Shtukenberg *et al.* *Cryst. Growth Des.*  
(in press)



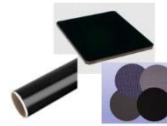
# A combined theoretical/experimental study of polymorphs of coumarin

A. Shtukenberg *et al.* *Chem. Sci.* **8**, 4296 (2017).



Phase       $\Delta G$  (rel. to I, kJ/mol),  $T = 54^\circ\text{C}$

Coumarin II  
(Opaque)



0.19

Coumarin IV  
(Blades)



0.26

Coumarin V

~0.84

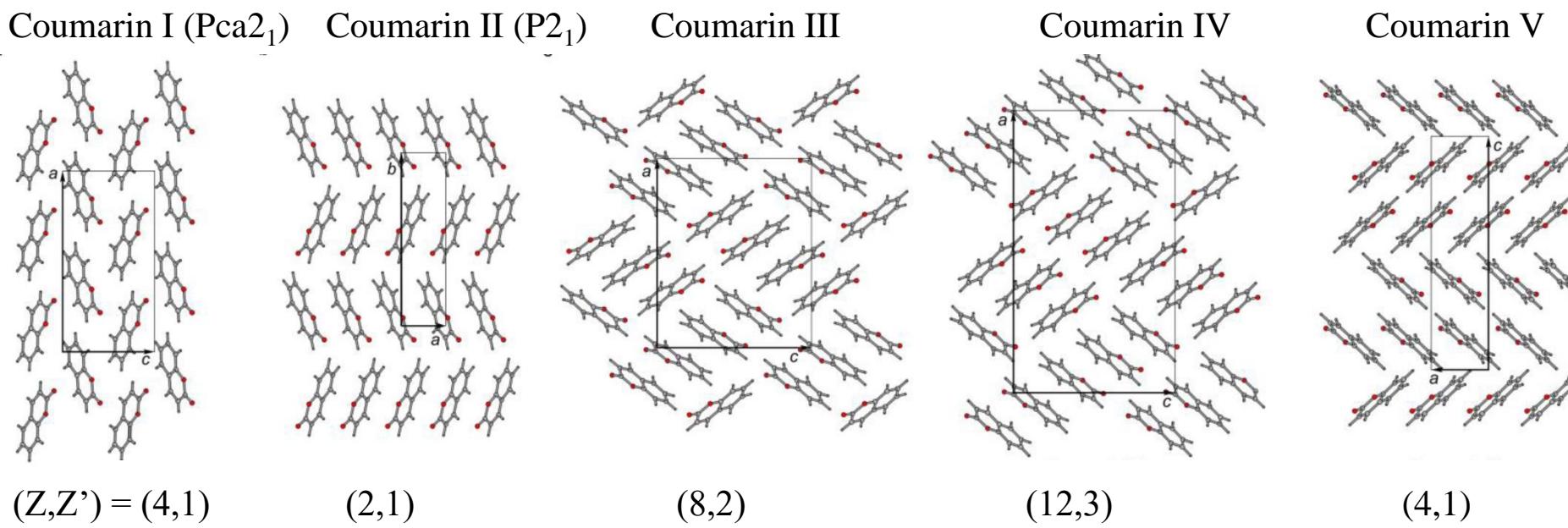


Coumarin III  
(Needles)

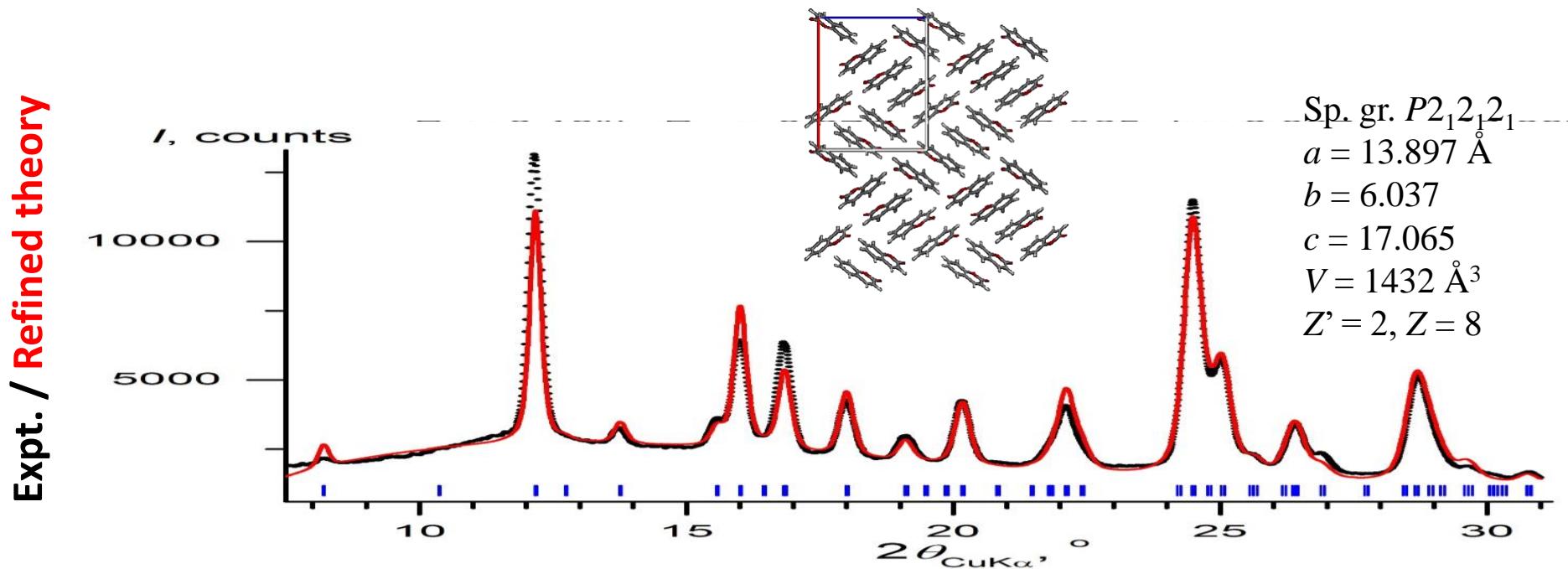


0.20

All polymorphs are metastable and convert to I within minutes. Can be stabilized by 10-30 % Canada balsam.



**Powder XRD pattern for Coumarin III phase ( $P2_12_12_1$ ,  $Z'=2$ )**





# Harmonic and quasi-harmonic approximations



Approximate free energy in the harmonic approximation:

$$G(P, T) = U_{\text{el}} + PV + F_{\text{harm-vib}}(T)$$

The harmonic vibrational frequencies are obtained by computing the lattice phonon frequencies at fully relaxed structures.

$$F_{\text{harm-vib}}(T) = \beta^{-1} \sum_{\alpha} \ln(\beta \hbar \omega_{\alpha})$$

In the quasi-harmonic approximation, we take into account the volume-dependence of the phonon frequencies and minimize the free energy over the volume:

$$G(P, T) = \min_V \left[ U_{\text{el}} + PV + \beta^{-1} \sum_{\alpha} \ln(\beta \hbar \omega_{\alpha}(V)) \right]$$

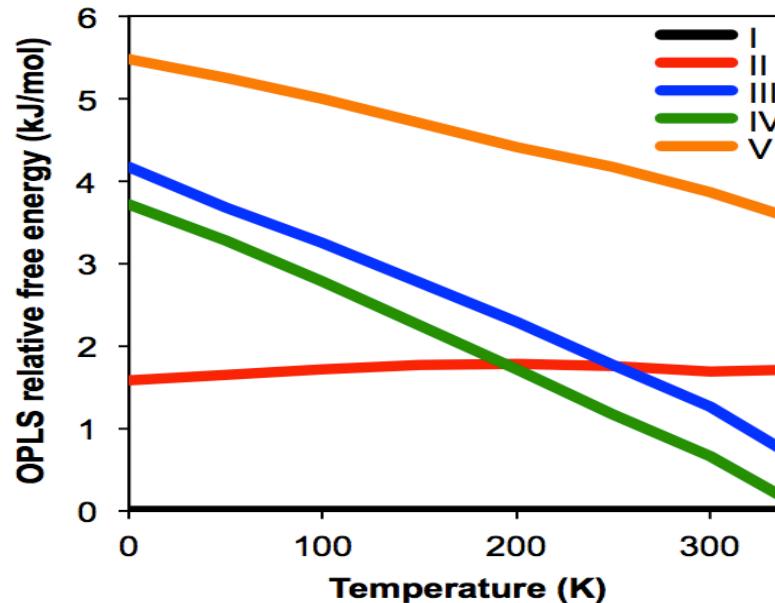
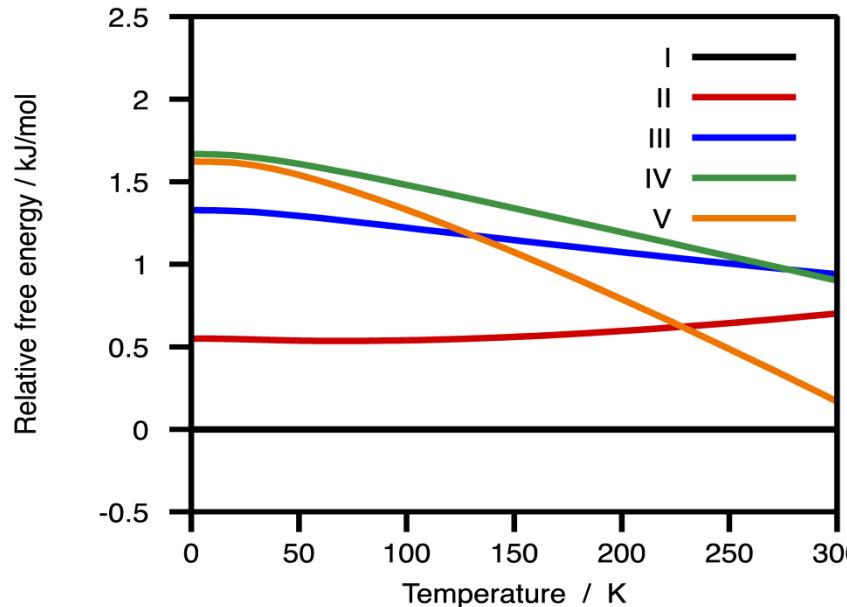


## Free energies from the harmonic approximation



**Harmonic  
free energies  
from PBE(0)+MBD**

**Harmonic  
free energies  
from force field**



Quasi-harmonic approximation does not improve on this problem.

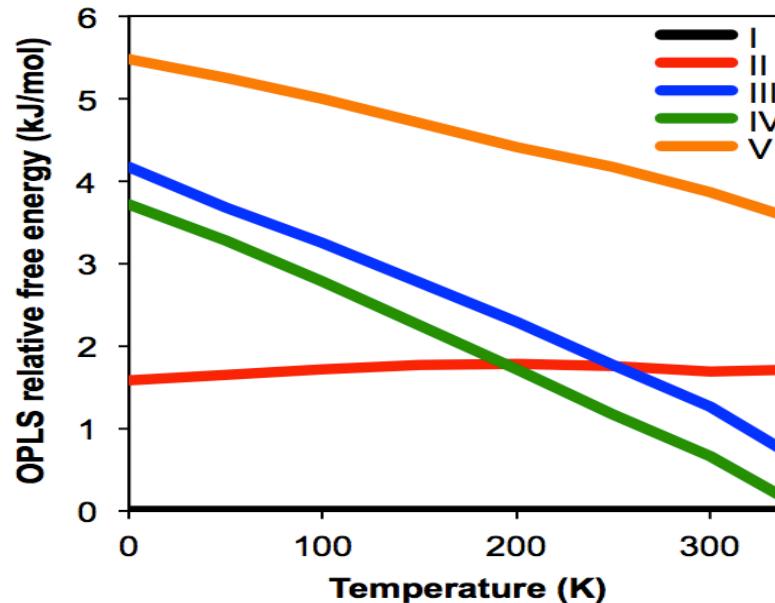
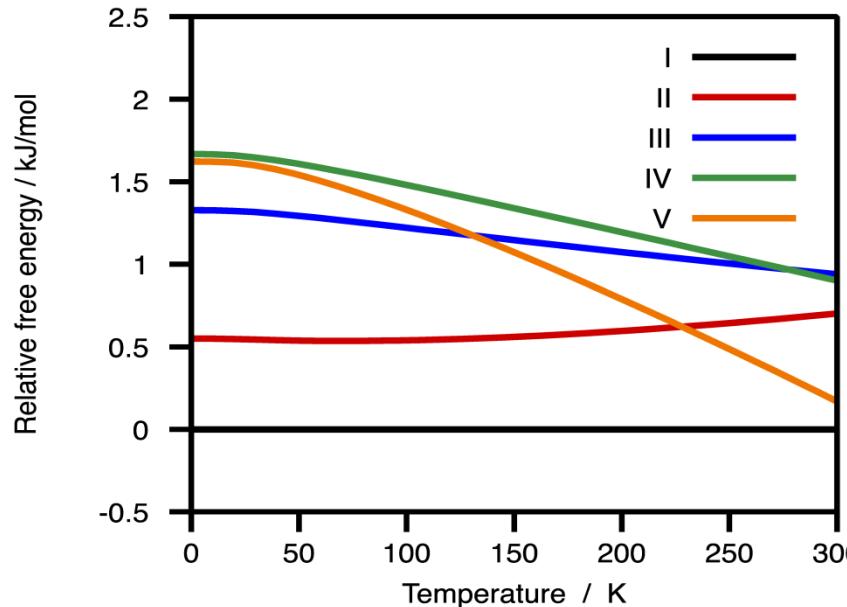


## Free energies from the harmonic approximation



**Harmonic  
free energies  
from PBE(0)+MBD**

**Harmonic  
free energies  
from force field**



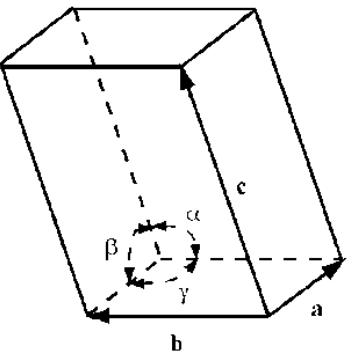
Quasi-harmonic approximation does not improve on this problem.



# Simple collective variables for exploration of crystal structures



**Cell shape:**



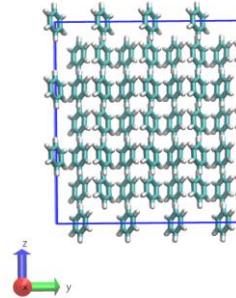
$$\mathbf{h} = \begin{pmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{pmatrix}$$

**Translational and rotational entropy:**

$$\frac{S}{Nk_B} = -\rho \int d\mathbf{r} \left[ g(\mathbf{r}) \ln g(\mathbf{r}) - (g(\mathbf{r}) - 1) \right]$$

**Spatial distribution:**

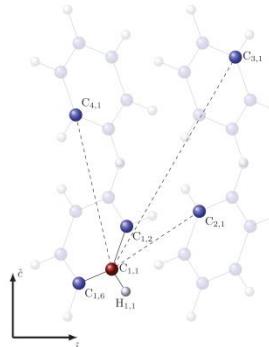
$$g(\mathbf{r}) = \frac{1}{4\pi^2 r^2 N \rho \sigma_r^2 \sigma_\theta^2} \sum_{i,j} e^{-(r-r_{ij})^2/2\sigma_r^2} e^{-(\cos\theta-\cos\theta_{ij})^2/2\sigma_\theta^2}$$



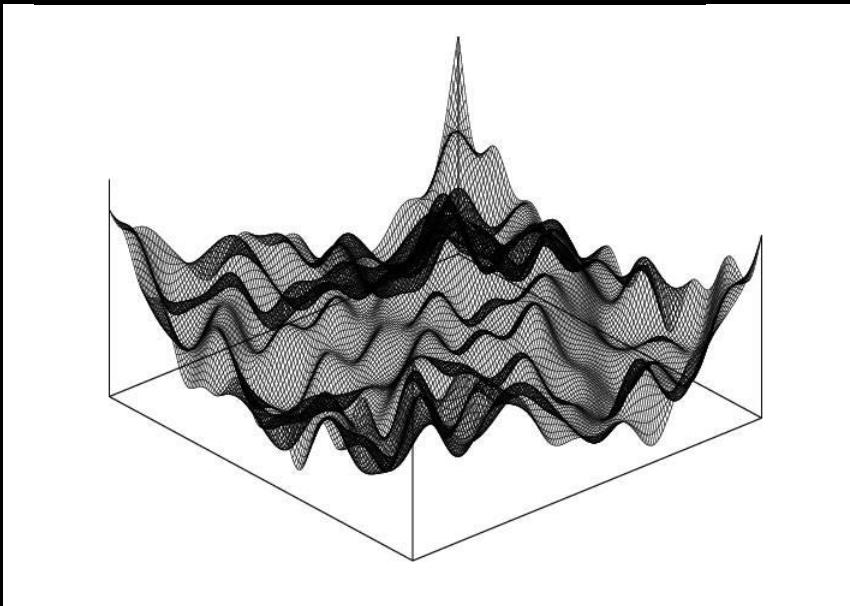
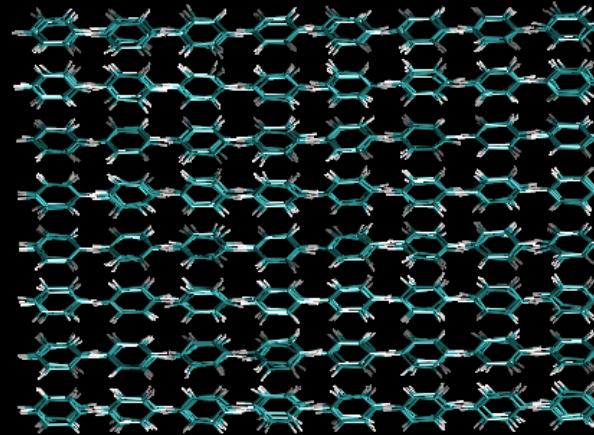
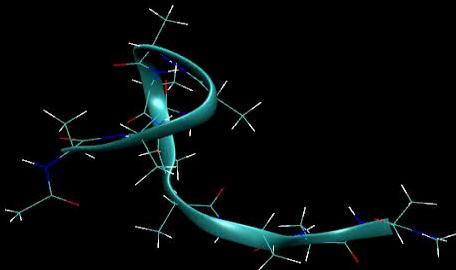
Piaggi and Parrinello

**Packing:**

$$Q_l^2 = \frac{4\pi}{(2l+1)N} \sum_{m=-l}^l \sum_{b=1}^{N_b} f(r_b) Y_{lm}(\theta_b, \varphi_b)$$



## Rough energy landscapes in proteins and crystals



Dunitz and Scheraga  
*PNAS* **101**, 14309 (2004)

## Adiabatic free energy (temperature-accelerated molecular) dynamics (AFED/TAMD)

L. Rosso, P. Minary, Z. Zhu, MET J. Chem. Phys. **116**, 4389 (2002)

Marglano and Vanden-Eijnden, Chem. Phys. Lett. **426**, 168 (2006); J. B. Abrams and MET, J. Phys. Chem. B **112**, 14752 (2008)

**For prediction of crystal structures:** Yu and MET Phys. Rev. Lett. **107**, 015701 (2011); Yu et al. J. Chem. Phys. **150**, 214109 (2014)

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

$$q_\alpha(\mathbf{r}), \quad \alpha = 1, \dots, n$$

Canonical probability distribution and free energy surface:

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

$$A(s_1, \dots, s_n, T) = -k_B T \ln P(s_1, \dots, s_n, T)$$

Write  $\delta$ -functions as product of Gaussians:

$$\delta(q_\alpha(\mathbf{r}) - s_\alpha) = \lim_{\{\kappa_\alpha \rightarrow \infty\}} \left( \frac{\beta \kappa_\alpha}{2\pi} \right)^{1/2} \exp \left[ -\frac{\beta \kappa_\alpha}{2} (q_\alpha(\mathbf{r}) - s_\alpha)^2 \right]$$

Dynamically sample Gaussian centers  $s_1, \dots, s_n$  by introducing a kinetic energy for them  
And extending the phase space, which gives the following Hamiltonian:

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

## Adiabatic free energy (temperature-accelerated molecular) dynamics (AFED/TAMD)

Ensure "on the fly" sampling of the marginal:  $m_\alpha \gg m_i$

Accelerate sampling via temperature:  $T_s \gg T$  for extended variables

**Adiabatically decoupled equations of motion:**

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

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

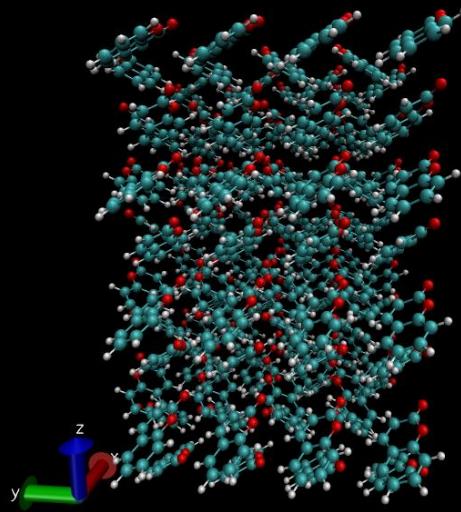
Under adiabatic conditions, we generate a distribution  $P_{\text{adb}}^{(\{\kappa\})}(s_1, \dots, s_n, T_s, T)$

$$\lim_{\{\kappa \rightarrow \infty\}} P_{\text{adb}}^{(\{\kappa\})}(s_1, \dots, s_n, T_s, T) = [P(s_1, \dots, s_n, T)]^{T/T_s}$$

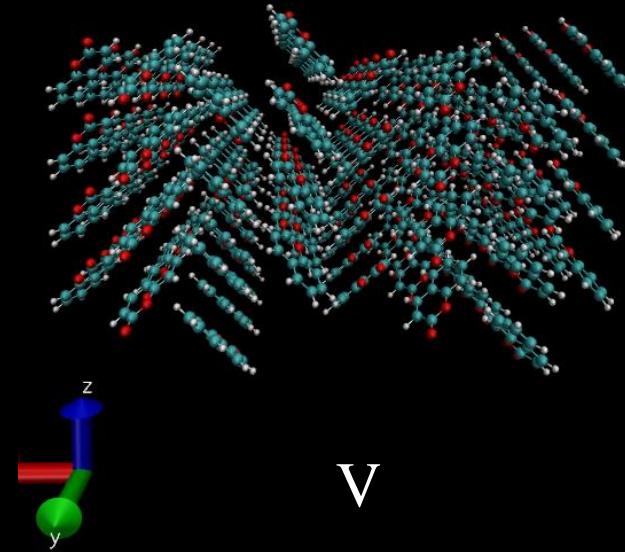
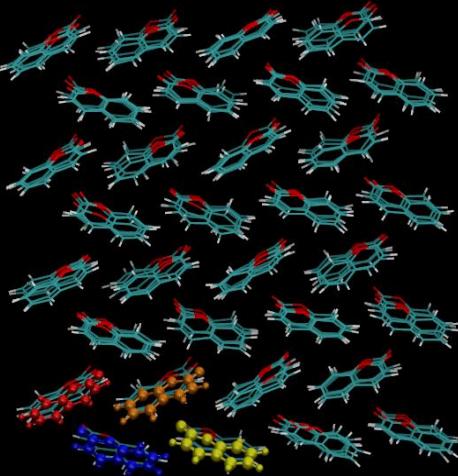
$$\lim_{\{\kappa \rightarrow \infty\}} \left[ \ln P_{\text{adb}}^{(\{\kappa\})}(s_1, \dots, s_n, T_s, T) \right] = \frac{T}{T_s} \ln P(s_1, \dots, s_n)$$

$$\lim_{\{\kappa \rightarrow \infty\}} \left[ -k_B T_s \ln P_{\text{adb}}^{(\{\kappa\})}(s_1, \dots, s_n, T_s, T) \right] = -k_B \cancel{\sum_s} \frac{T}{\cancel{T_s}} \ln P(s_1, \dots, s_n, T) = A(s_1, \dots, s_n, T)$$

# Probing structural transitions enhancing internal CVs

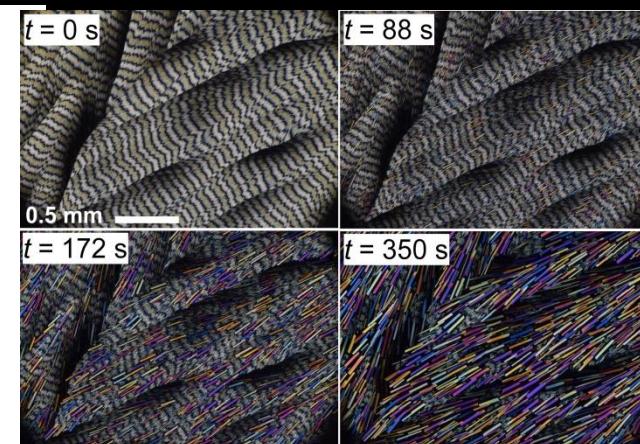


I

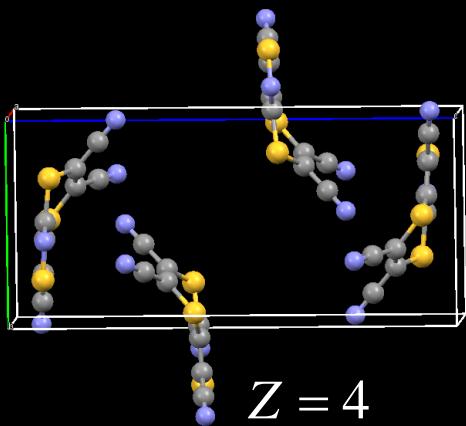


V

Temperature (K)	<b>I → II</b>	<b>I → V</b>	<b>II → V</b>
100	$3.2 \pm 0.7$	$15.0 \pm 1.6$	n/d
200	$3.9 \pm 0.7$	$15.5 \pm 1.6$	n/d
300	$4.5 \pm 0.7$	$16.0 \pm 1.6$	$11.6 \pm 0.7$

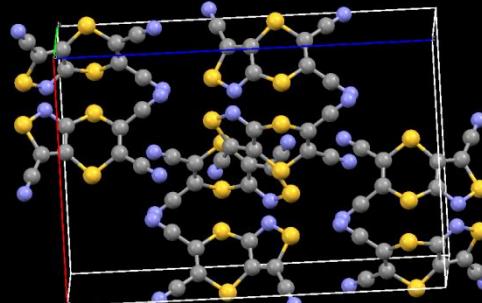


I:  $P2_1/n$



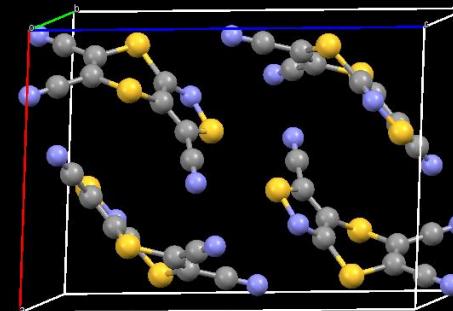
$Z = 4$

II:  $Pbca$ ,  $E = 3.46 \text{ kJ/mol}$

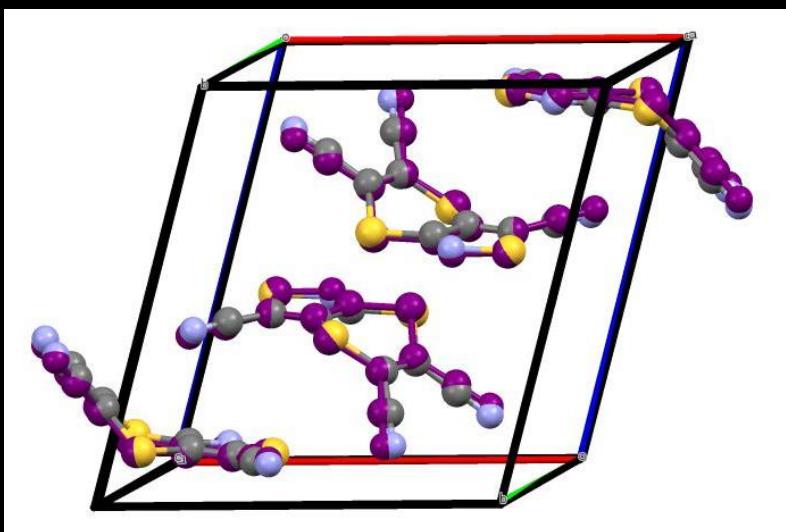


$Z = 8$

III:  $P2_1/c$ ,  $E = 5.37 \text{ kJ/mol}$

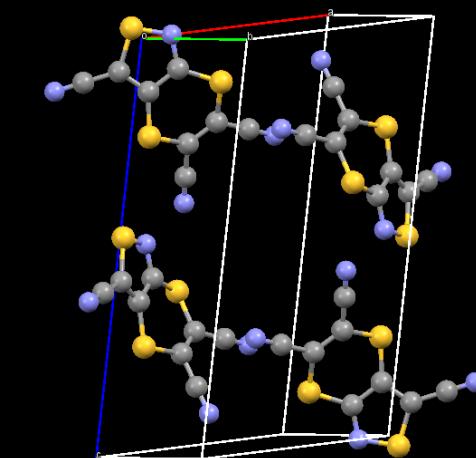


$Z = 4$



IV:  $P2_1/n$ ,  $E = 5.99 \text{ kJ/mol}$

$Z = 4$



V:  $P2_1/c$ ,  $E = 6.31 \text{ kJ/mol}$

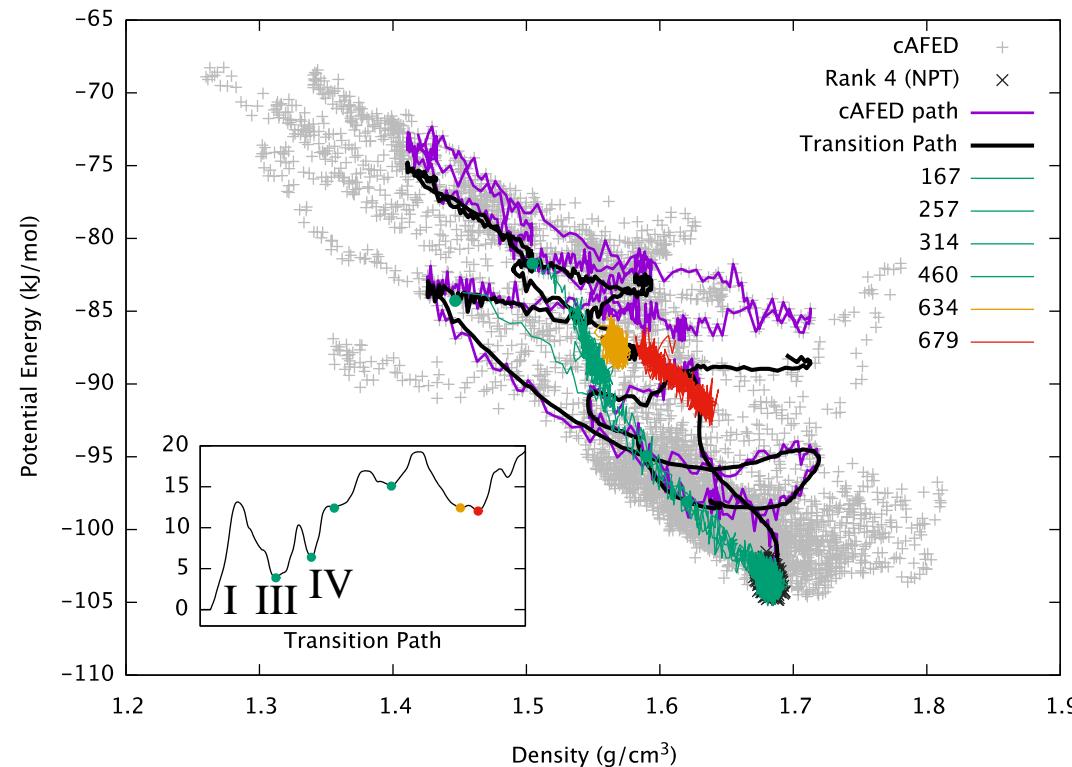
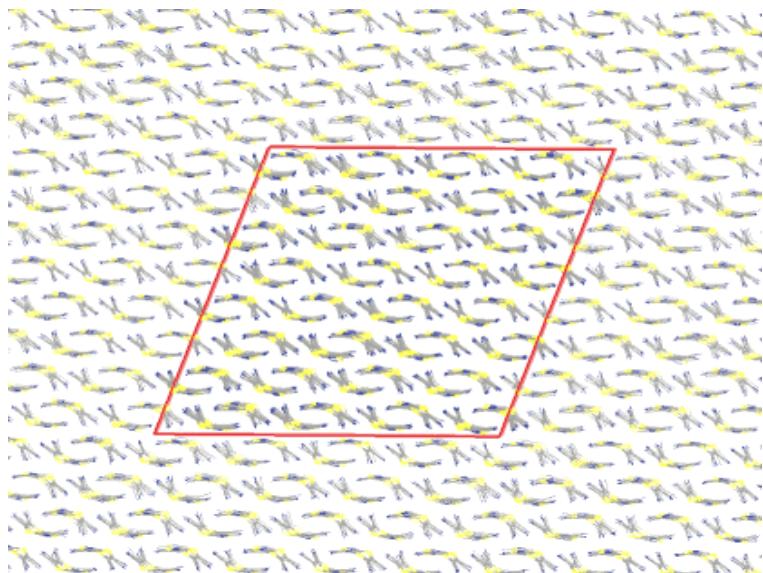
$Z = 4$



# Finding new structures using Crystal-AFED

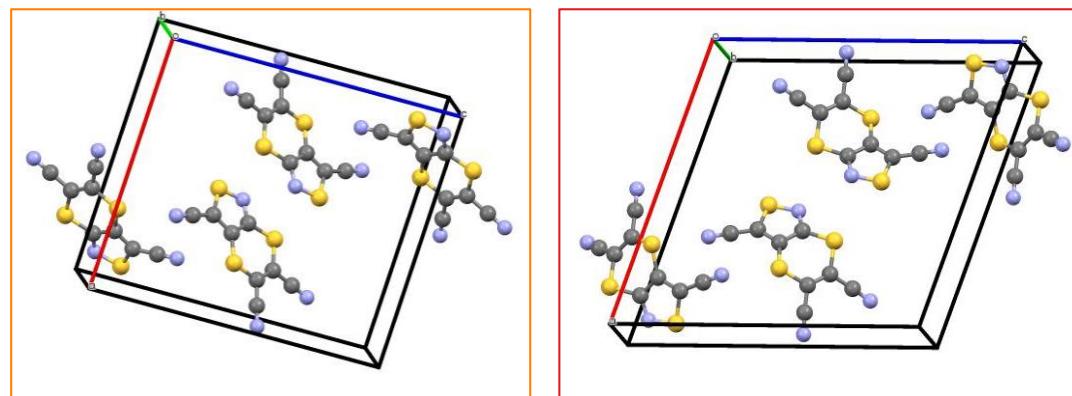


Enhanced Sampling using cell matrix as CV  
 $T = 150 \text{ K}$ ,  $T_s = 50,000 \text{ K}$ , 100 ps



Finds the monoclinic structures that are free-energy minima!

Space Group	$a$ [Å]	$b$ [Å]	$c$ [Å]	$\beta$
P21/n	13.610	4.951	15.591	91.5
P21/n	15.566	4.201	16.080	106.2

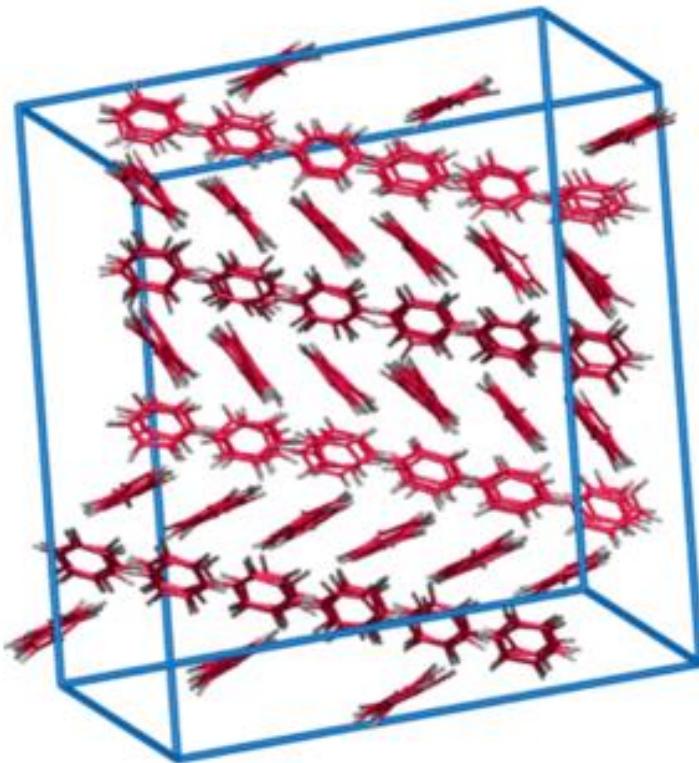




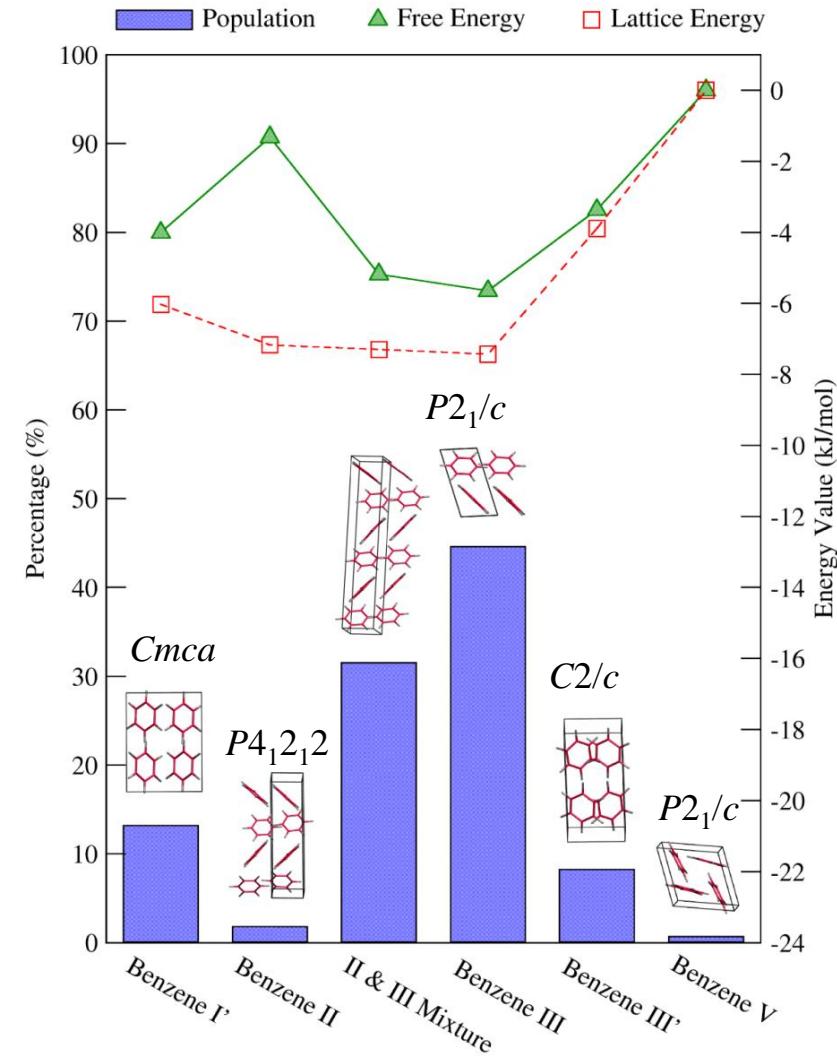
# Free energies of benzene polymorphs



Yu and Tuckerman *PRL* **107** 015701 (2011); Schneider\*, Vogt\*, Tuckerman *Acta Cryst. B* **2016**



GROMOS FF,  $T = 100$  K,  
 $T_s = 32,000 - 40,000$  K (CVs = cell matrix)  
 $N = 216$  ( $3 \times 3 \times 3$ ),  $P = 2$  GPa  
*Six-dimensional free energy surface!*

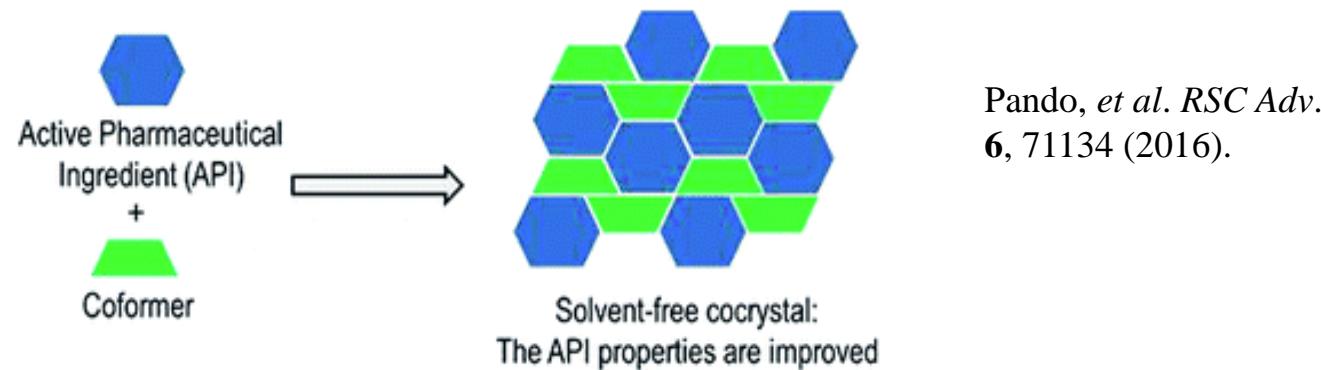




# Cocrystals and their role in pharmaceuticals



- Although the definition of a “cocrystal” depends on context, in molecular crystals, if two or more compounds crystallize together, they form a *cocrystal*.
- A 1:1 cocrystal, for example, can be thought of as a  $Z' = 2$  crystal in which the two monomers in the asymmetric unit are different. Some hydrates can also be thought of as cocrystals.
- In pharmaceutical preparations, the API may be crystallized with a coformer (forming a cocrystal) in order to improve the pharmacological properties of the API, such as solubility or chemical stability.

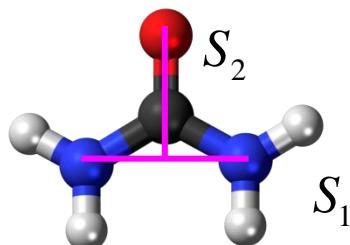
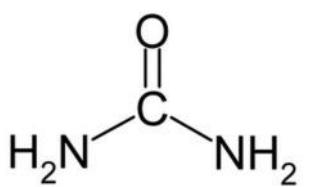


- In addition to the challenge of determining whether two monomers will cocrystallize, it is also important to predict the existence of polymorphism in these cocrystals.
- Heuristic prediction [Sandu, *et al. Cryst. Growth & Design* **18**, 466 (2018)]:

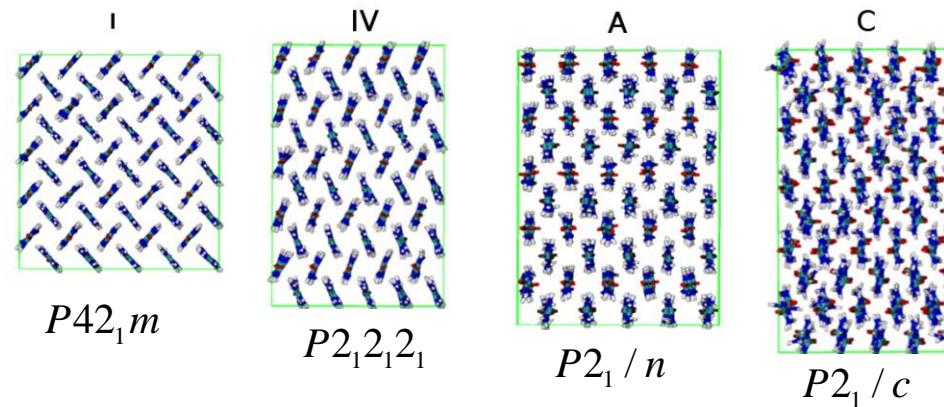
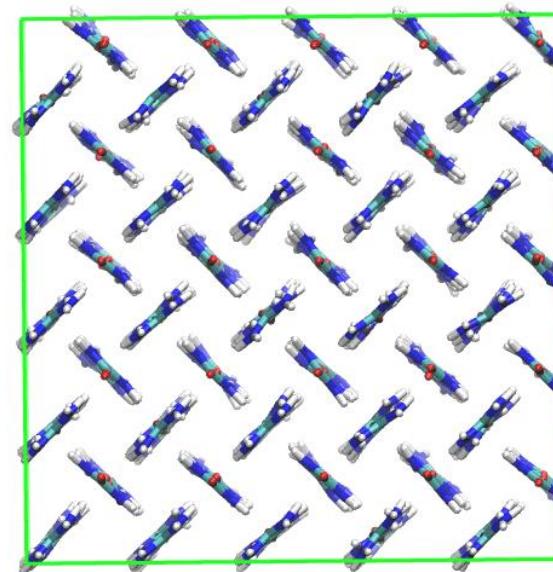
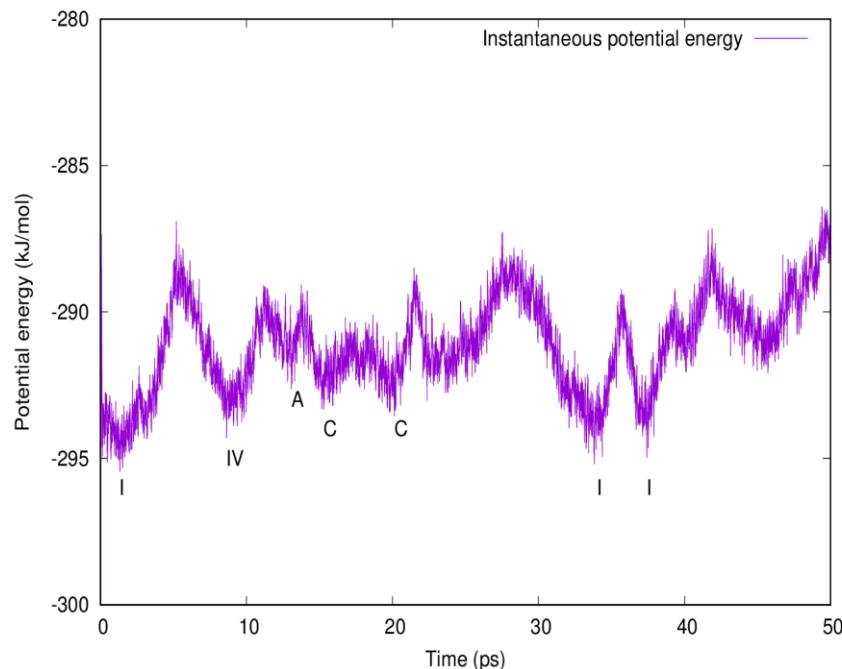
$$\text{MC} = (\text{heteromeric interaction propensity}) - (\text{homomeric interaction propensity})$$

Achieves roughly 50-80 % accuracy, depending on the compounds considered.

## CSP of urea using enhanced sampling



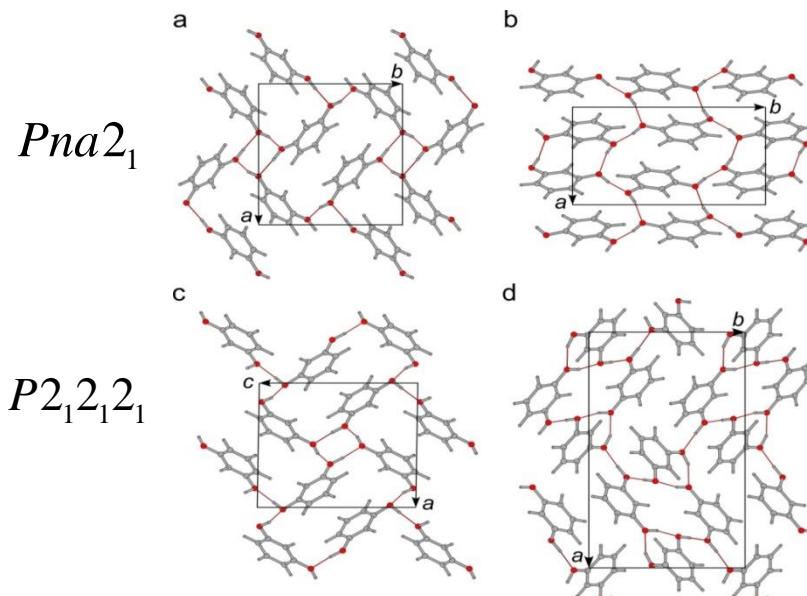
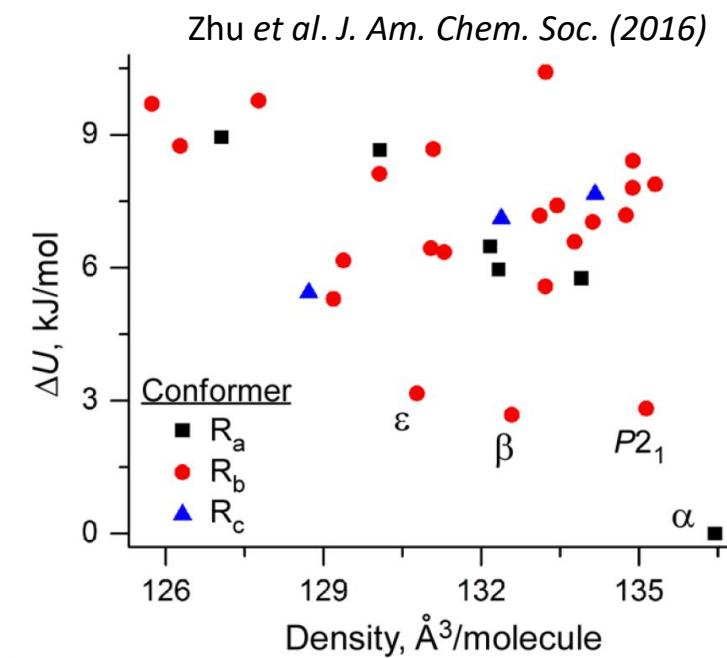
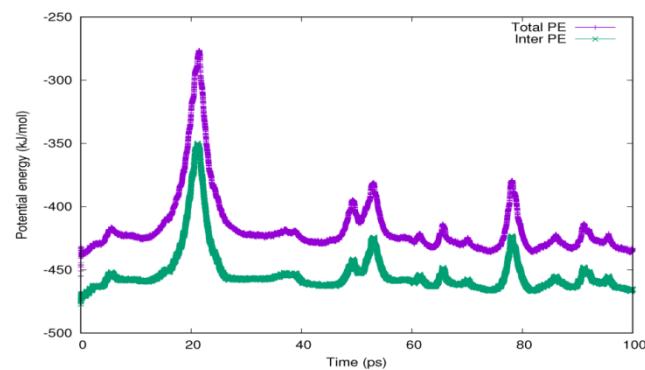
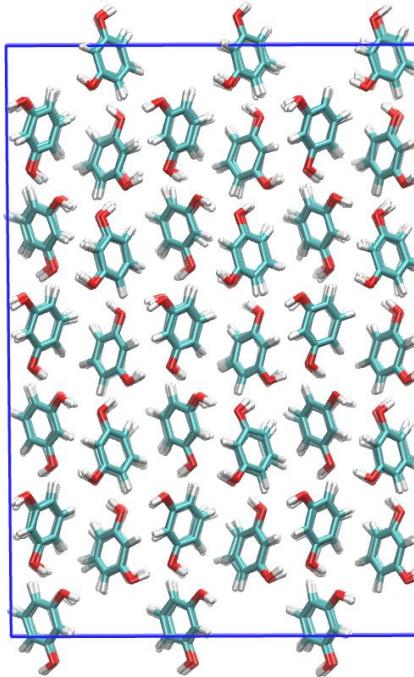
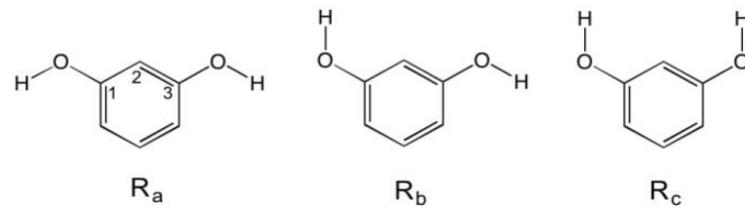
d-AFED simulations at 300 K, 1 bar,  $T_s = 10^4$  K  
OPLS FF with DFT charges. **Two entropy CVs.**



Study by Piaggi and Parrinello [arXiv:1806.06006]  
Used well-tempered metadynamics and required  
200 ns to find these urea polymorphs.  
[H. Song and MET (in preparation)]

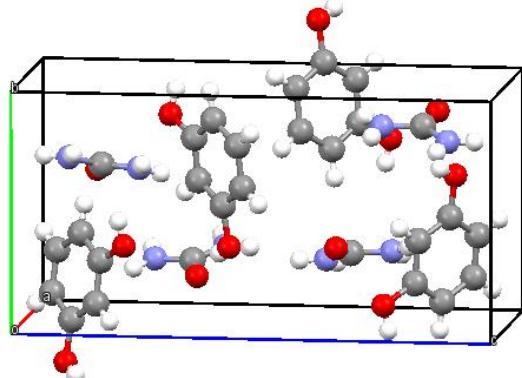
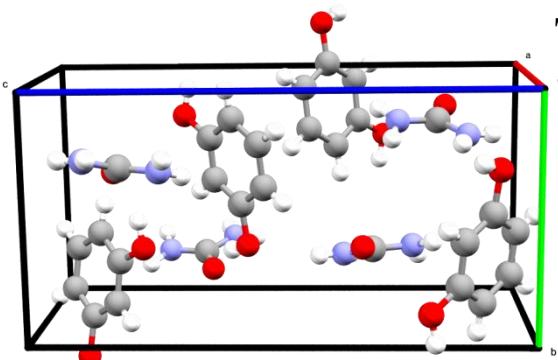
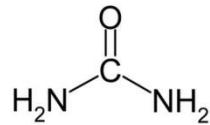
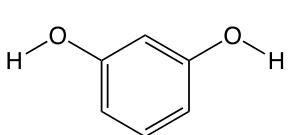


# CSP of resorcinol

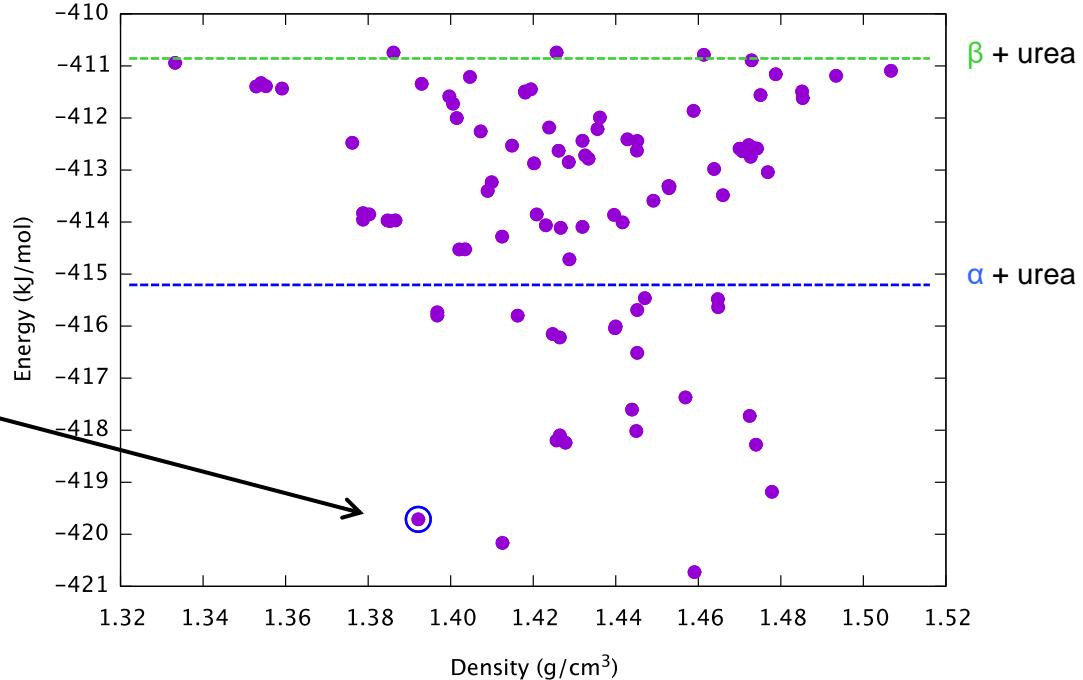




# Urea, resorcinol, and their 1:1 co-crystal



Perform CSP as a  $Z' = 2$  search



FF: OPLS + DFT charges

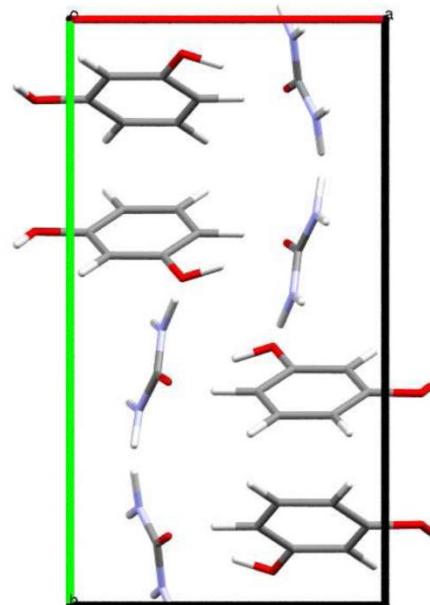
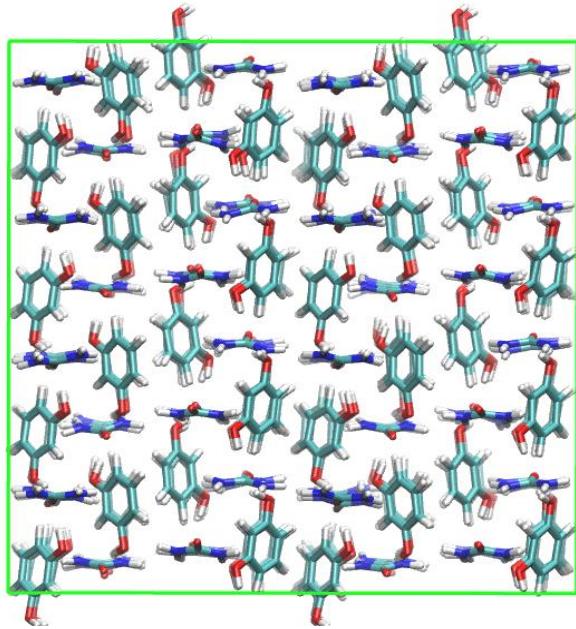
Molecular structure → Generate crystals → MD: Evaluate stability/energy ranking

Experimental structure: Pickering, Small Acta Cryst. B (1982); Matzger (2019)

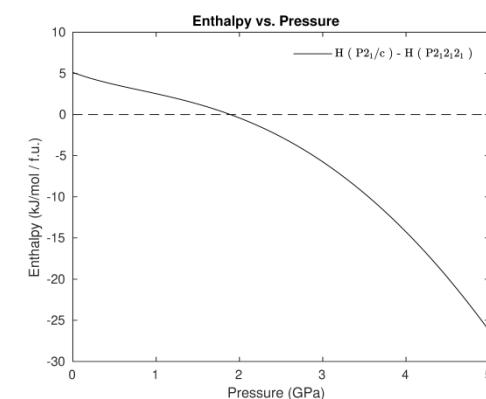
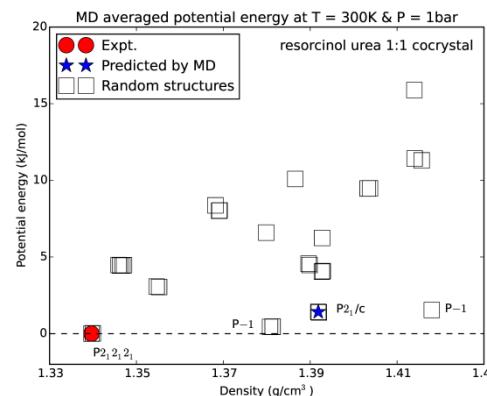
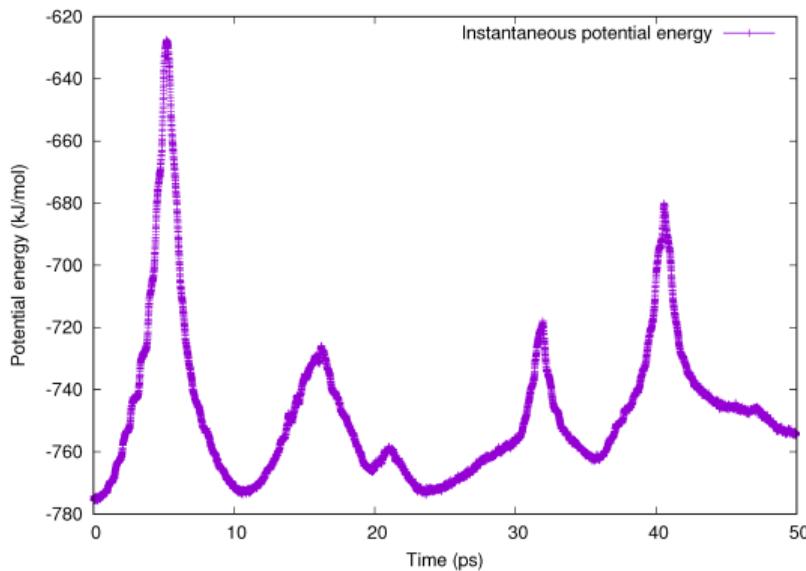
Space group:  $P2_12_12_1$     RMSD<sub>20</sub> = 0.47 Å



# Urea, resorcinol, and their 1:1 co-crystal



[001]



**P2<sub>1</sub>/c :**  
 $a = 7.9678 \text{ \AA}$   
 $b = 14.8209 \text{ \AA}$   
 $c = 6.9249 \text{ \AA}$   
 $\beta = 96.66^\circ$



## Why don't we see more predicted polymorphs?

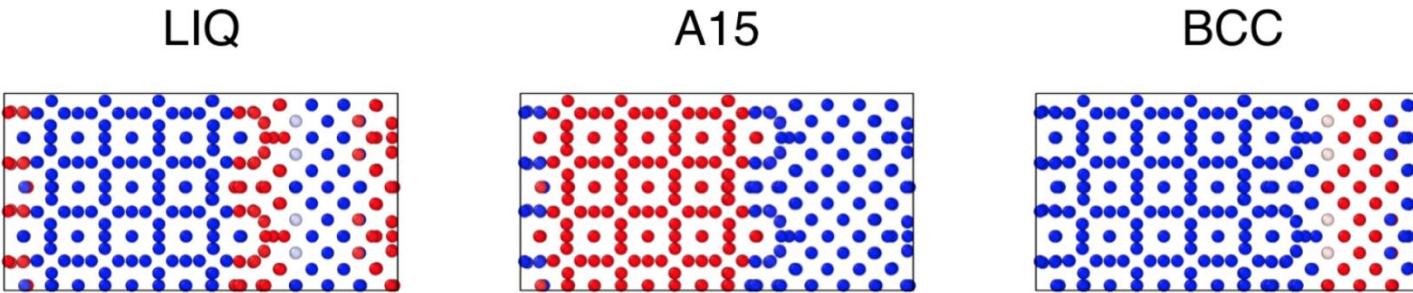


Some possible answers are [S. L. Price Acta Cryst. B (2013)]:

- Neglect of thermal effects in crystal packing calculations:  
*We saw that many false minima are eliminated when these are taken into account.*
- The approximations inherent in force field models are too inaccurate:  
*This needs further exploration, e.g., our FFs neglect manybody effects.*
- The method of crystallization restricts which arrangements are possible.
- Some structures (under certain conditions) cause dynamic or static disorder, leading to stacking faults, defects, differences in local domains:  
*High-pressure benzene example illustrates this.*
- Under the experimental conditions, they are only metastable. They might become stable under other conditions.
- The right experiment has not or cannot be performed.
- Some polymorphs simply escape detection or quickly disappear.
- They are not free-energy minima.



# Machine learning of CVs: Phase boundary migration in Molybdenum

ML-based design of a collective variable:

1. Start with a set of 14 primitive CVs:  $Q_6, Q_7, Q_8$ , and 11 symmetry functions of the form

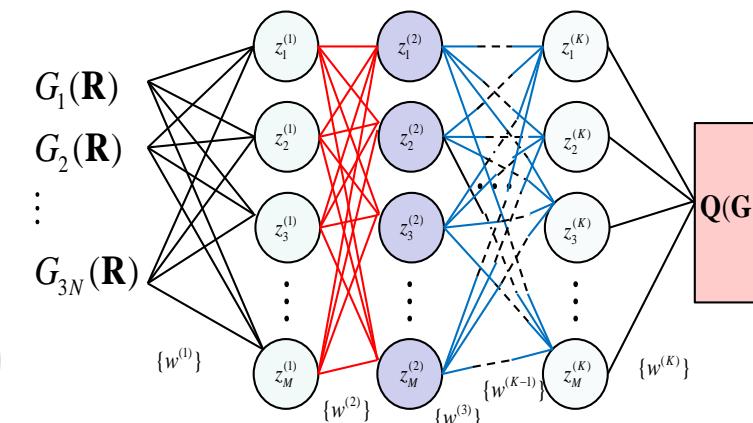
$$G_1(\mathbf{r}_i) = \sum_{j \neq i} e^{-\eta(|\mathbf{r}_{ij}| - R_s)^2} f_c(\mathbf{r}_{ij}), \quad G_2(\mathbf{r}_i) = \sum_{j \neq i} \cos(\kappa |\mathbf{r}_{ij}|) f_c(\mathbf{r}_{ij})$$

2. Sample the 14-dimensional space using TAMD/d-AFED
3. Use 3,000 samples generated to train a neural network to classify FCC, BCC, HCP, A15, and liquid (LIQ) phases. Outputs are structure classifiers  $Q^{\text{FCC}}, Q^{\text{BCC}}, Q^{\text{HCP}}, Q^{\text{A15}}, Q^{\text{LIQ}}$ .
4. Use classifiers  $\mathbf{Q} = (Q^{\text{A15}}, Q^{\text{BCC}})$  to construct a path variable in classifier space  
[Branduardi et al. JCP 126, 054103 (2007)]:

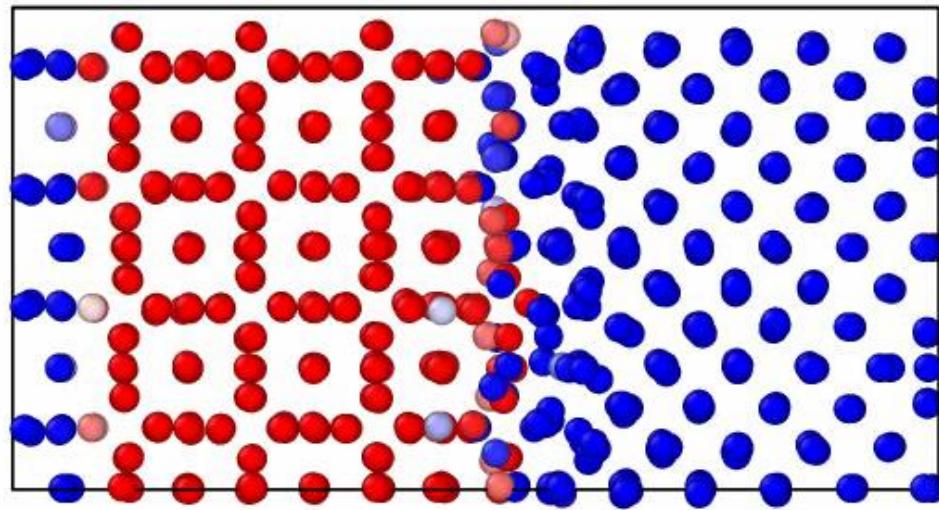
$$s(\mathbf{Q}) = \frac{1}{P-1} \frac{\sum_{i=1}^P (i-1) \exp(-\lambda(\mathbf{Q} - \mathbf{Q}(i))^2)}{\sum_{i=1}^P \exp(-\lambda(\mathbf{Q} - \mathbf{Q}(i))^2)}$$

5. Use orthogonal variable as an external bias

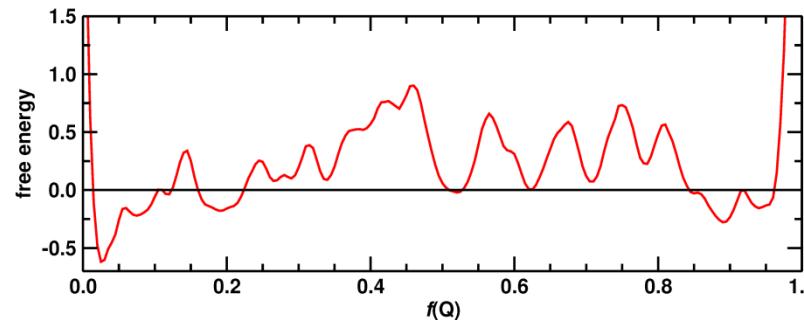
[Margul et al. JCP (2018)]:  $z(\mathbf{Q}) = -\frac{1}{\lambda} \ln \left( \sum_{i=1}^P \exp(-\lambda(\mathbf{Q} - \mathbf{Q}(i))^2) \right)$



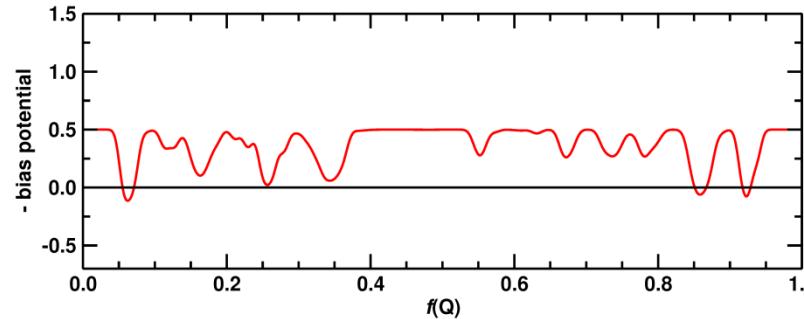
Time = 3480 ps



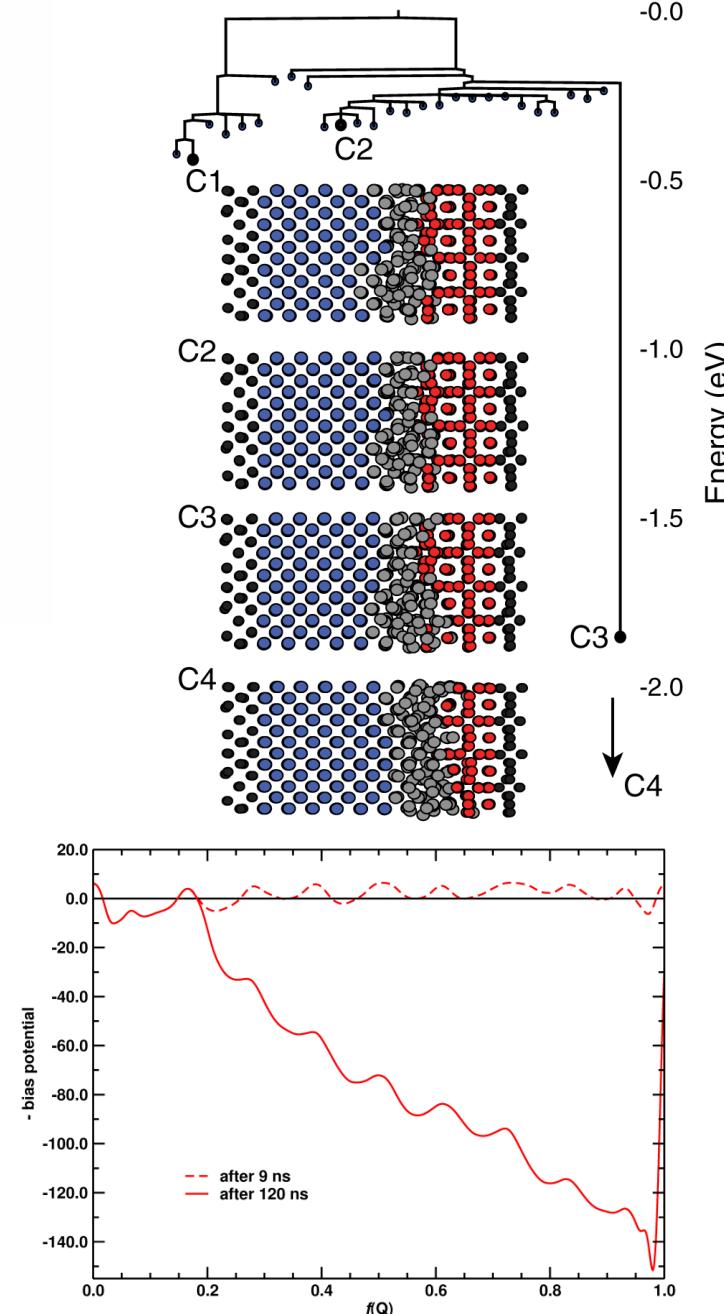
d-AFED



Meta-dynamics



Disconnectivity graphs from  
Kinetic MC: Duncan, *et al.* *Phys. Rev. Lett.* (2016)

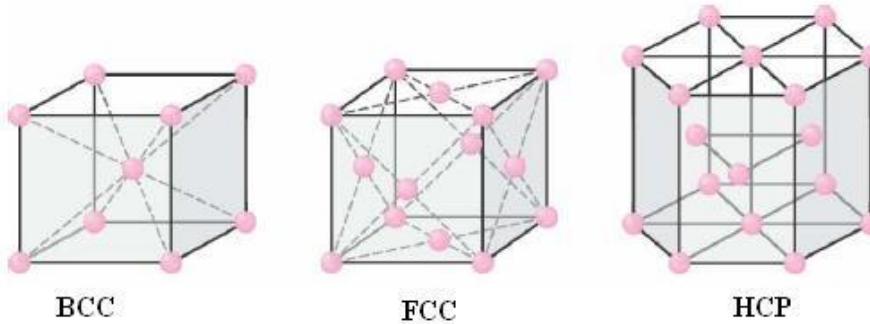




# Crystal structures of Xenon at 2700 K and 25 GPa



T. Q. Yu, P. –Y. Chen, A. Samanta, E. Vanden-Eijnden, MET J. Chem. Phys. (2014)



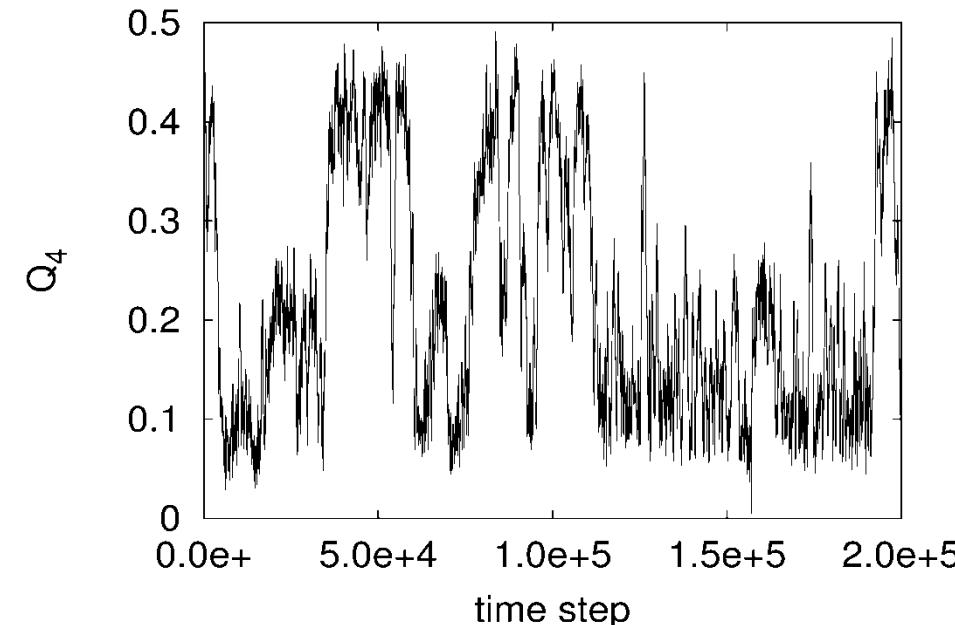
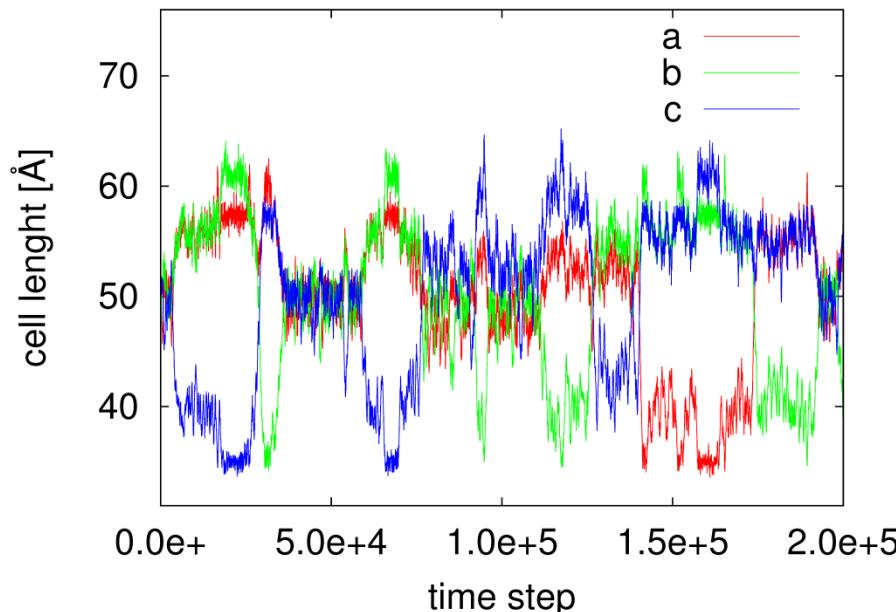
**Experimental evidence mixed fcc-hcp structures  
in pressure-induced Martensitic transitions  
from fcc to hcp as intermediates:**

Jephcoat *et al.* Phys. Rev. Lett. 59, 2670 (1987)

Cynn *et al.* Phys. Rev. Lett. 86, 4552 (2001)

Using  $Q_4$ ,  $Q_6$ , and the diagonal cell matrix  $\mathbf{h}_d$  as collective variables

$T_s = 10^5$  K,  $m_s = 10^8$  amu, 4,000 atoms

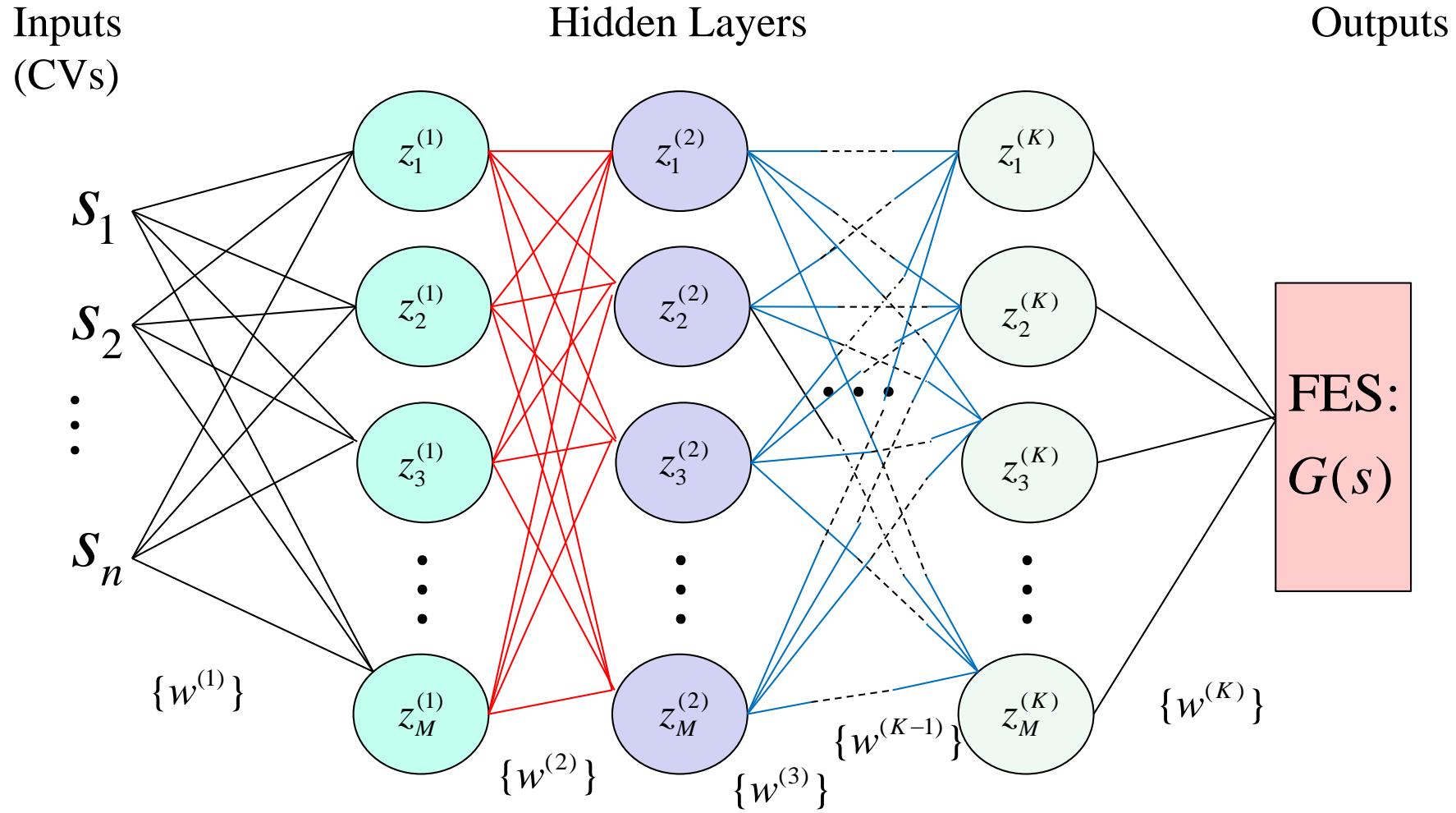




# Artificial neural networks (ANNs) and free energy landscapes



L. Dai, E. Schneider, C. Drechsel-Grau, MET Phys. Rev. Lett. (2017)



$$G_{\text{ANN}}(s_1, \dots, s_n, \mathbf{w}) \equiv G_{\text{ANN}}(\mathbf{s}, \mathbf{w})$$

$$= H \left( \sum_{j_K=1}^M h \left( \sum_{j_{K-1}=1}^M h \left( \dots \sum_{j_1=1}^M h \left( \sum_{\alpha=1}^n s_\alpha w_{\alpha j_1}^0 + w_{0 j_1}^0 \right) w_{j_1 j_2}^1 + w_{0 j_2}^1 \dots \right) w_{j_{K-2} j_{K-1}}^{K-1} + w_{0 j_{K-1}}^{K-1} \right) w_{j_{K-1} j_K}^K + w_{0 j_K}^K \right)$$



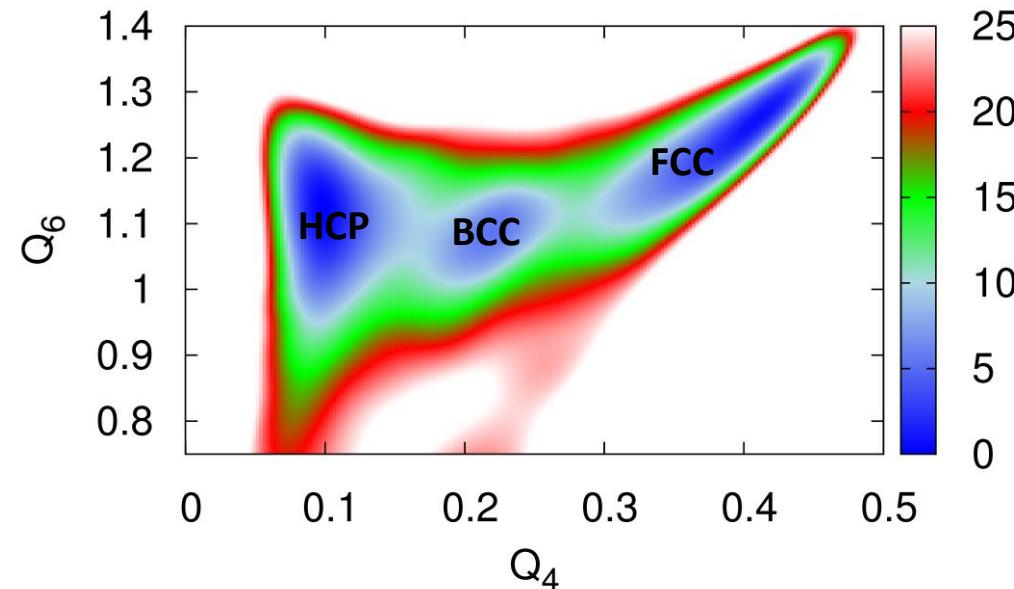
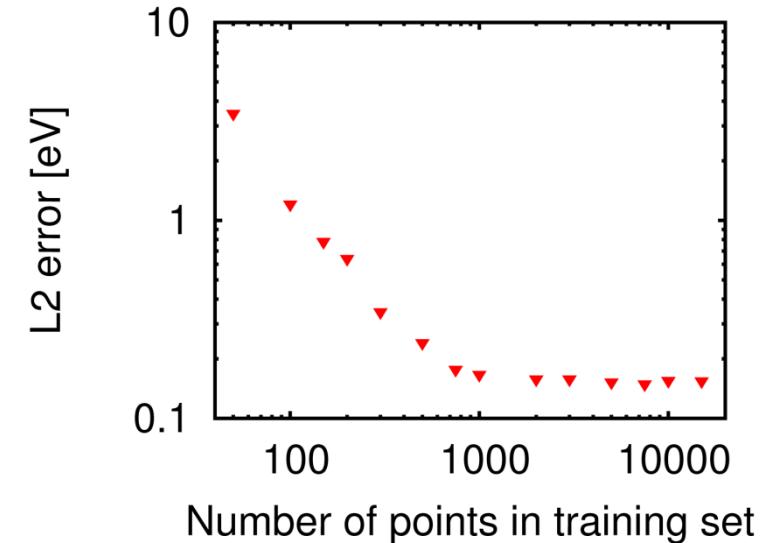
## Training a neural network for xenon crystal from enhanced MD: Two hidden layers, 20 nodes per layer.

Cost function:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{\lambda=1}^Z |G_{\text{ANN}}(\mathbf{s}_\lambda, \mathbf{w}) - G_\lambda|^2$$

Optimization:

$$\frac{\partial E}{\partial \mathbf{w}} = 0$$



Projected  
free-energy  
Surface  
(full FES is  
5 dimensional)



## Using the neural network to compute the isothermal compressibility



$$\kappa_T = -\frac{1}{\langle V \rangle} \frac{\partial \langle V \rangle}{\partial P} = \beta \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle}$$

Ensemble averages performed via a Monte Carlo calculation with the NN FES.

FES has an irregular shape. Hence, classification neural network to classify points as thermodynamically accessible or inaccessible.

Give each point  $s_\lambda$  a label  $y_\lambda$  that defines if the point is thermodynamically Accessible or not based on a thermodynamic threshold:

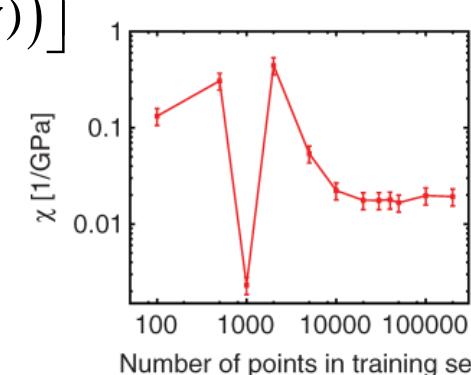
$$y_\lambda = \begin{cases} 1 & G_\lambda - G_{\min} < G_{\text{thresh}} \\ 0 & \text{Otherwise} \end{cases}$$

Cost function:

$$E(\mathbf{w}) = -\frac{1}{M} \sum_{\lambda=1}^M \left[ y_\lambda \ln(y_{\text{ANN}}(s_\lambda; \mathbf{w})) + (1 - y_\lambda) \ln(1 - y_{\text{ANN}}(s_\lambda; \mathbf{w})) \right]$$

Result:  $\kappa_T = 0.02 \pm 0.005 \text{ GPa}^{-1}$  at 2700 K

Experiment at 25 GPa and room temperature is  $0.008 \text{ GPa}^{-1}$





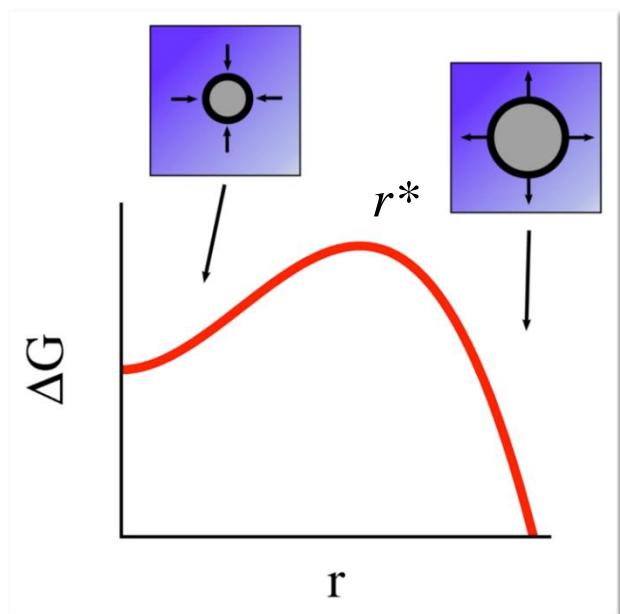
# The mechanism of equilibrium melting of a solid



A. Samanta, T.-Q. Yu, W. E, MET Science (November, 2014)

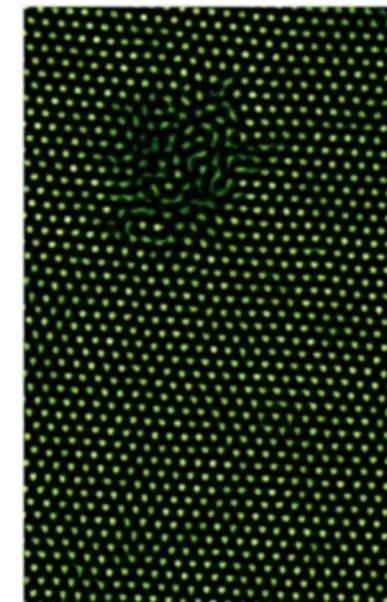
## Classical nucleation theory:

$$\Delta G(r) = \frac{4}{3} \pi r^3 \rho \Delta \mu + 4\pi r^2 \gamma_s, \quad \Delta \mu < 0$$



$$r^* = -\frac{2\gamma_s}{\rho\Delta\mu}$$

$$\Delta G(r^*) = \frac{16\pi\gamma_s^3}{3(\rho\Delta\mu)^2}$$





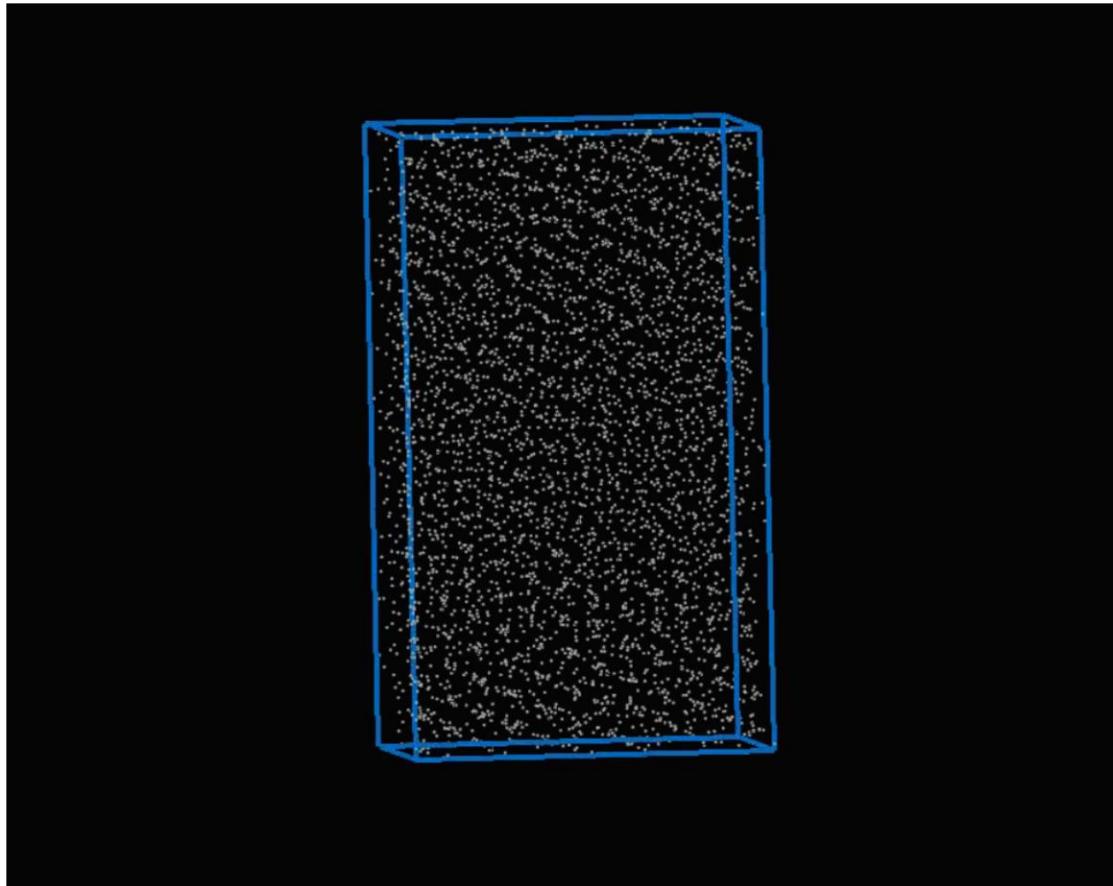
# Crystal d-AFED simulations of the melting of Cu



A. Samanta, T.-Q. Yu, W. E, *MET Science* (November, 2014)

Collective variables:  $\mathbf{h}_d, Q_4, Q_6, \quad T = T_m, \quad T_s = T_{\mathbf{h}} = 10^7 \text{ K}$

Potential model: Embedded Atom Model (EAM) of Mishin et al.  
*Phys. Rev. B* **63**, 224106 (2001).

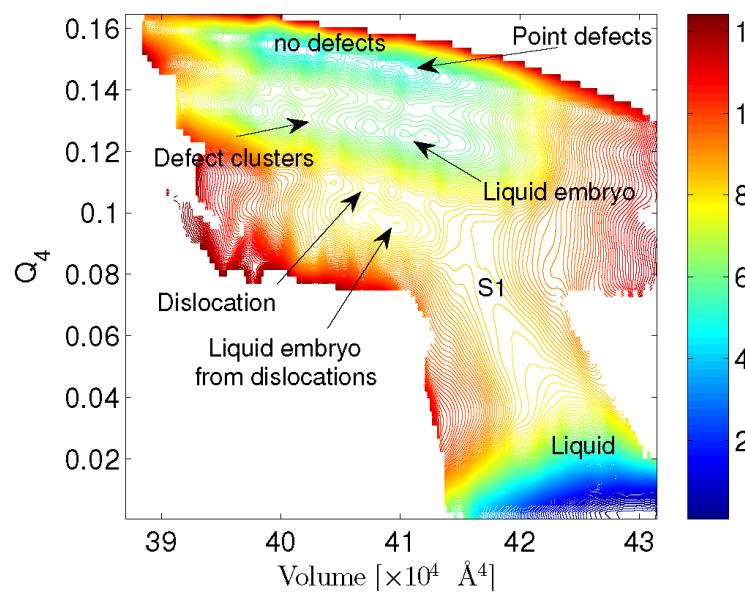




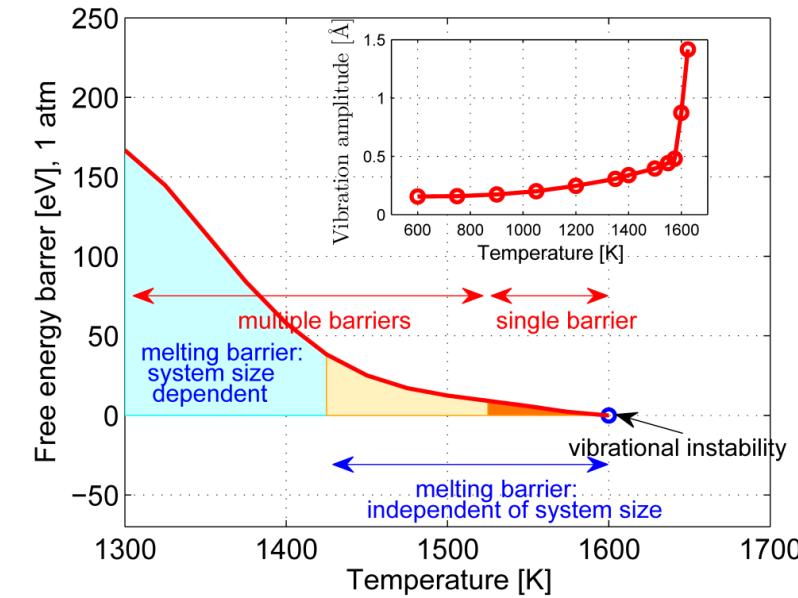
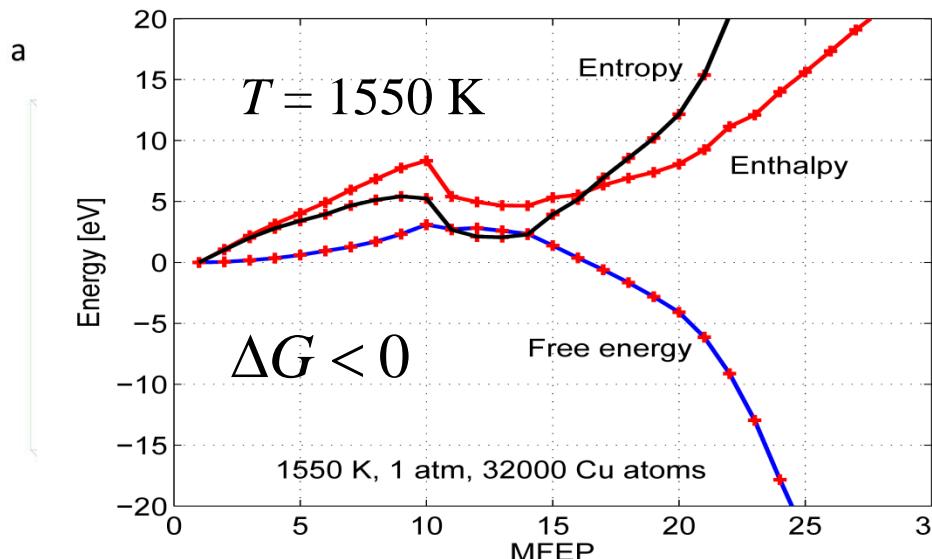
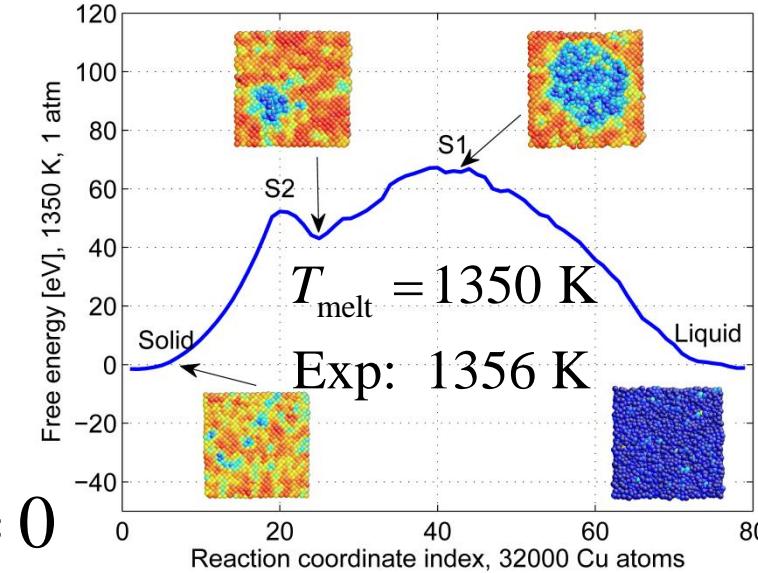
# Free energy landscape



One-dimensional profiles constructed using string method [Maragliano *et al.* *JCP* (2006)]



$$\Delta G = 0$$





# Many Challenges Remain



Faraday Discussions

Cite this: DOI: 10.1039/c8fd00033f



PAPER

[View Article Online](#)  
[View Journal](#)

## Crystal structure prediction is changing from basic science to applied technology

Jonas Nyman <sup>a,b</sup> and Susan M. Reutzel-Edens <sup>b</sup>

1. CSP methods still predict too many structures. Many predicted polymorphs not seen in experiment. Why not? Latest CCDC blind test allowed competitors to submit 100 candidate structures! How do we predict structures that can be observed and NOT those that cannot?
2. Will we stick with traditional structure generators, or will free-energy based exploration methods prevail on this problem? Which is better equipped to handle highly flexible molecules, predict the existence of disorder in crystals, or identify nonstoichiometric amounts of water?
3. How will machine learning impact the field on both the interaction model and structure prediction side?

*J. Chem. Phys.* perspective to be submitted by L. Vogt and MET will question this view.



# Acknowledgements



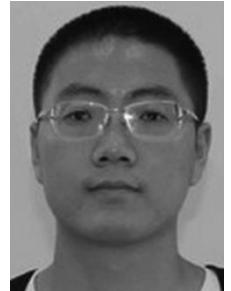
## External

- Krzysztof Szalewicz (U. Del.)
- Michael Metz (U. Del.)
- Bart Kahr (NYU)
- Qiang Zhu (UNLV)
- Jutta Rogal (RUB, ICAMS)
- Pance Naumov (NYU AD)
- Maria Baias (NYU AD)
- Serdal Kirmizialtin (NYU AD)
- Jörg Behler (Uni Göttingen)



## Postdocs

- Elia Schneider
- Leslie Vogt
- Nikolaos Galanakis



## Students

- Ming Chen (UC Berkeley)
- Tang-Qing Yu (CIMS)
- Hongxing Song
- Luke Dai (UC Berkeley)
- Alex Knoll (Uni Göttingen)

